



University
of Glasgow

Denwood, Matthew James (2010) *A quantitative approach to improving the analysis of faecal worm egg count data*. PhD thesis.

<http://theses.gla.ac.uk/1837/>

Copyright and moral rights for this thesis are retained by the Author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

A quantitative approach to improving the analysis of faecal worm egg count data

by

Matthew James Denwood



A thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)

Boyd Orr Centre for Population and Ecosystem Health
Faculty of Veterinary Medicine
University of Glasgow

May 2010

For Sarah,
and in memory of Richard.

Abstract

Analysis of Faecal Egg Count (FEC) and Faecal Egg Count Reduction Test (FECRT) datasets is frequently complicated by a high degree of variability between observations and relatively small sample sizes. In this thesis, statistical issues pertaining to the analysis of FEC and FECRT data are examined, and improved methods of analysis using Bayesian Markov chain Monte Carlo (MCMC) are developed.

Simulated data were used to compare the accuracy of MCMC methods to existing maximum likelihood methods. The potential consequences of model selection based on empirical fit were also examined by comparing inference made from simulated data using different distributional assumptions. The novel methods were then applied to FEC data obtained from sheep and horses. Several syntactic variations of FECRT models were also developed, incorporating various different distributional assumptions including meta-population models. The inference made from simulated data and FECRT data taken from horses was compared to that made using the currently most widely used methods. Multi-level hierarchical models were then used to partition the source of the observed variability in FEC using data intensively sampled from a small group of horses.

The MCMC methods out-performed other methods for analysis of simulated FEC and FECRT datasets, particularly in terms of the usefulness of 95% confidence intervals produced. There was no consistent difference in model fit to the gamma-Poisson or lognormal-Poisson distributions from the available data. However there was evidence for the existence of bi-modality in the datasets. Although the majority of the observed variation in equine FEC is likely a consequence of variability between animals, a considerable proportion of the variability is due to the variability in true FEC between faecal piles and the aggregation of eggs on a local scale within faeces.

The methods currently used for analysis of FEC and FECRT data perform poorly compared to MCMC methods, and produce 95% confidence intervals which are unreliable for datasets likely to be encountered in clinical parasitology. MCMC analysis is therefore to be preferred for these types of data, and also allows multiple samples taken from each animal to be incorporated into the analysis. Analysing the statistical processes underlying FEC data also revealed simple methods of reducing the observed variability, such as increasing the size of individual samples of faeces. Modelling the variability structure of FEC data, and use of the inferred parameter values in precision analysis and power analysis calculations, allows the usefulness of a study to be quantified before the data are collected. Given the difficulties with analysing FEC and FECRT data demonstrated, it is essential that such consideration of the statistical issues pertaining to the collection and analysis of such data is made for future parasitological studies.

Acknowledgements

The work contained within this thesis was funded by the DEFRA Veterinary Training and Research Initiative (VTRI) 0101. I am also grateful for the support and encouragement received from my supervisors, along with Louise Matthews, Iain McKendrick, Rowland Kao and Nils Toft for statistical and mathematical advice during my PhD, and to Mike Stear, Martin Nielsen, Stig Petersen and Chris Proudman for providing the sheep and equine data discussed.

Declaration

This thesis, and the work contained within it, was conducted from September 2006 to May 2010 by myself under the supervision of Prof. Sandy Love, Prof. Stuart Reid and Dr. Giles Innocent at the University of Glasgow.

Matthew Denwood

Contents

List of publications and presentations	vii
Glossary	ix
List of Figures	x
List of Tables	xv
Chapter 1: Introduction	1
1.1 Life cycle and epidemiology of the nematode parasite	3
1.2 Modelling parasite distributions	4
1.3 Advantages of statistical modelling as an inferential tool	18
1.4 Study objectives	19
Chapter 2: Selection, development and evaluation of a population model	21
2.1 Introduction	22
2.2 Defining parasite distributions	23
2.3 Selecting a model definition for the zero-inflated gamma Poisson model	29
2.4 Comparing prior distributions for the gamma Poisson model	34
2.5 Selecting model definitions for the lognormal Poisson model	39
2.6 Comparing JAGS and WinBUGS results	43
2.7 Discussion	46
2.8 Development of an R package to facilitate data analysis	47
Chapter 3: Distributional selection and model fit	49
3.1 Introduction	50
3.2 Comparison of MCMC and ML for analysis of ZIGP data	52
3.3 Use of a ZIGP model to analyse GP and ZIP data	58
3.4 Analysis of empirical fit as a criterion for model selection	68
3.5 Comparative analysis of ZILP and ZIGP simulated data	76
3.6 Discussion	82
Chapter 4: Analysis of field faecal egg count data	85
4.1 Introduction	86
4.2 Sheep FEC data	89
4.3 Equine FEC data	97
4.4 Empirical fit	103
4.5 Discussion	110

Chapter 5: Applying Bayesian MCMC to the Faecal Egg Count Reduction Test	113
5.1 Introduction	114
5.2 Development of an MCMC model for analysis of FECRT data	123
5.3 Comparison of three methods for analysis of FECRT data	125
5.4 Further Analysis of Non-Parametric Bootstrapping	135
5.5 Field data analysis	139
5.6 Meta-population modelling	147
5.7 Discussion	152
Chapter 6: Variability and repeatability of FEC	156
6.1 Introduction	157
6.2 Quantifying variability within an individual	158
6.3 Quantifying variability between individuals	172
6.4 Precision analysis	175
6.5 Discussion	187
Chapter 7: General discussion and conclusions	190
Bibliography	199
Appendix A: Model specifications	215
A.1 Faecal egg count models	216
A.2 Faecal egg count reduction test models	221
A.3 Models for quantifying variability of FEC	224
Appendix B: Derivation of the cv of a compound gamma distribution	229

Publications and presentations

Peer-reviewed publications

- Denwood, M. J., Reid, S. W. J., Love, S., Nielsen, M. K., Matthews, L., McKendrick, I. J. and Innocent, G. T. (2009). Comparison of Three Alternative Methods for Analysis of Equine Faecal Egg Count Reduction Test Data. *Preventive Veterinary Medicine*, **93**(4): pp. 316-23.
- Denwood, M. J., Stear, M. J., Matthews, L., Reid, S. W. J., Toft, N. and Innocent, G. T. (2008). The distribution of the pathogenic nematode *Nematodirus battus* in lambs is zero-inflated. *Parasitology*, **135**(10): pp. 1225-1235.

Oral presentations

- Denwood, M. J., Reid, S. W. J., Love, S., Nielsen, M. K., Matthews, L., McKendrick, I. J. and Innocent, G. T. (2009). Use and Validation of a Novel MCMC Method of Analysis for Faecal Egg Count Reduction Test Data. In *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine*.
- Denwood, M. J., Love, S., Reid, S. W. J. and Innocent, G. T. (2008). Investigating the reliability of a single faecal egg count to quantify the true egg shedding rate of a horse. In *Proceedings of the British Equine Veterinary Association*.
- Denwood, M. J., Love, S., Stear, M. J., Reid, S. W. J. and Innocent, G. T. (2008). Analysing the distribution of *Nematodirus battus* in lambs using Bayesian MCMC methods. In *Proceedings of the Association for Veterinary Teaching and Research Work*.
- Denwood, M. J., Love, S., Stear, M. J. and Innocent, G. T. (2007). Accounting for variability in faecal worm egg counts from groups of animals. In *Proceedings of the Association for Veterinary Teaching and Research Work*.

Poster presentations

- Denwood, M. J., Love, S., Matthews, J. B., Reid, S. W. J. and Innocent, G. T. (2008). Use of a novel Bayesian statistical model to improve detection of anthelmintic resistance in cyathostomins. In *Proceedings of the International Equine Parasite Drug Resistance Workshop* (Poster).
- Denwood, M. J., Reid, S. W. J., Toft, N. and Innocent, G. T. (2008). Beyond the gamma-Poisson: Improving the performance of Bayesian MCMC models for zero-inflated, overdispersed count data. In *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine* (Poster).

- Denwood, M. J., Stear, M. J., Love, S. and Innocent, G. T. (2007). Use of Bayesian MCMC to improve analysis of faecal egg count data. In *Proceedings of the World Association for the Advancement of Veterinary Parasitology* (Poster).
- Denwood, M. J., Toft, N., Love, S., Stear, M. J., Reid, S. W. J. and Innocent, G. T. (2007). Comparing the accuracy of maximum likelihood and MCMC estimates using a zero-inflated gamma Poisson model. In *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine* (Poster).

Software created as part of this work

- Denwood, M. (2010). bayescount: Analysis of faecal egg count distributions, faecal egg count reduction tests, and prospective power calculations using computationally intensive statistical techniques. *R package version 1.0.0*. Freely available to download from <http://cran.r-project.org/web/packages/bayescount>.

Glossary

<i>cv</i>	coefficient of variation
FEC	Faecal Egg Count
FECRT	Faecal Egg Count Reduction Test
GP	gamma Poisson
JAGS	Just Another Gibbs Sampler
LP	lognormal Poisson
MCMC	Markov chain Monte Carlo
RMSE	root-mean-square-error
ZIGP	zero-inflated gamma Poisson
ZILP	zero-inflated lognormal Poisson
ZINB	zero-inflated negative binomial (equivalent to ZIGP)
ZIP	zero-inflated Poisson

List of Figures

1.1	Stylised representation of the general parasitic nematode life cycle	3
2.1	The effect of altering larval challenge on the distribution of adult worms in a mathematical model of the distribution of parasites between hosts	27
2.2	The effect of altering the distribution of host and worm variability on the distribution of adult worms in a mathematical model of the distribution of parasites between hosts	27
2.3	The effect on the proportion of adult worms in each individual host of altering different parameters in a mathematical model of the distribution of parasites between hosts	28
2.4	The relationship between mean count and the shape parameter ($\frac{1}{cv^2}$) at each iteration using a ZIGP model with simulated data consisting of zero counts	37
2.5	Comparative estimates of mean count from the analysis of simulated data using equivalent JAGS and WinBUGS models	44
2.6	Comparative estimates of cv from the analysis of simulated data using equivalent JAGS and WinBUGS models	45
2.7	Comparative estimates of zero-inflation from the analysis of simulated data using equivalent JAGS and WinBUGS models	45
3.1	The relationship between the MCMC median estimates for mean and cv from simulated data with a sample size of 10	56
3.2	The relationship between the MCMC median estimate for mean and the empirical mean of the simulated dataset with a sample size of 10	56
3.3	Comparison of median estimates produced by ZIGP model (y axis) and GP model (x axis) for GP and ZIGP simulated data	62
3.4	Comparison of median estimates produced by ZIGP model (y axis) and ZIP model (x axis) for ZIP and ZIGP simulated data	64
3.5	Empirical cumulative distribution function for the lower (blue) and upper (red) 95% CI and median (black) estimates of cv produced by the ZIGP model from analysis of the GP simulated data	64
3.6	Empirical cumulative distribution function for the lower (blue) and upper (red) 95% CI and median (black) estimates of zero-inflation produced by the ZIGP model from analysis of the GP simulated data	65
3.7	The relationship between the true mean, true cv and lower 95% CI produced by the ZIGP model from the analysis of the ZIP simulated data	66
3.8	Comparison of an equivalent parametrisation of gamma (red), lognormal (blue) and truncated normal (black dashed) probability density functions	69
3.9	Comparison of an equivalent parametrisation of gamma (red), lognormal (blue) and truncated normal (black dashed) -Poisson probability mass functions (calculated using integration over lambda)	70

3.10	Comparison of the maximum log-likelihood produced by GP and LP models for GP and LP simulated data with a sample size of 10	72
3.11	Comparison of the maximum log-likelihood produced by GP and LP models for GP and LP simulated data with a sample size of 1000	73
3.12	The relationship between the proportion of maximum likelihoods produced by GP and LP models that indicated a better fit to the distribution from which the data were simulated (y axis - 95% CI shown dotted) and the sample size of the data (x axis)	74
3.13	Comparison of the proportion of maximum likelihoods giving a better fit to the distribution from which the data were simulated with and without zero-inflation in the GP and LP simulated data and models	75
3.14	Proportion of true parameters that fell above (negative value) and below (positive value) the 95% CI returned by the ZILP and ZIGP models in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions	78
3.15	Comparison of median estimates for cv produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data	79
3.16	Comparison of median estimates for zero-inflation produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data	80
3.17	Comparison of lower 95% CI estimates for zero-inflation produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data	80
3.18	Comparison of median estimates for mean produced by ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data	81
4.1	Comparison of the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with log-uniform prior on cv , from analysis of the distribution of <i>N. battus</i> and Nematodirus spp. between five cohorts of 200 lambs	90
4.2	Comparison of the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with uniform prior on cv , from analysis of the distribution of <i>N. battus</i> and Nematodirus spp. between five cohorts of 200 lambs	91
4.3	The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model from analysis of the distribution of <i>N. battus</i> and Nematodirus spp. between five cohorts of 200 lambs	92
4.4	The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZIGP model, with log-uniform prior on cv , from analysis of the distribution of <i>N. battus</i> and Nematodirus spp. between five cohorts of 200 lambs	92
4.5	The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZIGP model, with uniform prior on cv , from analysis of the distribution of <i>N. battus</i> and Nematodirus spp. between five cohorts of 200 lambs	92
4.6	Histogram of observed <i>Nematodirus battus</i> eggs enumerated in the faeces of 200 lambs in May 1995	95
4.7	Comparison of the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with log-uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	98

4.8	Comparison of the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	99
4.9	The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	99
4.10	The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZIGP model, with log-uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	100
4.11	The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZIGP model, with uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	100
4.12	Empirical cumulative distribution function of the lower and upper 95% CI estimates for zero-inflation produced by the ZILP model (blue), ZIGP model with log-uniform prior on cv (green) and ZIGP model with uniform prior on cv (orange) for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets	101
4.13	Maximum log likelihood produced using the ZIGP compared to ZILP, and GP compared to LP models, from analysis of all <i>N. battus</i> and <i>Nematodirus</i> spp. worm burdens and FEC in five cohorts of 200 lambs	104
4.14	Maximum log likelihood produced using the ZIGP compared to ZILP, and GP compared to LP models, from analysis of all 444 UK FEC, 64 Danish FECRT and 2 UK FECRT equine datasets	104
4.15	The relationship between the median estimate for cv produced by the ZILP model and the log of the likelihood ratio, for all equine (left) and all sheep (right) data	107
4.16	The relationship between the log of the likelihood ratio and the median estimate for <i>zero – inflation</i> produced by the ZILP model (left) or sample size (right)	108
4.17	The relationship between (log) maximum likelihood and ML estimate for zero-inflation produced by the ZIGP (red) and ZILP (blue) models for all equine and sheep data	108
5.1	The relationship between the simulated % reduction and the empirical mean % reduction from 3000 simulated FECRT datasets	130
5.2	The proportion of 95% confidence intervals containing the simulated true FEC reduction parameter for each method from the analysis of all simulated FECRT datasets (solid lines) and the 2900 simulated datasets that did not give an empirical mean reduction of 100% (dashed lines)	131
5.3	Effect of cv (left) and Δcv (right) on the proportion of bootstrapped 95% confidence intervals that contained the true efficacy parameter from 250 simulated FECRT datasets	137
5.4	Proportion of 95% confidence intervals produced using the bootstrap method that contained the true parameter from 1000 simulated FECRT datasets at each pre-treatment mean number of eggs counted (95% credible intervals shown shaded)	138
5.5	Median estimates of mean FEC reduction and median-relative 95% CI size produced for 64 Danish FECRT datasets using the single distribution (y axis) and paired (x axis) models	142

5.6	Effect of the amount of information in the data (relative size of 95% CI produced by the paired model) on the relative difference between the paired and single distribution median estimates of mean FEC reduction produced from 64 Danish FECRT datasets	143
5.7	Median estimates of mean FEC reduction and median-relative 95% CI size produced for 64 Danish FECRT datasets using the full dataset (y axis) and truncated dataset (x axis) with the paired model	144
5.8	The difference in median estimate of mean FEC reduction between the full and part data relative to the median estimate for the full data (y axis), plotted against the median estimate for the full data (x axis), produced for 64 Danish FECRT datasets	144
5.9	Median estimates of change in shape and median-relative 95% CI size produced for 64 Danish FECRT datasets using the full dataset (y axis) and truncated dataset (x axis) with the paired model	145
5.10	Median (black) and 95% CI (red-blue) estimates for zero-inflation produced by the FECRT model and the simple ZIGP model with the full pre-treatment data from 64 Danish FECRT datasets	145
5.11	Median (black) and 95% CI (red-blue) estimates for efficacy produced by the metapopulation model (y axis) and the individual paired models (x axis) from analysis of 64 Danish FECRT datasets	149
5.12	Posterior probability densities for the mean (blue) and <i>cv</i> (red) of the efficacy (top) and change in shape parameter (bottom) from metapopulation modelling of 64 Danish FECRT datasets	150
6.1	The required number of samples calculated by the Raftery and Lewis diagnostic for the hierarchical compared to non-hierarchical MCMC model for each of the 160 simulated hierarchical FEC datasets	162
6.2	The proportion of 95% CI containing the simulation parameter, mean relative bias, mean relative RMSE and mean relative 95% CI size of the estimate for each source of variability produced by the hierarchical (y axis) and non-hierarchical model (x axis), from analysis of the 160 simulated hierarchical FEC datasets	163
6.3	The <i>cv</i> associated with each of four sources of variability, all using a log-scale prior on the inverse shape parameter, observed in the FEC of four animals over a five day period	167
6.4	The <i>cv</i> associated with each of four sources of variability, all using a Uniform prior directly on <i>cv</i> , observed in the FEC of four animals taken over a five day period	168
6.5	The <i>cv</i> associated with each of four sources of variability, assuming replicate variability is fixed between animals, observed in the FEC of four animals taken over a five day period	169
6.6	Probability density of the combined daily and sample variability (left) and the total within-animal variability (right), assuming replicate variability is fixed between animals and 3g faeces are used per sample, in the FEC of four animals taken over a five day period	169
6.7	The relationship between the amount of faeces used and the effective <i>cv</i> associated with the variability in observed FEC between sub-samples from the same faecal sample	170

6.8	The relationship between the amount of faeces used and the total effective cv associated with the variability in observed FEC within an individual over a period of a few days	171
6.9	Estimates for the between animal variability (cv_b) obtained from 24 equine datasets with sample sizes of 50 or more, and within animal variability (cv_w) obtained from a total of 204 samples taken from four animals over a five day period	174
6.10	The effect of $R \times S \times \mu \times T^2$ on the probability that an observed mean FEC will lie within a tolerance (T) of the true mean FEC of an individual animal with a negligible cv_w of 0.001, calculated using the cumulative probability density function of the negative binomial distribution	180
6.11	The effect of $\mu \times S$ and cv_w on the $T^2 \times R$ required to maintain a 95% probability of observing a mean FEC within $\mu \pm T \times \mu$ (the true mean FEC) from an individual animal, calculated using an optimisation algorithm and the cumulative probability density function of the negative binomial distribution	181
6.12	The effect of varying the minimum egg detection threshold (S) on the required tolerance (T) to maintain a 95% probability of observing $\mu \pm T \times \mu$ (the true mean FEC) from an individual animal, calculated using an optimisation algorithm and the cumulative probability density function of the negative binomial distribution	182
6.13	Distribution describing the uncertainty in the power to detect an empirical mean reduction of less than 95% after treatment of 20 animals with a drug of true population efficacy 85%	186

List of Tables

2.1	The effect of altering either larval challenge (Dataset 2) or the distribution of host and worm susceptibility (Dataset 3) compared to the baseline (Dataset 1) on the distribution of worms in a mathematical model of the distribution of parasites between hosts	28
2.2	Number of simulated datasets successfully analysed by four variants of a ZIGP model	32
2.3	Proportion of true values above and below the 95% confidence intervals produced by four variants of a ZIGP model from the analysis of simulated data	32
2.4	Relative RMSE of the median estimate for each parameter produced by four variants of a ZIGP model from the analysis of simulated data	33
2.5	Number of simulated datasets successfully analysed by four variants of a ZILP model	36
2.6	Proportion of true values above and below the 95% confidence intervals produced by four variants of a ZILP model from the analysis of simulated data	36
2.7	Relative RMSE of the median estimate for each parameter produced by four variants of a ZILP model from the analysis of simulated data	37
2.8	Number of simulated datasets successfully analysed by two variants of a ZILP model, each with different priors, from the analysis of simulated data	41
2.9	Proportion of true values above and below the 95% confidence intervals produced by two variants of a ZILP model, each with two different priors for variability, from the analysis of simulated data	41
2.10	Relative RMSE of the median estimate for each parameter produced by two variants of a ZILP model, each with two different priors for variability, from the analysis of simulated data	42
2.11	Number of datasets successfully analysed by equivalent JAGS and WinBUGS models from the analysis of simulated data	44
3.1	Number of datasets returned by MCMC and ML techniques for each sample size from the analysis of simulated ZIGP data	54
3.2	Proportion of 95% confidence intervals produced by ML and MCMC techniques that contained the true parameter from the analysis of simulated ZIGP data	54
3.3	Relative RMSE produced by ML and MCMC techniques for each parameter and sample size from the analysis of simulated ZIGP data	55
3.4	Number of datasets returned by each combination of GP, ZIP and ZIGP model with GP, ZIP and ZIGP simulated datasets	61
3.5	Comparison of the relative RMSE, bias, proportion of true values above and below 95% CI and mean relative 95% CI size returned by each combination of parameter and model/simulated dataset	63

3.6	Number of datasets returned by the ZILP and ZIGP models in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions	77
3.7	Proportion of 95% CI returned by the ZILP and ZIGP models that contained the true parameter value in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions	78
4.1a	The lower and upper 95% CI estimates of zero-inflation produced for each dataset by the ZILP model and ZIGP model, with log-uniform prior (ZIGP model 1) and uniform prior (ZIGP model 2) on <i>cv</i> , from analysis of the distribution of <i>N. battus</i> and <i>Nematodirus</i> spp. between five cohorts of 200 lambs .	93
4.1b	The lower and upper 95% CI estimates of zero-inflation produced for each dataset by the ZILP model and ZIGP model, with log-uniform prior (ZIGP model 1) and uniform prior (ZIGP model 2) on <i>cv</i> , from analysis of the distribution of <i>N. battus</i> and <i>Nematodirus</i> spp. between five cohorts of 200 lambs (continued)	94
4.2	The proportion of maximum likelihoods that were higher for the LP compared to GP model (left) and for the ZILP compared to ZIGP model (right) for all equine and sheep data	105
4.3	Coefficients for mean, <i>cv</i> , zero-inflation and sample size in a GLM fit to the (log) likelihood ratios for <i>N. battus</i> and <i>Nematodirus</i> spp. FEC obtained from five cohorts of 200 lambs, fitted separately and together	106
4.4	Coefficients for mean, <i>cv</i> , zero-inflation and sample size in a GLM fit to the (log) likelihood ratios for Danish FECRT and UK FEC equine data, fitted separately and together	106
5.1	Terms and definitions used in describing anthelmintic resistance, taken from Dargatz et al. (2000)	118
5.2	A comparison of the FEC reduction (with confidence intervals) as calculated using four different techniques for three equine FECRT datasets, adapted from Torgerson et al. (2005)	120
5.3	Mean size of 95% confidence intervals relative to the median (MCMC and bootstrap) or mean (WAAVP) estimate of the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of 3000 simulated FECRT datasets	132
5.4	Mean size of 95% confidence intervals relative to the median (MCMC and bootstrap) or mean (WAAVP) estimate of the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of the 2900 simulated FECRT datasets that did not give an empirical mean reduction of 100%	132
5.5	Relative root-mean-square-error (RMSE) for median (MCMC and bootstrap) or mean (WAAVP) estimate for the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of 3000 simulated FECRT datasets	132
5.6	Relative root-mean-square-error (RMSE) for median (MCMC and bootstrap) or mean (WAAVP) estimate for the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of the 2900 simulated FECRT datasets that did not give an empirical mean reduction of 100%	132
6.1	Empirical summary of observed FEC from four animals over a five day period, with 18 samples each taken from Animals 1 and 2 and 84 samples each taken from Animals 3 and 4	166

6.2 Median and 95% CI estimates for the total within animal variability in each animal, assuming replicate variability is fixed between animals and 3g faeces are used per sample, in the FEC of four animals taken over a five day period 168

CHAPTER 1

Introduction

Introduction

Nematodes or roundworms comprise one of the most common phyla of animals, with a large number of different species of both parasitic and free-living organisms found in a diverse range of habitats from Antarctic soils (Yergeau et al., 2007) to the Chihuahuan Desert of New Mexico (Freckman and Virginia, 1989). Parasitic organisms within this phylum can be found in many plant and animal species (Blaxter et al., 1998), where they often greatly outnumber their hosts. Several of these are also of great significance as domestic animal pathogens; causing clinical symptoms that range from poor growth rates to anaemia and ischaemic colic (Uhlinger, 1991; Love, 1995; Urquhart et al., 1996). A great deal of effort has therefore been expended on studying the life cycle and pathogenesis of these parasites in an attempt to improve the control and treatment of helminth parasitism in domestic species. Despite this, nematode parasitism continues to be one of the main sources of economic loss due to poor growth rate in ruminants (Molento, 2009), and are also a clinical concern in equids in the developed world. This concern arises partly as a result of the development of anthelmintic resistance within the nematode population (Prichard et al., 1980; Sangster, 1999b; Dargatz et al., 2000) and is compounded by the difficulties in defining the distribution of parasites within and between hosts (Barger, 1985; Sréter et al., 1994; Stear et al., 1998).

These difficulties tend to be more pronounced in analyses of data obtained from horses. In these datasets, small groups of animals exhibit marked variability between parasite burdens as a result of heterogeneity in the host population. The available datasets are typically in the form of Faecal Egg Count (FEC) data, which represent the parasite egg shedding rate of an animal and are therefore an indirect measure of the parasite burden (Uhlinger, 1993). Application of more sophisticated statistical techniques to the analysis of parasite distributions, and further consideration of the statistical issues surrounding the collection and analysis of parasite burden data, would help to better understand these distributions, ultimately leading to more effective parasite control strategies (Barger, 1985; Sréter et al., 1994; Döpfer et al., 2004; Nielsen et al., 2006a).

1.1 Life cycle and epidemiology of the nematode parasite

Parasitic nematodes of various genera are commonly found in several species of domesticated animals, including horses and domestic ruminants such as cattle and sheep (Urquhart et al., 1996). Within these nematode species there is a wide range of physical morphology and pathogenicity, but with a few exceptions such as Ascarids and *Nematodirus battus*, all of these nematodes share the same basic life cycle shown in Figure 1.1. Each is obligated to go through part of its life cycle on pasture, first hatching from an egg and then undergoing a series of transformations known as moults between three distinct larval stages denoted L₁, L₂ and L₃, before being ingested by the definitive host species as infective L₃ larvae. Another series of moults through L₄ and L₅ then ensues before the nematode becomes a sexually mature male or female adult, and undergoes sexual reproduction with another nematode of the same species. Female worms then produce a large number of eggs which are passed out in faeces onto the pasture, continuing the life cycle into the next generation (Urquhart et al., 1996).

Although this basic life cycle is common to nearly all nematodes, there are key features of the

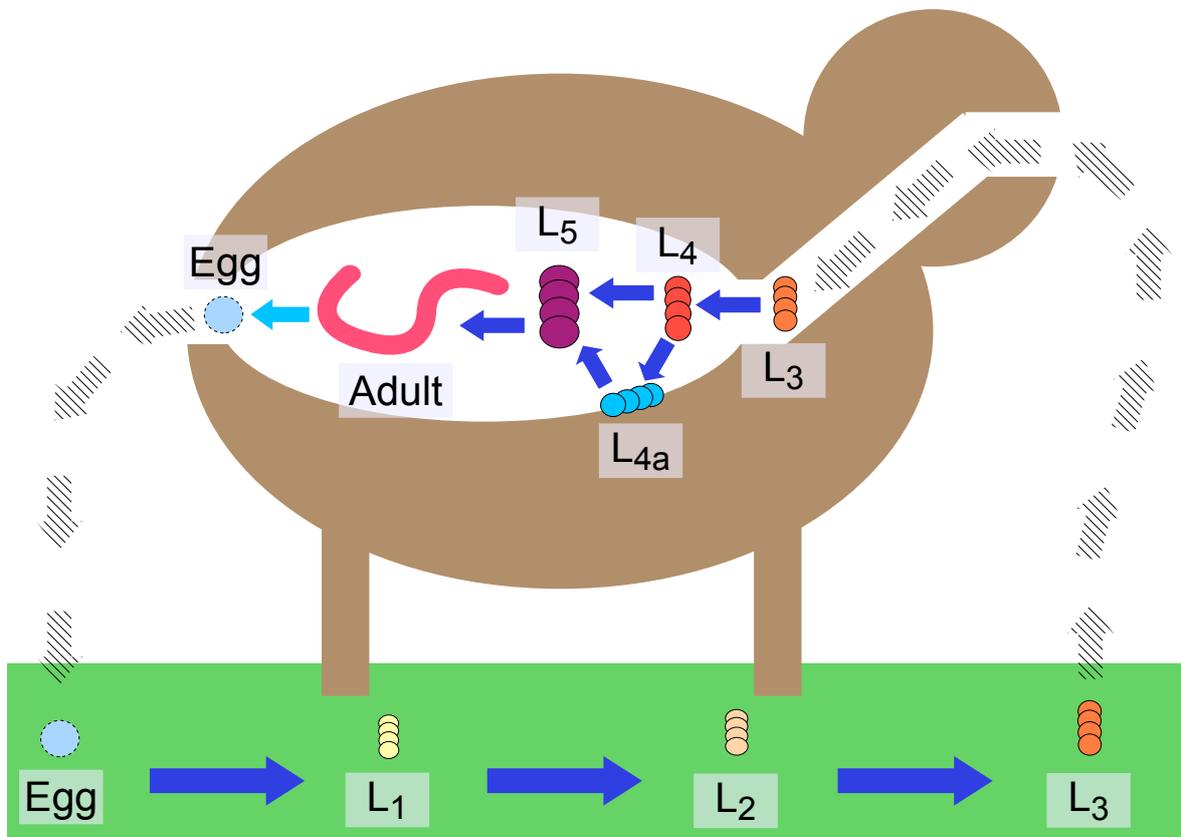


Figure 1.1: Stylised representation of the general parasitic nematode life cycle - larval stages represented with the prefix L and maturation stages shown using blue arrows

biology of some parasites that provide specific adaptation strategies. For example, L₄ stage larvae of small strongyle parasites of the domestic equid can undergo a period of hypobiosis in the large intestinal mucosa (denoted L_{4a} in [Figure 1.1](#)), during which they are less susceptible to anthelmintic drugs ([Uhlinger, 1991](#); [Love, 1995](#); [Mercier et al., 2001](#); [Duncan et al., 2002](#)). There is also a large variation in size between adult nematodes of different species, from a few millimetres to more than a meter in size, as well as differences in predilection sites along the alimentary canal of the host ([Urquhart et al., 1996](#)). Some species of nematode also undergo migratory stages through different areas of the body of the final host, such as the large strongyles of horses ([Urquhart et al., 1996](#)).

A combination of the life cycle of nematode parasites and the variability in host immunity results in an aggregated distribution of adult parasites between hosts ([Sinniah, 1982](#); [Barger, 1985](#); [Uhlinger, 1993](#); [Shaw and Dobson, 1995](#)). This frequency distribution influences the impact of parasites upon the host population ([Anderson and May, 1991](#)), and influences the success of potential control measures such as selective breeding. This variability in parasite burdens is also important to consider when modelling the host-parasite relationship ([Boag and Thomas, 1975](#); [May and Anderson, 1983](#); [Woolhouse, 1992](#); [Kao et al., 2000](#)). [Herbert and Isham \(2000\)](#) explicitly modelled both the larval and adult stages of the parasite as part of their stochastic host-parasite interaction model. However, it is more common to consider only the distribution of adult nematodes between hosts when analysing data using statistical modelling ([Wilson et al., 1996](#); [Wilson and Grenfell, 1997](#); [Nødtvedt et al., 2002](#); [Torgerson et al., 2005](#); [Vidyashankar et al., 2007](#)). This helps to simplify a highly complex system, but risks the loss of some potentially valuable information. Statistical analysis of both parasite distributions and FEC distributions has generally been achieved using relatively simple techniques such as linear regression and generalised linear models ([McCullagh and Nelder, 1989](#); [Wilson et al., 1996](#)), often after log-transforming the observed data ([Wilson and Grenfell, 1997](#)). However, more computationally complex methods such as maximum likelihood analysis have more recently become accepted as providing more useful results ([Bliss and Fisher, 1953](#); [Wilson et al., 1996](#); [Torgerson et al., 2005](#)).

1.2 Modelling parasite distributions

1.2.1 Defining the statistical distribution of nematode parasites

The number of parasites within an individual is a count observation, which would be expected to follow a Poisson distribution if parasites were truly randomly mixed within a population ([Elliott, 1977](#)). Equally, multiple FEC measurements from an individual obtained using a counting procedure such as the modified McMasters technique ([Mines, 1977](#)) would be expected to follow a Poisson distribution if eggs were randomly mixed within faeces. The

Poisson distribution is described by a single parameter, λ , which equates to both the mean and variance (Evans et al., 2000). However, it has been well established that parasites are statistically aggregated in many species, including humans (Simmiah, 1982; Churcher et al., 2005), sheep (Barger, 1985; Sréter et al., 1994; Morgan et al., 2005), goats (Hoste et al., 2001), birds (Elston et al., 2001), rats (Harvey et al., 1999), wildlife (Shaw et al., 1998), and not least in horses where there are commonly some animals with egg counts much greater than the rest of the herd (Uhlinger, 1993; Ihler, 1995; Chandler and Love, 2002; Döpfer et al., 2004; Torgerson et al., 2005). This aggregation is due to several factors (Crofton, 1962; Herbert and Isham, 2000), including immunity and antigenic heterogeneity (Galvani, 2003), and leads to FEC and parasite count data following a distribution which is over-dispersed with respect to the Poisson distribution (Elliott, 1977).

By far the most widely chosen distribution to represent parasite count data is the negative binomial (May and Anderson, 1978; Wilson and Grenfell, 1997; Shaw et al., 1998), first suggested by Fisher (1941) and later studied by Anscombe (1950). This distribution is defined as the number of failures before s successes, and was originally derived to model the waiting time in a series of Bernoulli trials such as tossing a coin (Evans et al., 2000). In this form, the negative binomial is parametrised by the two parameters s , the number of successes required, and p , the probability that each Bernoulli trial results in a success. However, it can be demonstrated that the negative binomial also generates the same probability mass function as the gamma-Poisson compound distribution (Venables and Ripley, 2002; Vose, 2004), where the shape and scale parameters α (sometimes denoted k) and β (sometimes denoted θ) of the gamma distribution are related to s and p by Equation 1.1.

$$\begin{aligned}\alpha &= s \\ \beta &= \frac{1-p}{p}\end{aligned}\tag{1.1}$$

The mean and variance of the gamma distribution are given by $\alpha\beta$ and $\alpha\beta^2$ respectively. Therefore, the mean of this compound distribution is also given by $\alpha\beta$, and the variance by the sum of the variances expected from the Poisson and Gamma parts of the distribution; $\alpha\beta + \alpha\beta^2$. This definition allows non-integer values of α so that both α and β are strictly positive continuous variables. It is this gamma-Poisson distribution that is relevant to parasite distributions, although when used within parasitology is typically parametrised using the mean μ , and an inverse measure of aggregation k , equal to α , rather than α and β (Fisher, 1941). The benefit of performing calculations using the probability mass function of the negative binomial is avoidance of the computationally complex integration of the lambda value of the Poisson distribution over the gamma distribution, which is necessary when performing calculations using the combined probability mass function of the gamma-Poisson compound distribution.

The gamma-Poisson distribution describes parasite distributions well (Wilson and Grenfell,

1997; Shaw et al., 1998) as it is highly flexible (Toft et al., 2006), although it does not always provide a good fit (Stear et al., 1995a). We can think of the Poisson distribution representing the counting error term (Wilson and Grenfell, 1997), while the gamma distribution reflects the true variability between samples. As the true difference between samples decreases, and parasite burdens become more homogeneous, this variability reduces until the distribution converges on the Poisson (Elliott, 1977).

While there are several studies demonstrating that an over-dispersed Poisson distribution is better suited to the analysis of FEC data than a Poisson distribution (Crofton, 1962; Pacala and Dobson, 1988; Grenfell et al., 1995), none of these have given a biological justification for the underlying gamma distribution. Crofton (1962) criticised the use of the logarithmic series empirical distribution, suggested by Williams (1947), on the basis that it was ‘empirical’ and had no theoretical background in parasitology. However a relation between the logarithmic, Poisson and negative binomial series is provided by Quenouille (1949). Crofton (1962) also justified the negative binomial on the basis that it was ‘fundamental’ and provided a series of biological conditions under which the distribution of parasites could become aggregated, but did not justify the use of the underlying gamma distribution to describe this aggregation. For example, one such potential justification is as follows. If ingestion of infective larvae is random, the time period between ingestion events would be described by an exponential distribution, giving rise to a gamma distribution for the group as a sum of the exponential distribution describing the larval intake for each animal. However, this represents only the intake of infective larvae, and therefore does not take into account the differing effects of host immune response and the distribution of worm genetics. Conceptually, any strictly positive distribution could be used in place of the gamma to describe the true variability between samples, although it is preferable to choose this distribution based on biological plausibility, rather than on the basis of empirical fit or computational convenience alone. The lognormal distribution has been previously suggested as an appropriate alternative to the gamma in parasitology (Elston et al., 2001), and has been found to give similar results to the gamma distribution for other biological problems (Thygesen and Zwinderman, 2006). A potential biological justification for the use of a lognormal distribution is outlined below.

According to the central limit theorem, summation of more than approximately 20 of any probability distributions will result in a distribution that is indistinguishable from the normal distribution. A lognormal distribution is a normal distribution when log transformed, as:

$$\sum (\log (distn_1) , \log (distn_2) \dots \log (distn_n)) = \log \left(\prod (distn_1, distn_2 \dots distn_n) \right)$$

The sum of a series of any distribution, and therefore the log of the same series of distributions, is normally distributed. The log of the product of these distributions, equal to the sum of the log of the series, must therefore also be normal. The exponent of this final distribution will therefore be lognormal. For worm burden data, these distributions could be a series of

distributions describing the survival rate from eggs on pasture to adult worms, with variation in the probability of survival at each stage introduced by variation between the grazing behaviour and immune defences of individuals, and worm genetics. These distributions of probability could represent various aspects of the host immune system and nematode ability to evade the host responses, each of which describes the variability in genetics and environmental influences between worms and hosts (Anderson and May, 1991). Such a system would result in a lognormal distribution of the mean number of worms per individual, although this system does not account for any possible negative or positive feedback on survival from other worms. It is also not clear what distribution would then apply to FEC as an additional process reflecting the egg shedding rates of worms is involved, which may vary over time and in response to positive or negative feedback influences of other parasites in the same host. The mean FEC has been shown to be related to the worm burden (Keymer and Hiorns, 1986; Cabaret et al., 1998), but not enough is known about the influences of competition on fecundity to be sure that this relationship is linear. There is also comparatively little information regarding the variability in fecundity of worms *in vivo*, and fecundity has been shown to be dependent on worm length (Stear and Bishop, 1999). Keymer and Hiorns (1986) also found that fecundity was dependent on the faecal volume produced by the host, which was itself dependent on worm burden. These complicating factors make speculation regarding the distribution of eggs within faeces from an individual difficult; it is commonly assumed that the distribution of FEC follows that of parasite counts, so this simplifying assumption will be maintained here.

The generalised Poisson distribution can also be used to model count data where the data are under- or over-dispersed relative to the Poisson distribution (Famoye, 1993), and has been shown to be a mixture of Poisson distributions with varying mean (Joe and Zhu, 2005), although the distribution of means is not easily defined. Other distributions resulting in a count observation such as the binomial may also have uses in special circumstances where there is a maximum number of observable parasites, but are less useful in the general case. Continuous distributions have also been used to model worm egg count data; Stear et al. (2006) considered the fit of gamma and lognormal distributions to FEC data obtained from sheep, Reinders et al. (2003) compared the fit of the lognormal distribution to the gamma-Poisson, and Gaba et al. (2005) found the Weibull distribution to provide a good fit to data. However, these distributions do not take into account the Poisson error structure inherent in the data, and comparisons of empirical fit using the probability mass at each integer produced by discrete distributions to empirical fit using the probability density produced by continuous distributions are invalid. The zero-inflated discrete Weibull distribution has also been used, and was shown to give a better fit than the zero-inflated Poisson distribution in recruitment modelling (Fortin and DeBlois, 2007), but also fails to take into account the Poisson error structure inherent in FEC data.

Mixture models have been used in parasitology to overcome the issues of pseudo-replication

(Paterson and Lello, 2003), and are also useful when data are comprised of two or more distinct but possibly overlapping groups. The group classification of each data point is not implicit in the data, so that use of a multinomial trial with associated probabilities is needed to model the proportion of the data that belongs to each group. In a simple application to FEC data, this mixture distribution consists of one group of ‘infected’ individuals which are described by a Poisson distribution compounded with any continuous distribution, and one group of ‘uninfected’ animals which have a FEC of zero. The proportion of the group which are infected represents the prevalence of parasitism, and provides an estimate of prevalence which is robust to the problems associated with integer population sizes reported by Jovani and Tella (2006). Zero-inflated models have the advantage that the mean and variance of the second distribution are defined as 0, restricting the number of degrees of freedom in the model relative to other mixture models. These have been used in the fields of parasitology (Nødtvedt et al., 2002; Jell et al., 2008; Walker et al., 2009), bovine mastitis (Rodrigues-Motta et al., 2007), forest science (Fortin and DeBlois, 2007), and medical epidemiology (Böhning et al., 2000), although their use has also been criticised in modelling road traffic accident analysis data (Lord et al., 2005). Zero-inflated distributions are useful when zero observations can arise from either count data (such as Poisson, lognormal-Poisson or gamma-Poisson distributions) or from a truly zero individual (Martin et al., 2005). Analysing the distribution effectively involves predicting the number of true zeros from the distribution of observed counts, and then inferring the number of ‘extra’ zeros from the number of ‘observed’ zeros minus the number of ‘expected’ zeros. Correct identification of the degree of zero-inflation (if any) will therefore depend on the correct choice of distribution for the infected group, so that the number of ‘expected’ zeros is accurate. It is also possible to use a zero-inflated distribution to approximate a sub-population which has a very low mean rather than a mean of zero. The advantage of using a zero-inflated distribution for this rather than a more accurate mixture model of two lognormal or gamma distributions is that the model contains fewer degrees of freedom, and will therefore be less subject to identifiability problems that often affect the application of mixture models (Marin et al., 2005; Peng and Zhang, 2008). The disadvantage is that the true mean of the animals in the low count group will be incorrectly identified as zero when it is in fact small but positive, although this should not have a large impact on the inferred overall mean of the groups as the value of this is dominated by the group with the larger mean.

1.2.2 Methods of model selection

For any biological dataset, there are often a number of different models that can be used to describe the data. These alternatives can either be nested, so that the most complex model contains the same terms as the simpler model but with the addition of another term, or non-nested so that each model contains different terms. An example of a nested pair

of models is the gamma-Poisson and zero-inflated gamma-Poisson, because the zero-inflated gamma-Poisson contains the gamma-Poisson distribution with an additional zero-inflation term. By contrast, the lognormal-Poisson and gamma-Poisson are not nested because each describes over-dispersion using a different distribution. Where the candidate models vary in their degree of complexity, as is always the case with nested models, the most complex distribution will describe the observed data more completely because the extra freedom in the model affords greater flexibility (Johnson and Omland, 2004). An extreme example of this is when a model contains as many degrees of freedom as there are data points, in which case the model is said to be saturated and the data are described completely. However, simpler models that describe the data nearly as well may be favoured over their more complicated counterparts when parsimony is required.

The origins of parsimony are often attributed to Occam’s (or Ockham’s) razor; “*entia non sunt multiplicanda praeter necessitatem*”, or “entities must not be multiplied beyond necessity”, although there is apparently little evidence that this quote was actually made by William of Ockham (Thorburn, 1918). Occam’s razor is central to scientific reasoning, and the concept is often simplified to mean that the simplest explanation that adequately describes the data should be preferred. However, this should require that the simplicity of the model and the fit to the data be quantified.

There are several methods of choosing between different models using likelihood driven model selection (Johnson and Omland, 2004). The likelihood-ratio test or ‘F’ test was originally developed for comparing nested models (Neyman and Pearson, 1928), and can be extended to the use of non-nested models under the asymptotic assumption that the sample size tends to infinity (Cox and Hinkley, 1974). This technique compares the ‘best fit’ of two nested models, choosing the more complex if the improvement in fit warrants the extra parameters of the more complex model. The Akaike’s Information Criterion (AIC) (Akaike, 1973) and similar Bayes Information Criterion (BIC) (Schwarz, 1978) along with Rissanen’s principle of minimum description length (Rissanen, 1978) account for the extra freedom in the model, producing a statistic that incorporates weighting for the number of parameters and can therefore be compared between models. The Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is a further method of comparing the fit of nested models which is particularly useful for hierarchical models estimated using Markov chain Monte Carlo (MCMC). Each of these techniques maintain the asymptotic assumptions of likelihood ratio tests, and can be problematic when the number of degrees of freedom in a model is not defined, as is the case with mixture models (Millar, 2009). The Vuong’s statistic (Vuong, 1989) has also been used to justify the selection of zero-inflated models (Miaou and Lord, 2003), although this test statistic does not apply a penalty for increasing model complexity (Vuong, 1989; Miaou and Lord, 2003).

An alternative approach to model selection is provided by the Bayesian philosophy. This

paradigm allows a posterior probability density describing uncertainty in a parameter value given the observed data to be derived from the probability of observing the data given a parameter value and some prior probability of the parameter value. Bayes' theorem, given below, is the basis of this philosophy, and has its origins in [Bayes \(1763\)](#).

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Where $P(A)$ is the prior probability of A , $P(B|A)$ is the probability of observing B given A , and $P(B)$ is the probability of observing B . In practise, $P(B)$ is often difficult to calculate since it involves integration over all possible values of A , and as this is a normalising constant that is equal for all possible values of A it is often ignored so that:

$$P(A|B) \propto P(B|A) P(A)$$

Bayes' theorem is usually applied to making inference on parameter values in a fixed model, where A is the parameter value and B is the data ([Raftery, 1995](#)). However, A can also be thought of as the entire model, so that A encompasses all possible values of the model parametrisation rather than a single combination of parameter values. If M_1 represents the posterior probability of model one, and M_2 the posterior probability of model two, then the probability that M_1 is the correct model (given that one of M_1 or M_2 is correct) is given by [Raftery \(1995\)](#) as:

$$P(M_1|D) = \frac{P(D|M_1) P(M_1)}{P(D|M_2) P(M_2) + P(D|M_1) P(M_1)}$$

This can also be written as the posterior odds for preferring model two over model one as follows:

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(D|M_1)}{P(D|M_2)} \times \frac{P(M_1)}{P(M_2)}$$

The ratio of $P(D|M_1)$ to $P(D|M_2)$ is called the Bayes factor ([Raftery, 1995](#)), and is similar to the likelihood ratio test except that it considers the likelihood of the model over *all* possible parameter values rather than just the *most likely* combination of parameter values, and so corrects for the extra flexibility of more complex models in a more intuitive manner. The ratio of $P(M_1)$ to $P(M_2)$ is the prior odds, and is often equal to 1, indicating no prior model preference, but can be weighted towards one or other model to reflect prior belief. Bayes factors can be considered as an 'automatic Occam's razor' ([Smith and Spiegelhalter,](#)

1980), but under certain circumstances can be approximated using BIC (Smith and Spiegelhalter, 1980; Kass and Raftery, 1995; Spiegelhalter et al., 2002) and sometimes AIC (Smith and Spiegelhalter, 1980). Indeed, Jefferys and Berger (1991) provide a rationale as to how Occam’s razor itself is a conclusion of this Bayesian reasoning, and give several examples of simple model comparisons within this framework. Bayes factors can also be used as the basis of hypothesis testing (Goodman, 2005; Louis, 2005). The major disadvantage of this approach is that the marginal likelihood of the data given the model ($P(D|M_i)$) is calculated by integration of the likelihood over the distribution of values for all parameters, which is often highly dimensional (Andrieu et al., 2001). The use of approximations such as BIC have an advantage in the regard that this complex integration is not required (Kass and Raftery, 1995), although BIC has been shown to provide a poor approximation to Bayes factors under certain circumstances (Stone, 1979; Berger et al., 2003). DIC has been used to compare the fit of the zero-truncated negative binomial and negative binomial distributions (Kuhnert et al., 2005), but has also been shown to be unreliable compared to Bayes factors for comparison of the fits of zero-inflated gamma-Poisson and lognormal-Poisson distributions (Millar, 2009). In addition, the statistical methodology underlying DIC is not well understood compared to that of AIC and BIC, and is a topic for ongoing discussion (Alston et al., 2005).

A major problem inherent with likelihood-based model selection techniques is that, even when used appropriately, they provide only a probability that a certain model is to be preferred rather than the certainty that it is the correct model. Reporting results from only the ‘best fit’ model will therefore increase both type I and type II errors associated with periodic selection of the ‘incorrect’ model. Model averaging techniques (Hoeting et al., 1999; Wasserman, 2000) attempt to overcome this by incorporating information from several models, although logically it seems that inference obtained in this way must always be inaccurate as all but one of the candidate models must be incorrect. A simpler and possibly safer approach is to report results from all possible models along with their associated fit statistic. Reversible jump MCMC methods allow models with an uncertain number of parameters to be fitted (Green, 1995), and can therefore be used for simultaneous estimation of several models using a single algorithm (Black and Craig, 2002; Alston et al., 2005). However, the complexity and substantial computational burden involved with the sampling algorithm is a major disadvantage of this technique.

An alternative approach to model selection is to prefer a particular model formulation based on biological plausibility or prior experience with datasets of a similar nature. This is conceptually equivalent to adjusting the ratio of $P(M_1)$ to $P(M_2)$ in the prior odds given above, and allows a preferred model to be favoured with a given prior probability before combining with the Bayes factor to obtain the posterior belief in the model. Given sufficient evidence in the data the non-preferred model would still be indicated as more likely using this technique, although the model with the greatest prior probability will be chosen given sparse evidence in the data. As this approach is, with the exception of the prior odds, no different

to the use of a Bayes factor, the computational disadvantages are also the same. However, this reasoning can be extended to the extreme situation where the prior evidence for one model is so strong based on biological reasoning or experience of similar datasets that the prior odds are sufficient to overwhelm any reasonable Bayes factor, so that calculation of the Bayes factor is unnecessary. Therefore, model selection based on biological plausibility alone is justifiable as long as a sufficiently strong prior belief is reasonable. Selection of a single state model in preference to a zero-inflated model on the basis of theoretical justification has been advocated by [Lord et al. \(2007\)](#), and [Miaou and Lord \(2003\)](#) warn against selection of statistical distributions purely on the basis of empirical fit. [Warton \(2005\)](#) used AIC to compare the fit of the zero-inflated negative binomial and negative binomial distributions, but also concluded that a theoretical justification for a dual state model over a single state model is essential. Model selection on biological plausibility alone can be tempered by use of model validation techniques to verify that the data are sufficiently explained by the model. One approach to model validation is to compare the observed data to potential realisations of the data on the basis of the fitted model, to check that the observed data are not extreme given the model ([Meng, 1994](#); [Gelman et al., 1996](#); [Gibson et al., 2006](#)). This helps to ensure that the model sufficiently describes the observed data without the need to solve high-dimensional integrals, and represents an elegant method of embedding classical statistical methods within a Bayesian analysis. [Streftaris and Gibson \(2004\)](#) have also used similar methods to quantify the length of infection chains in the analysis of foot and mouth disease epidemics.

The approaches to model selection discussed above all work under the assumption that one of the candidate models is correct, with the others being incorrect. While in some applications this is valid, it is unlikely to be the case in fitting parasite distribution datasets where multiple potential sources of heterogeneity between hosts and parasites are likely to result in complex processes underlying the observed counts ([Anderson and May, 1982](#)). In this situation, it may be more useful to knowingly select a simpler model on the basis of tractability or improving the interpretation of the inference. This is in line with the observation that ‘all models are wrong but some models are useful’, attributed to [Box \(1976\)](#), among others. In the nested case, a distribution that allows greater freedom in the model may also be preferred to a more confined distribution on the basis that it is in some ways more conservative, i.e. is likely to provide larger confidence intervals for uncertain parameters. [Brame and Groer \(2002\)](#) preferred the use of a negative binomial over a Poisson distribution for this reason. This further complicates the use of likelihood based model selection techniques, as the ‘best fitting’ model may not be the most useful if another model provides better insight into the simplified system or produces more conservative parameter estimates, lending further credence to model selection on biological plausibility.

1.2.3 Statistical techniques for fitting distributions to data

When analysing parasitological data, the most common practise for determining the mean FEC of an observed dataset is based on calculation of the empirical arithmetic or geometric mean of the data. This is also recommended by [Coles et al. \(1992\)](#) to calculate the mean Faecal Egg Count Reduction Test (FECRT) reduction in an observed dataset. Such a measure provides an asymptotic estimate of the true mean of the population, but the possible discrepancy between the true population mean and observed sample mean is often not considered. The use of inferential statistics to determine the true characteristics of a population given a representative sample from that population provides a method of quantifying the uncertainty about true population parameters, and is central to most applications of statistics to biological problems. Parametric statistics such as the t-test are invalid when applied to right-skewed FEC data ([Wilson and Grenfell, 1997](#)). In order to overcome this, log transforming FEC data to improve normality is usual ([Wilson and Grenfell, 1997](#)), often after adding one to all observed counts to remove zero counts that cannot be log transformed ([Gregory and Woolhouse, 1993](#); [Fulford, 1994](#)). However, this often fails to normalise FEC data with a small mean ([Wilson and Grenfell, 1997](#)) as the true sample means described by an approximately lognormal distribution, which would be normal if log transformed, are overlaid by the Poisson distribution which has a larger relative effect on the combined distribution when the mean is small. This Poisson distribution is also the source of the zero counts which cause problems when log transforming the data, and would not be observed from the strictly positive lognormal or gamma distributions.

In order to correctly analyse FEC, somewhat more sophisticated statistical tools are therefore required. One such alternative is provided by likelihood-based techniques, which use a likelihood function to calculate the probability of obtaining the observed data from a given model with any combination of parameters. Generalised linear models (GLM) are likelihood-based generalisations of parametric statistical techniques such as the t-test that allow the error structure of the data to be separated from a link function which associates the response variable (such as FEC) with the explanatory model ([Wilson and Grenfell, 1997](#)). The likelihood of obtaining the response variable from the distribution describing the error term is maximised by altering the parameters of the link function. The error distribution can be one of a series of distributions, including the Poisson or negative binomial distributions ([Wilson and Grenfell, 1997](#)), and allow the zero counts frequently observed with FEC data to be modelled more correctly than simply adding an arbitrary constant to all the data. Generalised linear models have been used extensively in parasitology ([McCullagh and Nelder, 1989](#)), and have been shown to produce superior results to linear regression on transformed data ([Wilson et al., 1996](#); [Wilson and Grenfell, 1997](#)). The use of a GLM with a binomial error structure and logit link, known as logistic regression, is also useful in analysis of risk factors for a binary outcome, and has been used to analyse the risk factors associated with cyathostomiasis in

the horse (Reid et al., 1995).

While GLM are more flexible than parametric statistics, the specification of link function and error term is still quite limited. Alternative likelihood-based techniques overcome this limitation by considering the likelihood of the full model rather than just that of the error term: they not only describe the error structure as parametric distributions, but also use distributions to describe the process underlying the data. Likelihoods can therefore be calculated for virtually any conceivable model structure, and allow any reasonably well behaved linear or non-linear model to be fitted using whatever error structure is appropriate (Williams and Dye, 1994). These more complex likelihood-based techniques are therefore more sophisticated than simple GLM (Wilson and Grenfell, 1997) and have been shown to produce superior estimates when compared to least squares (Williams and Dye, 1994). The simplest application of likelihood principles is the use of an optimisation algorithm to find the combination of parameters that maximises the likelihood of observing the data. This is termed the maximum likelihood, and represents the combination of all model parameters that gives the greatest probability of obtaining the observed data from the model. The maximum likelihood can also be calculated using conditional, marginal or profile likelihoods, which seek to reduce the impact of ‘nuisance’ parameters by writing the likelihood as a function of only the parameters of interest (McCullagh and Nelder, 1989). Such maximum likelihood techniques are widely accepted in parasitology (Bliss and Fisher, 1953; Wilson et al., 1996; Torgerson et al., 2005), and can be easily achieved using proprietary software, as well as with models written in statistical programming languages such as R (R Development Core Team, 2009). Although maximum likelihood methods have mainly desirable properties (Edwards, 1992), such methods usually also carry the asymptotic assumption that the distribution of error for each parameter is normally distributed around the maximum likelihood estimate for that parameter. This may not be valid for parameters that are difficult to estimate, such as the dispersion parameter of the negative binomial for datasets typical in parasitology (van de Ven, 1993; van de Ven and Weber, 1999). For these types of dataset, the number of data points required for the asymptotic assumption of normality to hold may be far in excess of the data available (Stear et al., 2004). The results of a simulation study suggested that maximum or conditional likelihood produced the best parameter estimates when used with the negative binomial distribution (van de Ven, 1993).

It is also possible to use likelihood-based methods to examine the full distribution of likelihoods given different parameter combinations, also known as the likelihood curve. Likelihood profiling is a computationally intensive method of examining the relationship between parameter values and the likelihood of obtaining the observed data, which calculates the likelihood at every possible combination of parameter values. This method is feasible with models containing relatively few parameters, but not with high dimensional models. One alternative to likelihood profiling is to use Markov chain Monte Carlo (MCMC) to derive a distribution describing the uncertainty in each parameter estimate. Although MCMC is applicable using

frequentist principles, the technique is more often implemented within a Bayesian framework as this allows calculation of the complex integral representing the probability of the data to be omitted. MCMC works on the principle of Monte Carlo integration, which allows an undefinable distribution to be evaluated by drawing samples from the likelihood or posterior distribution (Gilks et al., 1998), approximating the population distribution using a sample distribution. The common method of drawing these samples is using a Markov chain (Gilks et al., 1998). Using this method, the value of the next sample to be drawn in a chain will depend only on the value of the sample drawn immediately before it, so that the chain of samples will after a period of time ‘forget’ its starting position. This period is known as the ‘burn-in’ period. The Markov chain will then converge on a stationary distribution that is independent of the starting values used, from which samples representing the desired distribution can be drawn (Gilks et al., 1998). The major disadvantage of a Markov chain is that convergence on this stationary distribution may be slow, resulting in a large amount of computational effort being expended before the chains can be sampled from. The number of samples required before a chain will converge is affected by the auto-correlation dependence, which is a measure of how ‘quickly’ the chain will move between extreme values (Gilks et al., 1998). If the ‘proposal distribution’ of possible values for a parameter at the next iteration is heavily dependent on the value at the previous iteration, then the chain will move slowly between extreme values. If the proposal distribution is relatively unaffected by the current value of the parameter then the chain will move more quickly between extreme values, and is said to be well mixed. Auto-correlation dependence also affects the Monte-Carlo error associated with a sample of values, in that a fixed number of iterations will better approximate to the true distribution when auto-correlation dependence is low (Gilks et al., 1998). Other techniques such as random sampling or Latin Hypercube sampling are generally more efficient, but cannot be likelihood driven. There are several different possible algorithms for generating a Markov chain (reviewed by Tierney (1994)), but the most commonly used are importance sampling (Ripley, 1987; Geweke, 1989), the Metropolis-Hastings method (Metropolis et al., 1953; Hastings, 1970) and the Gibbs sampler (Geman and Geman, 1990). Each of these different mechanisms are commonly associated with different problems, and are useful for different applications (Tierney, 1994). The Metropolis-Hastings algorithm is very flexible in it can be used with any function for which a likelihood can be calculated, but it can be inefficient as a result of the acceptance/rejection step of the algorithm. The Gibbs sampler is less flexible as it requires the prior distribution for each parameter to be defined as a parametric distribution, but is more efficient as no acceptance/rejection step is required. Further details of these sampling algorithms are given in Gilks et al. (1998). Each of these algorithms could be implemented for individual models using any appropriate computer programming language. However, software to implement these techniques on user-specified models is available in the form of the Bayesian Inference Using Gibbs Sampling (BUGS) project (Lunn et al., 2000; Thomas et al., 2006) and the similar Just Another Gibbs Sampler (JAGS) software (Plummer, 2008).

The major disadvantages of MCMC analysis include the relative computational complexity associated with the technique, as well as the need to ensure that the Markov chain has converged on the stationary distribution before sampling [Gilks et al. \(1998\)](#). A failure of the sampling algorithm to adequately converge on the stationary distribution will distort the posterior values for all parameters, and therefore bias the results ([Toft et al., 2007](#)). One of the most convenient ways of checking for convergence is to sample two or more parallel Markov chains with differing initial conditions, and use the Gelman-Rubin statistic to assess convergence between the chains. As the two chains are given the same model, and the stationary distribution of the two chains should be unaffected by the starting position, both chains will be sampling from the same distribution if converged. The Gelman-Rubin statistic compares the variance between chains to the variance within the chains to ensure that this is the case ([Gelman and Rubin, 1992a](#); [Brooks and Roberts, 1998](#)). A potential scale reduction factor of close to one is demonstrative of adequate convergence ([Gelman and Rubin, 1992a](#)), although the exact figure is somewhat arbitrary. It has been recommended to use starting values that are over-dispersed with respect to the stationary distribution to ensure the parameter space is fully explored before convergence ([Gelman and Rubin, 1992a,b](#)). It is also important to ensure that sufficient samples are drawn from the stationary distribution to reduce inaccuracies associated with Monte Carlo error [Gilks et al. \(1998\)](#). This can be done by comparing the parameter estimates between chains and drawing more samples if the estimates are not sufficiently similar, but can also be done using calculations on pilot chains using the Raftery and Lewis diagnostic as described by [Raftery and Lewis \(1995\)](#). Diagnostic statistics such as the Gelman-Rubin statistic and Raftery and Lewis diagnostic are not fool-proof, and visual assessment of trace plots to compare the stationary distribution of the two chains is always advisable. However, for large scale simulated data analysis where manual checking of each analysis would be laborious, automated functions such as the ‘runjags’ package are available ([Denwood, 2008](#), freely available to download from CRAN at <http://cran.r-project.org/web/packages/runjags/>). The ‘autorun.jags’ function of this package allows automatic calculation of chain convergence using the Gelman-Rubin statistic, and will extend a model in increments of 10,000 iterations, or a given user-specified value, until the Gelman-Rubin statistic is sufficiently reduced or the specified time limit is exceeded. Run-length diagnostics using the Raftery and Lewis’s diagnostic are then used to calculate the required number of sampling iterations, and the model extended further if required. This function, along with the other functions in the runjags package, can be used to implement any user-specified model in JAGS and return the Markov chains as an object within R.

One of the requirements of Bayesian analysis is that a distribution describing prior belief about each parameter must be chosen. This prior represents our knowledge about a parameter before data are collected, and is updated according to the data using the likelihood to obtain a posterior distribution for the parameter ([Goodman, 2005](#)). The use of prior distributions is somewhat controversial, and the Bayesian approach is often criticised for a perceived lack of objectivity. However, results obtained using frequentist methods are also subjective

in that the inference depends on the perceived experimental design (Berger and Berry, 1988). The advantage of the Bayesian approach is that the subjectivity is constrained to the choice of priors, rather than the frequently more opaque subjectivity issues associated with some frequentist methods. Techniques for attempting to make the Bayesian technique objective include robust Bayesian analysis (Berger et al., 1994), model averaging using different priors (Wasserman, 2000), establishment of ‘reference priors’ that are thought to be minimally informative, and the attempt to define objective ‘Jeffreys’ priors that contain no information for parameter combinations where the likelihood is informative (Jeffreys, 1961). These techniques are mostly quite technically complex and are difficult to implement in software such as BUGS and JAGS, and have been criticised by subjective Bayesians as unnecessary. The selection of priors is inherently subjective, similarly to many other statistical assumptions, and could be considered as a useful opportunity to introduce information into the model based on known limits of the biological system. In general, the effect of prior distribution selection is important when dealing with limited datasets, but has little effect with larger sample sizes (Raftery, 1995).

Each of the statistical techniques outlined above depends on the use of a parametric distribution to describe the data, or at least the error structure. The technique of non-parametric bootstrapping involves re-sampling and summarising the observed data to demonstrate empirically the distribution of the parameters of interest (Efron, 1979). As such it does not require knowledge of the distribution or processes generating the data (Mooney and Duval, 1993). Non-parametric bootstrapping approaches are therefore useful when the distribution of data is unknown. The assumption underlying this approach is that the data obtained are completely representative of the full distribution describing the population. This assumption is potentially violated when dealing with small sample sizes leading to misleading results. Computationally intensive parametric methods relax this assumption by fitting the observed data to a distribution, so that not all possible values have to be observed in order to be sampled. It is also possible to perform parametric (or semi-parametric) bootstrapping, although the advantages of distributional independence are then lost. A non-parametric bootstrapping approach has recently been suggested as an appropriate method to generate confidence limits from equine FECRT data (Vidyashankar et al., 2007), although the small sample sizes typically involved in equine FECRT data may be of concern when using this technique.

Although conceptually quite separate from inferential modelling, some discussion of simulation modelling in parasitology is warranted here. Mathematical simulation models have been used to develop the understanding of the host-parasite relationship (Boag and Thomas, 1975; May and Anderson, 1983; Kao et al., 2000), to analyse the aggregation of parasites among hosts (Woolhouse, 1992) and to examine the problem of anthelmintic resistance development (Gettinby et al., 1989; Gettinby, 1989; Barnes and Dobson, 1990; Smith, 1990). Such models can be either specific if formulated to model a single system very closely, or generic if designed to simplify a system using biologically meaningful parameters (Smith and Grenfell, 1994).

Simulation models are extremely useful for determining the effects of changing parameters such as parasite control measures and management approaches on known systems, but are less useful for estimating parameter values.

1.3 Advantages of statistical modelling as an inferential tool

The use of inferential statistics as outlined above represents an improvement on the calculation of summary statistics, as it allows inference to be made at the population level rather than merely summarising the sample. As such, more useful information is obtained from the datasets available, which allows a greater understanding of the distribution of parasites between hosts to be developed. Such techniques could also be used to underpin prospective sample size calculations to ensure that future studies involve sufficient data collection to provide the information required (Cohen, 1988). Some more robust statistical analysis of FEC data (Wilson et al., 1996) and FECRT data (Torgerson et al., 2005; Vidyashankar et al., 2007) has been advocated, although the current WAAVP guidelines for evaluating the efficacy of equine anthelmintics suggest using both parametric and non-parametric statistical techniques in an effort to corroborate each other (Duncan et al., 2002). Fitting observed data to a distribution also allows inference to be made on other characteristics of the data, such as variability between animals and possible multi-modality in the resultant data.

Statistical modelling approaches can also be used to quantify the repeatability of tests for quantifying parasite burdens, and for examining the complex variability patterns of parasitism within and between a group of animals. The most commonly used technique for quantifying the parasite burden of an individual or group of animals is the McMasters method (Whitlock and Gordon, 1939), the method of which is given in MAFF (1986). The technique is based on the floatation dilution principle described by Stoll (1930), and involves counting the number of eggs in a fixed volume of faecal suspension in saturated saline solution. However, quantifying the FEC output of an animal does not give a true indication of worm burden, as it reflects only the number of adult females (Sinniah, 1982), and does not take into account possible effects of density dependence or worm length on worm fecundity (Stear and Bishop, 1999), or the possible effect of worm senescence on fecundity. As a result, the relationship between worm count and FEC is poor (Smeal et al., 1977; Uhlinger, 1993).

These techniques can be used for identifying animals that are contributing to pasture contamination, and are therefore useful for controlling the underlying source of a group infection. The process is subject to variability introduced by the process of counting eggs in a sample (Sinniah, 1982), along with any extra variability introduced by inadequate mixing of the sample, or laboratory error in weighing faeces, or measuring solutions. However, FEC have generally been found to have a reasonable heritability (Doeschl-Wilson et al., 2008) and

have been used to estimate heritable traits of host resistance (Bisset et al., 1992; Vanimisetti et al., 2004). The precision of the FEC measurement is known to be related to the number of eggs counted (Hunter and Quenouille, 1952), and therefore the mean FEC and minimum egg detection threshold of the technique used. Attempts have been made to quantify the observed variability between FEC using the McMasters technique (Levine et al., 1960; Dunn and Keymer, 1986; Rossanigo and Gruner, 1991), to derive the optimal number of samples to take (Hunter and Quenouille, 1952; Gasbarre et al., 1996), and to calculate the power of a FECRT (Gill et al., 1986). However, none of these studies have used a modelling approach to attempt to partition the variance.

The seasonal variation in parasite burden and mean FEC has been well described in ruminants (for example Parnell, 1962; Reid and Armour, 1972; Eysker and Vanmeurs, 1982; Smeal et al., 1977), and has also been reported in horses (Ogbourne, 1971; Herd, 1986; Langrová, 1998; Baudena et al., 2000). However less is known about the variation in true mean FEC over a shorter period of time. It has been suggested that observed FEC vary over the course of a day due to factors related to feed intake and gut mobility of the host (Uhlinger, 1993), although Bennett (1990) found no evidence for the diurnal variation in FEC of horses and Keymer and Hiorns (1986) found no measurable bias in FEC in a study in mice, suggesting that observed variability is truly random. Attempts have also been made to evaluate the use of pooled FEC samples (Eysker et al., 2008), although again without much consideration to the statistical processes underlying the procedure. Alternative methods based on similar principles to the McMasters technique have also been proposed (Mes, 2003; Presland et al., 2005; Cringoli, 2006), and in some cases the observed variability using the newer technique has been shown to be less than that obtained using the McMasters technique (Mes, 2003; Presland et al., 2005). However, none of the studies have identified the true source of the observed variability, and some of the statistical conclusions made by Mes (2003) have been questioned (Morrison, 2004). Use of more rigorous statistical approaches to partition the observed variability would therefore be of benefit. In addition, by improving the understanding of the statistical processes underlying the test, it may be possible to make recommendations that would improve the repeatability of the test.

1.4 Study objectives

The overall scope of this thesis was to use a quantitative approach to improve the analysis of faecal worm egg count data. More specific objectives were to investigate statistical issues pertaining to the analysis of FEC and FECRT data, in order to improve analysis techniques for existing data, and then to develop sample size calculation techniques with which to design future studies. In order to facilitate this, novel methods for analysis of FEC data were first developed and validated using simulated data before being applied to biological datasets.

The progression of this thesis towards achieving these goals is outlined below.

Analysis of parasitological datasets is usually mainly concerned with providing inference on the true mean FEC, although the variability of the distribution is often also quantified using one of several possible measures. The biological relevance of each of these measures is discussed, and the properties of the chosen measure demonstrated. Asymptotic assumptions are likely to be violated with FEC data, therefore methods of analysis using Bayesian MCMC were explored. Several different syntactic variations of MCMC models representing different biologically plausible distributions were compared using simulated datasets to find the parametrisation that produced the most appropriate inference for each distribution. The inference made, particularly from sparse datasets, was substantially affected by the choice of MCMC model parametrisation.

Issues pertaining to the choice of statistical technique and model selection were then explored using simulated datasets to ascertain the consequences of making incorrect asymptotic and distributional assumptions. The usefulness of model selection based on empirical fit was also examined using similar methods. As the zero-inflated and over-dispersed distributions discussed are variously nested around the Poisson distribution, the inference made using under- and over-specified models was compared, and the potential consequences where the true distribution is unknown are discussed. The validated MCMC methods were then applied to FEC data obtained from horses and sheep to examine the parameter distributions and distributional fit in field data, and the inferences compared to those made using the existing methods of analysing FEC data examined in the simulation studies. The FEC data used were from parasites with completely distinct fundamental biology as well as different host species. The inference made from each type of data, including the difference in empirical fit to contrasting frequency distributions, is discussed.

A further set of models to address the specific issue of the FECRT were then developed and compared to existing methods using simulation studies, before being applied to biological datasets. The FECRT data were from diverse sources and relate specifically to screening for anthelmintic resistance within cyathostomin populations in horses. Statistical issues pertaining to appropriate methods of FECRT data collection and analysis were then discussed, and potential biases of the currently most widely used procedure elucidated.

Finally, a multi-level MCMC model was used to analyse FEC data obtained by repeat sampling of horses so that the effect of sample variability between observed FEC could be quantified, and recommendations to improve the repeatability of the McMaster's technique are made. Using this information, methods of performing prospective precision analysis for FEC sampling, and power analysis calculations for hypothesis driven FECRT procedures, were developed and parametrised using the above inference on the variability structure of equine FEC data.

CHAPTER 2

**Selection, development and evaluation of a
population model**

Selection, development and evaluation of a population model

2.1 Introduction

Population modelling is a useful way of analysing the distribution of parasites (Elliott, 1977; Nødtvedt et al., 2002; Morgan et al., 2005), and has been shown to give more useful information than the traditional method of simply calculating the group arithmetic or geometric mean (Torgerson et al., 2005). The advantages are that more information can be extracted from the data by defining each count as part of a group distribution, and by treating each FEC as a sample from a Poisson distribution, it is also possible to account for the known counting errors associated with the McMaster technique (Döpfer et al., 2004). The main disadvantage of this approach is that it requires the use of a parametric distribution to describe the process underlying the data. This may produce misleading results if the chosen distribution is inappropriate.

The distribution of parasites between hosts can be quantified by simply counting the number of parasites on/within hosts, and recording the number found on/in each host. For internal parasites such as nematodes, a proxy for worm counts that does not require a post-mortem examination of the hosts such as FEC may be preferable. Both measures are count observations, and therefore might be expected to fit a Poisson distribution if parasite distributions between hosts were random (Elliott, 1977), i.e. all animals had the same expected number of parasites. However, almost all studies of such distributions have found them to be highly aggregated (Crofton, 1962; May and Anderson, 1978; Shaw et al., 1998), so that the true distribution of parasites is over-dispersed with respect to the Poisson. This implies that the observed variation is more than would be expected from a truly random process with uniform mean, and necessitates the use of a further distribution to describe the variation in true mean number of parasites between hosts. The distribution most widely used to describe this variation in true mean is the gamma distribution, usually formulated as the negative binomial (see Section 1.2.1), although the lognormal distribution may be more appropriate. Extension of these gamma-Poisson and lognormal-Poisson distributions to the equivalent zero-inflated gamma Poisson (ZIGP) and zero-inflated lognormal Poisson (ZILP) mixture models would

also be useful if the data are considered to be bi-modal. This allows the observed data to be derived from one of two groups; either the ‘infected’ continuous distribution or a group with a true mean of zero which are truly uninfected (Martin et al., 2005). Alternatively, the distribution could be used to approximate a bi-modal distribution with one high mode and one low mode if the majority of counts from the latter are expected to be zero.

Comparison of the inference made using each distributional assumption first requires that the optimal parametrisation for each of the models be defined. However, no such pre-defined models are readily available for use with software capable of implementing an analysis using MCMC, such as WinBUGS (Lunn et al., 2000) and Just Another Gibbs Sampler (JAGS) (Plummer, 2008). The overall aim of the work described here was to compare the accuracy of several variants on each parameterisation using simulated data. The specific objective was to identify the formulations with the best defined properties, for use in later chapters of this thesis.

2.2 Defining parasite distributions

2.2.1 Introduction

There are three factors that are generally of interest when examining the distribution of parasites between and within hosts. The most obvious is the mean, which can either be the mean egg count or mean parasite count. This represents the expected number of parasites or parasite eggs in an animal, and is related to the total number of parasites or eggs by the number of animals in the group. It is therefore arguably a more useful measure of the mid-point of the distribution than the median or mode, neither of which are frequently used in the parasitology literature. The mean can either be thought of as the ‘population mean’, defined here as the mean of the distribution from which all animals in the infected group are derived, or as the ‘true sample mean’, defined here as the average of the true mean egg shedding rate, or parasite burden, of the sampled animals. The two measures are identical when all animals from the population are sampled, but when only a sample is taken from the population the two are distinct, as the former describes a population from which we only have a sample, and the latter describes the observed mean of the sample of animals. The uncertainty associated with estimating the population mean will always be greater than that for the true sample mean, as the population mean also includes animals which were not sampled. For the majority of the work here the mean referred to is the population mean; the concept of the true sample mean will be revisited in [Chapter 6](#).

The second variable of interest is the variability, or dispersion in parasite burden or parasite egg shedding rate. This can be subdivided into the known Poisson variability due to the

counting process, variability between animals, the variability within individual animals over short periods of time, and the aggregation of eggs in a single faecal pile. Quantifying the amount of the total variability that is attributable to each of these factors is likely to be very difficult unless large numbers of repeat samples are taken, and is impossible if only one sample is taken per animal. Consequently, the variability is usually considered as a single value describing all the sources of variability between samples, excluding that due to the counting process.

When using zero-inflated distributions, the proportion of the population that produce zero counts in excess of that expected from the distribution used to describe the majority of the counts is also of interest. One interpretation of this value is as one minus the prevalence of infection in the population. It is possible for the prevalence to be 100%, so that there are no animals with a true mean of zero, in which case the zero-inflated distribution would collapse to the simpler distribution describing the infected animals given sufficient evidence for this in the data. In this situation, results obtained using a zero-inflated distribution and the non zero-inflated equivalent would be equivalent.

While the mean and zero-inflation of a distribution are more clearly defined, the variability can be measured in different ways. The most obvious is to examine the absolute standard deviation or variance between samples, but it makes more sense to use a measure that is standardised in terms of the mean of the distribution because the absolute variance of a sample is likely to be heavily affected by the mean. One possibility is to use the *variance : mean* ratio, also known as the *coefficient of dispersion*, as a measure of the variability. This has the advantage of describing the amount of the total variability between observed counts that can be attributed to the variability of the counting process. The variance of a Poisson process is equal to its mean; therefore if the variance between counts is greater than the mean then the distribution must be over-dispersed with respect to the Poisson. In practise, this measure is not particularly helpful as the vast majority of parasite distributions are already known to be over-dispersed. It is also possible to alter this measure slightly so that it reflects the ratio of the extra-Poisson variance, which is equal to the variance between samples, to the mean. In this case, any value close to zero represents a distribution with little over-dispersion. If using the gamma distribution to describe the variability between samples, the extra-Poisson *variance : mean* ratio can be derived as:

$$\frac{\text{total variance} - \text{Poisson variance}}{\text{mean}} = \frac{(\lambda + \alpha \times \beta^2) - \lambda}{\alpha \times \beta} = \beta$$

Where α and β are the shape and scale parameters of the gamma distribution respectively, and λ is the mean of the Poisson distribution. This measure is therefore convenient for the gamma distribution as it is exactly equivalent to the scale parameter, β .

The second possible measure of variability is to use the ratio of the extra-Poisson standard deviation to the mean, also known as the *coefficient of variation (cv)*. This measure is commonly used for data measured on a ratio scale, such as positive counts, although it has no meaning for data measured on an interval scale. Values of coefficient of variation (*cv*) of close to zero indicate a Poisson distribution, and the variability between samples increases as the *cv* increases towards Infinity. If using the gamma distribution, the *cv* would be equal to:

$$\begin{aligned}
 cv &= \frac{\text{standard deviation of gamma distribution}}{\text{mean}} \\
 &= \frac{\sqrt{(\text{total variance} - \text{Poisson variance})}}{\text{mean}} \\
 &= \sqrt{\left(\frac{\text{total variance} - \text{Poisson variance}}{\text{mean}^2}\right)} \\
 &= \sqrt{\frac{\lambda + \alpha \times \beta^2 - \lambda}{\alpha^2 \times \beta^2}} \\
 &= \sqrt{\frac{1}{\alpha}}
 \end{aligned}$$

The *cv* is therefore related to the shape parameter α , which is often used in the parasitology literature as the inverse measure of aggregation k .

If using the lognormal distribution with mean and variance on the log scale to describe the variability between samples, the *cv* is related to the standard deviation of the log scale distribution alone (Limpert et al., 2001). This can also be shown from the following derivation. The variance of the lognormal distribution is given by $(\exp(l\mu))^2 \times \exp(l\sigma^2) \times (\exp(l\sigma^2) - 1)$ and the mean by $\exp(l\mu) \times \exp\left(\frac{l\sigma^2}{2}\right)$, where $l\mu$ is the mean and $l\sigma$ the standard deviation of the distribution on the log scale (Evans et al., 2000). The ratio of the standard deviation to mean (*cv* on the exponent scale) is therefore given by:

$$\begin{aligned}
 cv &= \frac{\sqrt{(\exp(l\mu))^2 \times \exp(l\sigma^2) \times (\exp(l\sigma^2) - 1)}}{\exp(l\mu) \times \exp\left(\frac{l\sigma^2}{2}\right)} \\
 &= \frac{\exp(l\mu) \times \sqrt{\exp(l\sigma^2) \times (\exp(l\sigma^2) - 1)}}{\exp(l\mu) \times \exp\left(\frac{l\sigma^2}{2}\right)} \\
 &= \frac{\sqrt{\exp(l\sigma^2) \times (\exp(l\sigma^2) - 1)}}{\sqrt{\exp(l\sigma^2)}} \\
 &= \sqrt{\exp(l\sigma^2) - 1}
 \end{aligned} \tag{2.1}$$

In deciding which of these measures to use, it is important to consider what is being quantified. In this case, it is most relevant to measure the ‘difference’ in worm burden or egg shedding rate between samples, due to the differences in host susceptibility and worm survival and fecundity. Therefore, the measure of variability that would be most useful would remain constant when the larval challenge changes, but differences in animal susceptibility and worm survivability are constant. Which of the different measures of variability best fulfils these requirements is assessed using a rudimentary simulation model to describe, in very basic terms, the variability in worm burden between animals. This is based on each host animal being assigned a probability of removing a larva before development to an adult parasite, and each ingested larva being assigned a probability of developing to an adult. These probabilities are varied between hosts and larvae to simulate variability between animals and worms, and stochastic variability in the number of adult worms between animals given a fixed larval challenge is examined.

2.2.2 Materials and methods

A mathematical model representing the processes discussed was written in the R statistical programming language (R Development Core Team, 2009). The number of ingested larvae for each of 10,000 animals was fixed at the same value for each simulation. The variability in host immune defence was simulated using a beta distribution, so that each animal was given a distinct probability of eliminating each larva before development to an adult parasite. The variability in worm resistance to the host immune response was modelled similarly. The survival of each larva was then modelled using a Bernoulli trial using the combined probabilities derived from host immunity and worm resistance, and the resultant number of adult parasites in each host was calculated. The observed coefficient of variation and coefficient of dispersion were calculated as $\frac{\sigma}{\mu}$ and $\frac{\sigma^2}{\mu}$ respectively, where σ is the standard deviation of the number of adult worms per animal, and μ is the mean number of adult worms per animal. The model was run a total of three times, with the second and third simulations using altered values for the larval challenge and altered values for the variation in host immunity and worm resistance respectively.

2.2.3 Results

The effect of doubling the larval challenge while retaining the same values for host and worm susceptibility distributions, simulating the effect of moving the same group of animals onto a field with a different larval challenge, is shown in [Figure 2.1](#). The mean number of worms per adult animal is increased, but the distribution retains a similar shape when the x-axis scale is increased proportionately with the increase in larval challenge. In effect this can be thought

of as doubling the expected number of parasites in each animal within the population, without changing the proportion of worms in each individual animal. Conversely, altering the distributions describing the host and worm susceptibilities while retaining the same value for larval challenge alters the shape of the distribution, but not the mean (Figure 2.2). In this case the variation in the proportion of worms in each animal is changed, altering the shape of the distribution. In Figure 2.3, this effect is shown using an empirical cumulative distribution function, plotting the proportion of the total worm population within each animal. The cor-

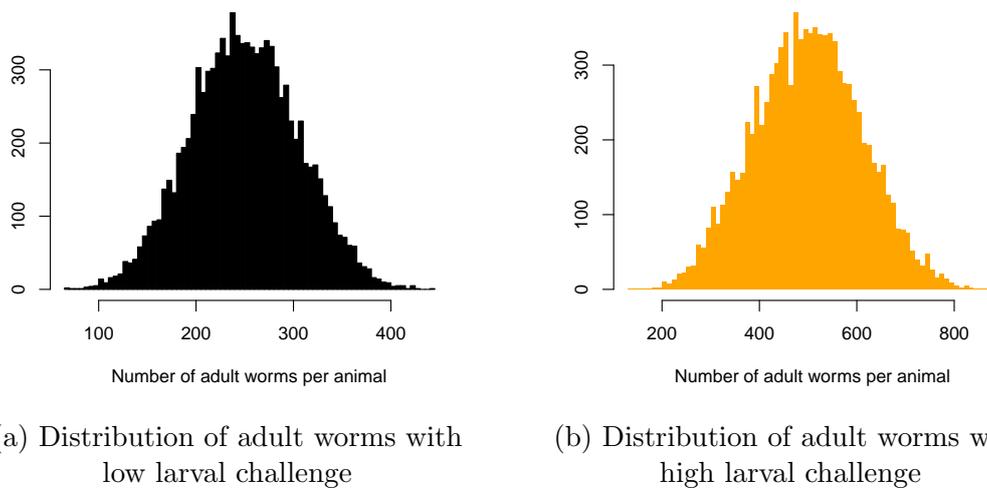


Figure 2.1: The effect of altering larval challenge on the distribution of adult worms in a mathematical model of the distribution of parasites between hosts - the x-axis scale is increased proportionately to the increase in larval challenge

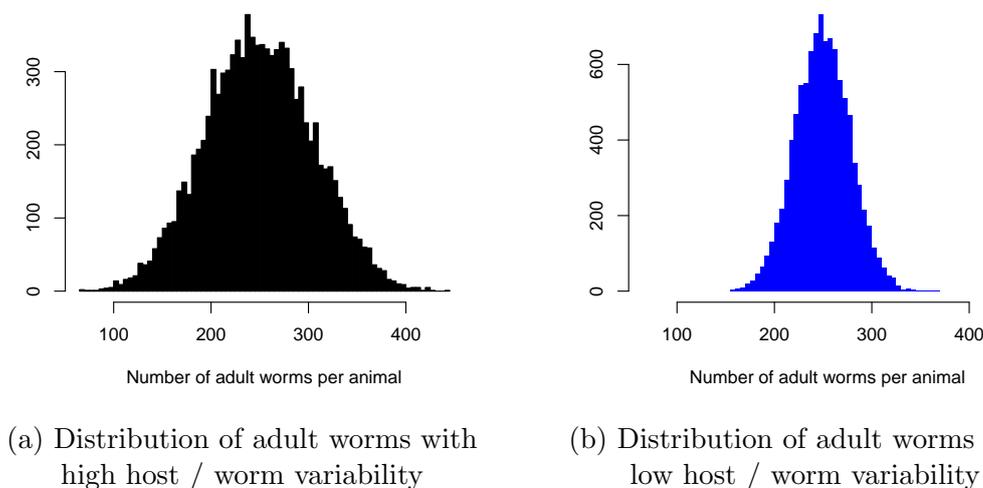


Figure 2.2: The effect of altering the distribution of host and worm variability on the distribution of adult worms in a mathematical model of the distribution of parasites between hosts - the x-axis scale is fixed as the larval challenge is unchanged

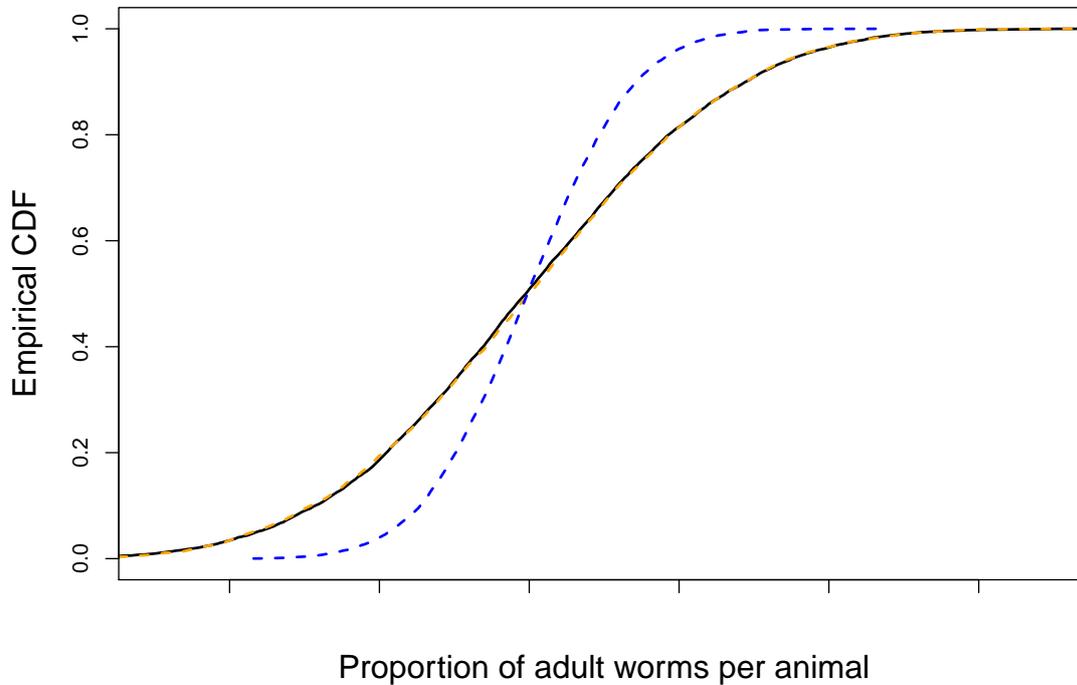


Figure 2.3: The effect on the proportion of adult worms in each individual host of altering different parameters in a mathematical model of the distribution of parasites between hosts - the effect of changing larval challenge (orange) and the distribution of host and worm susceptibility (blue) shown relative to the baseline parameters (black).

Table 2.1: The effect of altering either larval challenge (Dataset 2) or the distribution of host and worm susceptibility (Dataset 3) compared to the baseline (Dataset 1) on the distribution of worms in a mathematical model of the distribution of parasites between hosts

	Coefficient of variation	Coefficient of dispersion
Dataset 1	0.220	12.158
Dataset 2	0.220	24.285
Dataset 3	0.113	3.208

responding *coefficient of variation* and *coefficient of dispersion* for each of these distributions is shown in [Table 2.1](#).

2.2.4 Discussion

In order to assess changes in host and/or parasite susceptibility independently of variable larval challenge, a useful measure of variability would remain constant for the two distributions shown in [Figure 2.1](#) but change between the distributions shown in [Figure 2.2](#). Based on the results of this simulation exercise, it appears that the *cv* is a suitable marker to measure aggregation for the purposes of these studies. This is also consistent with the existing wide-spread use of the parameter ‘k’, which is an equivalent, although indirect, measure of aggregation. This measure is therefore used exclusively in this thesis.

2.3 Selecting a model definition for the zero-inflated gamma Poisson model

2.3.1 Introduction

There are several subtly different ways of specifying a Bayesian gamma Poisson model using the BUGS syntax. Although each represents an equivalent model, the syntactic choice of both potential sampling parameters and prior distributions may impact on the efficiency of the sampling algorithm which could reduce the speed of convergence ([Kass et al., 1998](#)). Several different model specifications were tested for convergence, reliability and accuracy by comparing the outputs of each model to the known parameters used to generate a simulated dataset.

2.3.2 Materials and methods

2.3.2.1 Model Specifications

A description of the processes in each variation of the ZIGP model is as follows. The Poisson part of each model specification is identical, as each data point must be defined as part of a Poisson process with a unique value for lambda. Similarly, the group to which each count belongs is determined in each variant by a Bernoulli trial, the output of which is either 0 (indicating the count is part of the true zero group) or 1 (indicating the count is part of the gamma Poisson distribution).

The gamma distribution can be defined in one of several ways, shown in [Models 2.1, 2.2 and 2.3](#) (see [Appendix A](#) for model code). For each of these, the prior distributions used allow sampled values of mean to be between 0.001 and 1000, and independently, sampled values for the cv to be between 0.01 and 10. Zero-inflation is sampled from between 0 and 100%. The parameters used to assess convergence are $mean$ (the mean of the gamma distribution), $\frac{1}{a}$ (the inverse of the shape parameter) and $1 - \frac{zero\ inflation}{100}$ (the proportion of the group that are defined by the gamma distribution) for all models. In [Model 2.1](#), the lambda value of each true Poisson process is gamma distributed with two parameters; a (the shape parameter) and b (the rate parameter or inverse scale parameter). The mean of this is therefore equal to $\frac{a}{b}$, and the cv to $\sqrt{\frac{1}{a}}$, which is calculated from a using the output chains. The prior is on $\frac{1}{a}$ rather than a because $\frac{1}{a}$ is closer to the value of interest (cv). Zero-inflation is calculated as $100(1 - p)$ after running the model. In [Model 2.2](#), the lambda value of each true Poisson process is defined by the multiplication of a single mean value with a dispersal parameter equal to $gamma$ for each count. This $gamma$ comes from a gamma distribution with a single unknown parameter; a and a mean of $\frac{a}{a} = 1$. Although the mean of the gamma distribution has been altered, only the scale parameter has been changed so the shape parameter remains unaltered. The cv is calculated from a , and zero-inflation from $prob$, as before. [Model 2.3](#) is a slight variation of [Model 2.2](#) in that the mean of the gamma distribution is defined as the mean of the data rather than being equal to 1. The prior is put directly on the mean as with [Model 2.2](#) as before, and b is derived from $\frac{a}{mean}$.

Although it has been stated that in each model variant zero-inflation is determined from a Bernoulli trial for each count, the probability parameter of the Bernoulli trials can be defined in one of two ways. In the [Models 2.1 to 2.3](#), the probability of each count being part of the true gamma Poisson distribution (ie. $lambda$ being multiplied by one) is defined as a single unknown parameter $prob$. It is also possible to allow each count to have a separate value for this probability, so that the group value for zero-inflation is taken from the mean of $prob$. [Model 2.4](#) incorporates this modification into [Model 2.3](#).

2.3.2.2 Data Simulation

A total of one thousand datasets were each produced with 10 and 50 observations using R. Each dataset was generated using a ZIGP model with parameters randomly selected from a $\log_{10}(U(-1, 2))$ distribution, $U(0, 0.75)$ distribution, and $\log_{10}(U(-2, 0.70))$ distribution for mean count, the proportion of the population that is uninfected, and the cv respectively (where U represents the Uniform distribution). Because of the greater difficulty inherent in analysing datasets with lower values for mean count, this parameter was taken from a log uniform distribution to maximise the number of datasets with lower mean counts. The cv was also taken from a log uniform distribution to increase the number of datasets that were close to a Poisson distribution.

Each dataset was then analysed separately using [Models 2.1 to 2.4](#) detailed above. In order to aid the analyses, a function was written by the author in the R programming language that allowed the analyses to be automated, with R calling JAGS to run each simulation as necessary. A burn-in period of 5,000 iterations was run before sampling for 10,000 iterations. Convergence was assessed by calculating the Gelman-Rubin statistic using two chains, with any models that failed to converge being re-run for 10,000 iteration intervals until convergence was achieved, up to a maximum allowed processor time of one hour per dataset. The required sample length was then calculated using the Raftery and Lewis's diagnostic, and the simulation extended if necessary. The lower and upper 95% highest posterior density credible interval estimates along with the median estimate for each model and each dataset were output to a text file for each model.

2.3.2.3 Statistical analyses

The final outputs for each model as well as the simulation parameters were then input into R for comparative analysis. Because some models were unable to return usable analyses for some datasets, because of poor convergence or a persistent error when running the model for that dataset, and in order to make sure that each model was evaluated equally, the results for any datasets that were not returned for any individual model were discounted for all models. To assess the accuracy of the median estimates, the relative root-mean-square-error (RMSE) was calculated using the simulated, or true, value for each parameter. The relative RMSE can also be thought of as the standard deviation of the ratio between each median estimate and the simulated values; although it should be noted that this is not equivalent to the accepted meaning of the term 'standard deviation'. The term 'relative RMSE' is used exclusively here to avoid confusion.

2.3.3 Results

The number of datasets that were successfully analysed by [Models 2.1 to 2.4](#) at each sample size are shown in [Table 2.2](#). [Model 2.1](#) returned the most successful analyses, with [Models 2.2, 2.3](#) and [2.4](#) returning similar numbers of errors and failures to converge within one hour for the larger sample size. [Model 2.4](#) returned comparatively few successful analysis results with a sample size of ten.

In [Table 2.3](#) the percentage of the simulated parameters that fell outside the 95% CI for each combination of model, parameter and sample size used, are shown. The percentage of simulated parameters that fell outside the 95% confidence interval for zero-inflation with [Model 2.4](#) was over 50% at both sample sizes. An even higher percentage were outside the 95% confidence interval for mean count and cv with [Model 2.1](#). [Models 2.2](#) and [2.3](#) produced less

2.3 SELECTING A MODEL DEFINITION FOR THE ZERO-INFLATED GAMMA POISSON MODEL

Table 2.2: Number of simulated datasets successfully analysed by four variants of a ZIGP model

	Model 2.1	Model 2.2	Model 2.3	Model 2.4
Total successfully analysed	891	674	716	415
Model returned an error	11	201	170	326
Failed to converge	0	27	16	161
Dataset all 0	98	98	98	98
(a) Sample size 10				
Total successfully analysed	936	781	763	808
Model returned an error	42	185	195	165
Failed to converge	17	29	37	22
Dataset all 0	5	5	5	5
(b) Sample size 50				

results outside the 95% confidence interval for mean count than the other models. However, both models seemed to consistently under-estimate the cv , especially at the smaller sample size.

The root-mean-square-error at each combination of model, parameter and sample size is shown in [Table 2.4](#). The relative RMSE for all models was comparatively large with the

Table 2.3: Proportion of true values above and below the 95% confidence intervals produced by four variants of a ZIGP model from the analysis of simulated data

	Model 2.1	Model 2.2	Model 2.3	Model 2.4
Above mean count 95% CI	55.9%	0.3%	0.6%	0.8%
Below mean count 95% CI	0%	0%	0%	76.6%
Above variability 95% CI	47.9%	22.9%	24%	19.8%
Below variability 95% CI	25.3%	0%	0%	1.1%
Above zero-inflation 95% CI	3.9%	0.3%	0.3%	55.1%
Below zero-inflation 95% CI	0%	1.4%	1.4%	1.9%
(a) Sample size 10				
Above mean count 95% CI	68.5%	1.3%	1.2%	13.1%
Below mean count 95% CI	0%	1.6%	1.6%	73.2%
Above variability 95% CI	44.2%	25.4%	25.3%	23.7%
Below variability 95% CI	23.8%	1.2%	1.3%	6.5%
Above zero-inflation 95% CI	5.2%	0.1%	0.1%	58.4%
Below zero-inflation 95% CI	0.1%	2.6%	2.3%	17%
(b) Sample size 50				

smaller sample size, although this improved for all except **Model 2.1** with the larger sample size.

2.3.4 Discussion

Of the types of errors examined, low confidence of 95% confidence intervals has the most potential to result in inaccurate conclusions to be drawn from analyses, as a tendency to produce inappropriately small confidence intervals will increase the Type I error rate associated with the method. The relative RMSE is also important as it reflects the accuracy of the ‘most likely’ estimate, although would not necessarily result in false conclusions being drawn. The proportion of datasets returned successfully may be related to the parameters of the distribution of the data, with for example datasets with low means and high variability resulting in more convergence failures. For this reason, only datasets that were successfully analysed by all models discussed were included in the calculation of confidence intervals and relative RMSE.

On the basis of the number of datasets successfully analysed, **Model 2.1** outperformed the others. This was largely due to a reduced proportion of the datasets causing the model to return an error associated with extreme values for the Poisson means. More of the datasets failed to converge within the required time span with the larger sample size for **Models 2.1** to **2.3**, reflecting the longer running time to achieve the same number of iterations associated with larger datasets. Interestingly, **Model 2.4** converged more frequently with the larger sample size, possibly indicating an identifiability issue with the smaller datasets. **Models 2.2** and **2.3** out-performed the other models on the basis of the 95% confidence intervals and relative RMSE for the mean and zero-inflation, although had a tendency to over-estimate

Table 2.4: Relative RMSE of the median estimate for each parameter produced by four variants of a ZIGP model from the analysis of simulated data

	Model 2.1	Model 2.2	Model 2.3	Model 2.4
Mean count	260.720	109.310	110.757	95.031
Variability	0.562	6.768	6.804	6.469
Zero-inflation	5.531	4.407	4.400	11.948
(a) Sample size 10				
Mean count	393.041	0.905	1.047	3.428
Variability	0.607	2.460	2.501	2.013
Zero-inflation	5.482	3.318	3.329	9.297
(b) Sample size 50				

2.4 COMPARING PRIOR DISTRIBUTIONS FOR THE GAMMA POISSON MODEL

the variability parameter. This is probably due to the choice of prior for the variance parameter compounded with the difficulty in estimating this parameter from the data tending the sampler towards larger values, which may also explain the greater proportion of datasets that returned an error with these models compared to [Model 2.1](#). [Model 2.1](#) failed to return appropriate 95% confidence intervals for the mean parameter, indicating that placing the prior on the mean and shape parameters produces better results than placing the prior on the shape and scale parameters and inferring the mean from these. This is likely to be because sampling the shape and scale parameters produces good estimates of these parameters independently, but not necessarily the dependent parameter that is the mean.

In spite of the comparatively poor results for the variability parameter, it is evident that placing the prior directly on the mean rather than on the scale parameter produces more reliable inference on the mean, and that a single value of zero-inflation produces a more identifiable model than using individual probabilities. The effect of altering the prior distribution on the variance parameter in the preferred [Models 2.2](#) and [2.3](#) is explored in [Section 2.4](#).

2.4 Comparing prior distributions for the gamma Poisson model

2.4.1 Introduction

The results obtained in [Section 2.3](#) indicate that [Models 2.2](#) and [2.3](#) produce the most accurate 95% confidence intervals when analysing simulated data, although the results from both models indicates a tendency to over-estimate the cv . This is most likely due to the choice of prior distribution for the shape parameter having an effect on the posterior distribution, given the lack of information on the true value of the cv in the data. As a result, the uniform prior placed on the inverse shape parameter in the model would tend towards a posterior distribution that was closer to being uniform than the log-uniform distribution from which the cv was simulated. This is because, on average, the prior would tend towards values that are higher than the simulation value. This effect might be reduced by using a prior distribution for the inverse shape parameter that is log-uniform. The effect of the prior for cv on the inference made using [Models 2.2](#) and [2.3](#) is analysed by comparing the results obtained using a uniform prior (presented in [Section 2.3](#)) to results obtained using a log-uniform prior for analysis of the same data.

2.4.2 Materials and methods

[Models 2.2a](#) and [2.3a](#) are an adaptation of models [Models 2.2](#) and [2.3](#), with the prior cv specified on the log scale. For each model, the the log of the inverse shape parameter is

2.4 COMPARING PRIOR DISTRIBUTIONS FOR THE GAMMA POISSON MODEL

sampled from a uniform scale with a lower limit of -9.21 ($= \log 0.0001$) and an upper limit of 4.6 ($= \log 100$). This provides prior distributions with the same limits as before, but with a different shape. [Models 2.2a](#) and [2.3a](#) are used to analyse the data discussed in [Section 2.3](#). The results of the analysis, along with the existing analyses produced using [Models 2.2](#) and [2.3](#), are presented as before. [Model 2.2a](#) is also used to analyse a dataset consisting of ten observations with no observed eggs.

2.4.3 Results

The number of datasets that were successfully analysed by [Models 2.2a](#) and [2.3a](#) at each sample size are shown in [Table 2.2](#), along with the results for [Models 2.2](#) and [2.3](#) from [Section 2.3](#) for comparison. [Model 2.2a](#) returned the most successful analysis results, defined by the model converging and not returning an error, and [Models 2.2](#) and [2.3](#) the least.

[Table 2.3](#) shows the percentage of the simulated parameters that fell outside the 95% CI for each combination of model, parameter and sample size used. The percentage of simulated parameters that fell outside the 95% confidence intervals was closer to the 5% assumed value for all parameters, except zero-inflation at sample size 10, for [Models 2.2a](#) and [2.3a](#) than [Models 2.2](#) and [2.3](#). The relative RMSE for the cv and mean count at the smaller sample size is reduced for [Models 2.2a](#) and [2.3a](#) compared to [Models 2.2](#) and [2.3](#) ([Table 2.7](#)). relative RMSE for zero-inflation and mean count with the larger sample size are comparable between all models, although slightly higher for [Models 2.2a](#) and [2.3a](#).

The analysis of the zero count dataset produced very large confidence intervals reflecting the prior distribution for the mean count and zero-inflation parameters, although the posterior distribution for the variability parameter tended towards higher values of cv compared to the prior. The relationship between the mean count and shape parameter estimates at each iteration are shown in [Figure 2.4](#).

2.4.4 Discussion

The prior chosen for the inverse shape parameter appears to have a dramatic effect on the inference for cv , indicating that the prior is more informative than the data for this parameter. This could be due to the prior being relatively informative, but is more likely to be because the likelihood curve is relatively uninformative due to the parameter being difficult to estimate from data. The prior for the inverse shape parameter also appears to affect the inference for the other parameters, although to a lesser extent. Even when the prior for cv is on the same scale as the sampled simulation parameters, the 95% confidence intervals for this parameter still give a lower true confidence than expected. A prior distribution of very

Table 2.5: Number of simulated datasets successfully analysed by four variants of a ZILP model

	Model 2.2	Model 2.3	Model 2.2a	Model 2.3a
Total successfully analysed	674	716	801	739
Model returned an error	201	170	101	160
Failed to converge	27	16	0	3
Dataset all 0	98	98	98	98
(a) Sample size 10				
Total successfully analysed	781	763	870	830
Model returned an error	185	195	119	112
Failed to converge	29	37	6	53
Dataset all 0	5	5	5	5
(b) Sample size 50				

Table 2.6: Proportion of true values above and below the 95% confidence intervals produced by four variants of a ZILP model from the analysis of simulated data

	Model 2.2	Model 2.3	Model 2.2a	Model 2.3a
Above mean count 95% CI	0.6%	0.6%	1.3%	1.7%
Below mean count 95% CI	0%	0%	1.7%	1.9%
Above variability 95% CI	26.7%	27.6%	0.4%	0.4%
Below variability 95% CI	0%	0%	4.6%	4.1%
Above zero-inflation 95% CI	0.2%	0.2%	0.2%	0%
Below zero-inflation 95% CI	3%	3.2%	0%	0%
(a) Sample size 10				
Above mean count 95% CI	1.5%	1.4%	3.6%	3.9%
Below mean count 95% CI	1.4%	1.4%	1.4%	1.4%
Above variability 95% CI	23.8%	23.7%	0.3%	0.4%
Below variability 95% CI	1.2%	1.4%	7.3%	8.3%
Above zero-inflation 95% CI	0.3%	0.3%	2.9%	2.6%
Below zero-inflation 95% CI	2.5%	2.2%	0.4%	0.4%
(b) Sample size 50				

Table 2.7: Relative RMSE of the median estimate for each parameter produced by four variants of a ZILP model from the analysis of simulated data

	Model 2.2	Model 2.3	Model 2.2a	Model 2.3a
Mean count	91.825	95.070	29.234	30.245
Variability	7.80	7.98	1.38	1.06
Zero-inflation	4.15	4.15	4.69	4.71
(a) Sample size 10				
Mean count	1.420	1.460	1.550	1.540
Variability	2.260	2.280	0.426	0.426
Zero-inflation	3.920	3.930	5.360	5.340
(b) Sample size 50				

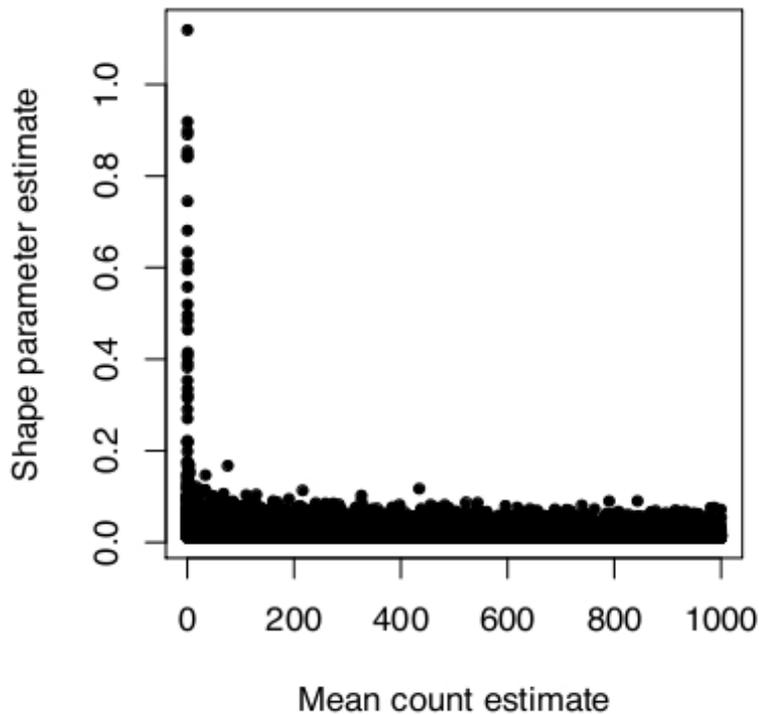


Figure 2.4: The relationship between mean count and the shape parameter ($\frac{1}{cv^2}$) at each iteration using a ZIGP model with simulated data consisting of zero counts - Model 2.2a analysis of a dataset consisting of 10 zero counts

2.4 COMPARING PRIOR DISTRIBUTIONS FOR THE GAMMA POISSON MODEL

similar characteristics to the sampling distribution may be required before the 95% confidence intervals perform as expected for this parameter with small datasets.

Choice of the sampling distribution for the mean count parameter appears to be less important, as the performance of the estimates for mean count have consistently been relatively good with all models examined. However, the choice of sampling distribution for mean count has a greater effect when the data consist of all zero counts. The extreme lack of information in the data results in the ZIGP model returning the prior distribution for mean, and a posterior distribution for the variance parameter that is weighted towards larger values of cv . The posterior distribution favouring a high mean seems counter-intuitive, but is a result of the prior for mean being biased towards higher values. In order to account for this high mean, the estimate for variability is increased to allow for the possibility of observing all zero-counts. This results in an estimate for variability that is linked to the estimate for mean, as seen in [Figure 2.4](#). The causes and implications of this phenomenon are discussed further in [Section 2.7](#).

From the analysis presented here, it appears that the models with the log uniform priors produce superior results, but it should be remembered that this is a function of the way in which the data were generated. A prior that is closest to the population distribution of the parameter that it represents will always produce superior results to a prior that is not close to the population distribution, because the true value of the parameter will on average be favoured more by this prior. Repeating this over several datasets results in a lower relative RMSE and more accurate 95% confidence intervals in the model with the prior that better represents the true population. This is one of the strengths of a Bayesian approach, as long as the population distribution of the parameter is known. Where information in the data is sparse, the posterior distribution will be heavily influenced by the prior distribution, even if the prior distribution is minimally informative. One solution to this could be the use of a Jeffreys prior for the inverse shape parameter. This type of prior ensures that the posterior is data driven, by ensuring that the prior is uninformative in areas of the likelihood where the data is informative. The Jeffreys prior is derived from the Fisher information matrix, and one of the key properties is that it is invariant under re-parametrisation such as log transformation. This means that the prior will be locally uninformative on any scale, so can be used to ensure that the posterior is likelihood driven at a given location in parameter space. Disadvantages of the Jeffreys prior include computational complexity in calculating the prior this way, and the potential to lead to improper posterior distributions. A subjective Bayesian would also argue that there is no situation where there is sufficiently little prior belief regarding a parameter to justify its use. For these reasons and because the aim here is to eventually be able to incorporate some prior belief about the likely value of the cv , the use of a Jeffreys prior will not be explored further.

Analyses produced by [Model 2.2](#) were quite similar to those produced by [Model 2.3](#), and

those produced by [Models 2.2a](#) and [2.3a](#) were also similar. This implies that there is little practical difference in using a gamma distribution centred on the mean rather than a gamma distribution centred on one. However, of the four models examined in this chapter, [Model 2.2a](#) seems to marginally outperform the others in terms of the number of datasets that returned an error or failed to converge. This model specification was therefore chosen to be used for further data analyses. For instances where a uniform prior for the inverse shape parameter would be more appropriate, the equivalent [Model 2.2](#) should be used.

2.5 Selecting model definitions for the lognormal Poisson model

2.5.1 Introduction

Similarly to the definition of a ZIGP model, there is more than one potential definition of a ZILP model possible using the BUGS syntax. For consistency between models the zero-inflated distribution is specified as for [Models 2.1](#) to [2.3a](#), but there are several different possibilities for specifying the parameters of the lognormal distribution. The lognormal distribution arises as a multiplication of a series of any distribution, and is therefore normal when log transformed (as discussed in [Section 1.2.1](#)). The mean and standard deviation of the normal distribution on the log scale ($l\mu$ and $l\sigma$) can be converted to/from the mean and standard deviation of the lognormal distribution (μ and σ) via the following calculations (derived from [Evans et al. \(2000\)](#)):

$$\begin{aligned} \mu &= \exp\left(l\mu + \frac{l\sigma^2}{2}\right) & \sigma &= \sqrt{\exp(2 \times l\mu + l\sigma^2) \times (\exp(l\sigma^2) - 1)} \\ l\mu &= \log \mu - \sqrt{\log\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right)} & l\sigma &= \sqrt{\log\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right)} \end{aligned}$$

The standard deviation of the distribution on the log scale ($l\sigma$) is an equivalent measure of variability to the cv as shown in the following equation (derived from [Equation 2.1](#)):

$$l\sigma = \sqrt{\log(cv^2 + 1)}$$

$l\sigma$ is also the standard measure to use to measure the ‘spread’ of the lognormal distribution. It therefore makes sense to place the prior on this in the model specification, although the alternative would be to place a prior on the standard deviation on the exponent scale. Similarly, a prior could be placed on the mean of the distribution on the log scale, which is equal to the log of the geometric mean of the exponent distribution, or the arithmetic mean of the exponent distribution. In this section, a comparison of models specified using these different parameters is made.

2.5.2 Materials and Methods

Models 2.5 to 2.8 represent four slightly different ways of parametrising the same model. In **Model 2.5** the priors are placed on $l\mu$ and $l\sigma$, in **Model 2.6** they are placed on μ and σ , and in **Model 2.7** the priors are placed on μ and $l\sigma$. In **Model 2.8**, the priors are again placed on $l\mu$ and $l\sigma$, but this time the prior on $l\sigma$ is log uniform rather than uniform (similar to **Model 2.2a**). In all models, in common with the ZIGP models discussed in **Section 2.3**, μ is sampled from between 0.001 and 1000, and cv from between 0.01 and 10. In models where the prior is placed on $l\mu$, this requires the lower and upper limits for the prior to be calculated so that the value for μ , which is calculated from the chains after the simulation is completed, cannot be sampled from outside the limits of 0.001 and 1000. The priors placed on $l\sigma$ and σ in each case allow the value for cv to vary between 0.01 and 10.

A total of one thousand datasets were each produced with 10 and 50 observations using R. Each dataset was generated using a ZILP model with parameters randomly selected from a $\log_{10}(\text{Uniform}(-1, 2))$ distribution for mean count, $\text{Uniform}(0, 0.75)$ distribution for the proportion of the population that is uninfected, and $\log_{10}(\text{Uniform}(-2, 0.70))$ distribution for cv respectively. Values for mean count were again taken from a log normal distribution to maximise the number of datasets with lower mean counts, and cv was also taken from a log normal distribution to increase the number of datasets that were close to a Poisson distribution. Analysis of each dataset with **Models 2.5 to 2.8**, and comparative analysis of the final outputs and simulation parameters, was performed as in **Section 2.3**.

2.5.3 Results

Table 2.8 shows the number of models successfully analysed by each model. Unlike the ZIGP models, none of the ZILP models returned an error for any of the datasets analysed. All models achieved convergence within an hour less frequently with the larger sample size, as was seen with the ZIGP models. Of the four ZILP models, **Models 2.5 and 2.8** returned the most successful analysis results.

The 95% confidence intervals produced by **Model 2.6** failed to contain the simulation parameter for a large proportion of datasets (**Table 2.9**). Those produced by **Models 2.5 and 2.7** contained the true parameter at least 95% of the time. For **Model 2.8** only, there appeared to be a tendency to over-estimate the variability. The relative RMSE in the median estimates for each model is shown in **Table 2.10**. **Models 2.5 and 2.8** produced similar relative RMSE for mean count and zero-inflation at each sample size, although the relative RMSE for variability was smaller for **Model 2.8** than **Model 2.5**. The relative RMSE produced by models **Models 2.6 and 2.7** were similar to or greater than those produced by **Models 2.5 and 2.8** in each case.

Table 2.8: Number of simulated datasets successfully analysed by two variants of a ZILP model, each with different priors, from the analysis of simulated data

	Model 2.5	Model 2.6	Model 2.7	Model 2.8
Total successfully analysed	923	851	922	913
Model returned an error	0	0	0	0
Failed to converge	0	72	1	10
Dataset all 0	77	77	77	77
(a) Sample size 10				
Total successfully analysed	916	871	944	828
Model returned an error	0	0	0	0
Failed to converge	77	122	49	165
Dataset all 0	7	7	7	7
(b) Sample size 50				

Table 2.9: Proportion of true values above and below the 95% confidence intervals produced by two variants of a ZILP model, each with two different priors for variability, from the analysis of simulated data

	Model 2.5	Model 2.6	Model 2.7	Model 2.8
Above mean count 95% CI	0.1%	42.9%	0.5%	0.5%
Below mean count 95% CI	1.8%	0.1%	0.2%	3.6%
Above variability 95% CI	0.6%	63.8%	2.4%	0%
Below variability 95% CI	0.5%	0.1%	0.4%	5.7%
Above zero-inflation 95% CI	0.1%	0.6%	3.9%	1%
Below zero-inflation 95% CI	0.1%	1.5%	0.1%	0.1%
(a) Sample size 10				
Above mean count 95% CI	0.9%	21.6%	0.9%	1%
Below mean count 95% CI	2.3%	0.7%	1%	3.7%
Above variability 95% CI	1.8%	27.9%	1.8%	0.9%
Below variability 95% CI	1.9%	0.4%	1.2%	5.7%
Above zero-inflation 95% CI	0%	0%	0.9%	1.2%
Below zero-inflation 95% CI	0.9%	16%	0%	0.1%
(b) Sample size 50				

Table 2.10: Relative RMSE of the median estimate for each parameter produced by two variants of a ZILP model, each with two different priors for variability, from the analysis of simulated data

	Model 2.5	Model 2.6	Model 2.7	Model 2.8
Mean count	1.757	52.522	27.000	1.651
Variability	3.518	412.179	7.858	0.872
Zero-inflation	7.141	9.100	11.387	7.327
(a) Sample size 10				
Mean count	0.882	9.815	1.356	0.848
Variability	1.820	1322.350	2.055	0.698
Zero-inflation	3.274	3.035	4.141	4.033
(b) Sample size 50				

2.5.4 Discussion

In contrast to the ZIGP models, none of the ZILP models returned an error. The ZIGP models appeared to be more likely to return an error in datasets with high cv than low cv , suggesting an issue with sampling of extreme values of this parameter. One potential explanation for the ZILP models not returning an error could be the shape of the lognormal distribution with high variance compared to the shape of the gamma distribution with a small shape parameter - the gamma distribution with these parameters is a single curve with a mode close to 0, whereas the lognormal retains the skewed peak shape. This results in extremely small values being more likely to occur with the gamma distribution than the lognormal distribution.

As with the ZIGP models, the number of datasets that achieved convergence within the one hour time frame was quite different between model specifications. There are several possible explanations for a failure to converge such as bi-modality and unidentifiability. The most likely explanation is auto-correlation which highlights the impact that subtly different model formulations can have on the degree of auto-correlation within the resultant MCMC chains. **Models 2.6** and **2.8** returned the fewest converged datasets, suggesting a higher degree of auto-correlation. This indicates that the model converged most easily with a prior on the $l\sigma$. In contrast, there did not appear to be a large discrepancy between the number of datasets that achieved convergence between **Models 2.5** and **2.7**. However time to convergence is dependent on the most heavily auto-correlated parameter which is invariably the variability parameter, so that differences in mean parameter auto-correlation between these models would not impact on the convergence time. **Models 2.5** and **2.7** also produced similarly defined 95% CI, although the relative RMSE was lower for **Model 2.5** in every case. This

suggests that a prior placed on $l\mu$ directly should be preferred to calculation of $l\mu$ from a prior placed on the mean on the exponent scale.

Based on the results of this simulation, using priors based on $l\mu$ and $l\sigma$ produces superior results to priors based on μ and σ . In contrast to the ZIGP model formulation, there did not appear to be an advantage to using a log-uniform distribution (Model 2.8) over a uniform distribution (Model 2.5) from which to sample $l\sigma$, possibly because the distribution of cv when using a uniform distribution for $l\sigma$ is more heavily weighted to lower values than when using a uniform distribution for the inverse shape parameter of the gamma distribution. Therefore, the model specification given by Model 2.5 was used for further analyses.

2.6 Comparing JAGS and WinBUGS results

2.6.1 Introduction

Bayesian analysis using MCMC can be achieved in a number of ways. Specific software to implement Bayesian MCMC models using Gibbs samplers include JAGS (Plummer, 2008), and the better known BUGS project of WinBUGS (Lunn et al., 2000) and OpenBUGS (Thomas et al., 2006). A brief comparison of the results obtained using JAGS with results from an identical model implemented in WinBUGS is warranted, to ensure that inference made is independent of the choice of MCMC implementation.

2.6.2 Materials and methods

A ZIGP model was implemented in JAGS and an identical model was created in the WinBUGS language, with the same log uniform prior distribution for variability. One hundred datasets were generated with a sample size of 50, as described in Section 2.3, and analysed with each model. Unlike JAGS, WinBUGS cannot easily be interfaced with R on Unix systems, so a different method of running both models was employed. The WinBUGS model was run using the batch script feature of WinBUGS along with the coda output facility, and R was used to check the results for convergence using the Gelman-Rubin statistic, and to analyse the results to produce upper and lower 95% confidence intervals and median estimates. The model was run using two chains, and 5,000 burn-in iterations followed by 10,000, and if necessary 100,000 or 500,000 sampling iterations in order to achieve convergence in the sampled chains. For ease of comparison the same procedure was followed for the JAGS model. Both sets of analyses were saved to a text file and input into R for comparison.

2.6.3 Results

The number of datasets successfully analysed by both models (Table 2.11) was very similar. In addition, the datasets that were not analysed successfully with one model tended to not be analysed successfully with the other model. Figures 2.5 to 2.7 show the correlation between the JAGS and WinBUGS estimates for mean count, cv and zero-inflation. Except for a small amount of variation between the two models for estimates for cv in some datasets, the estimates were very similar between the models.

Table 2.11: Number of datasets successfully analysed by equivalent JAGS and WinBUGS models from the analysis of simulated data

		WinBUGS			Total
		Success	Error	Not conv.	
JAGS	Total number successfully analysed	86	0	4	90
	Model returned an error	1	7	0	8
	Failure to converge	1	0	1	2
	Total	88	7	5	100

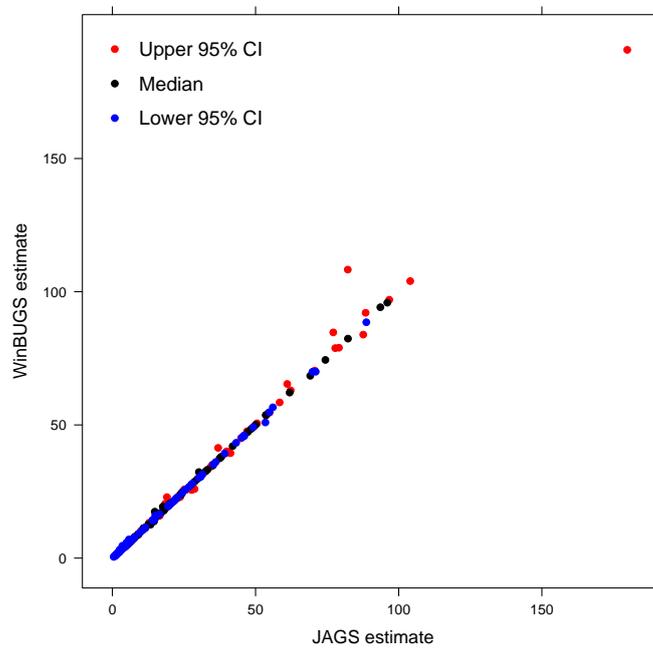


Figure 2.5: Comparative estimates of mean count from the analysis of simulated data using equivalent JAGS and WinBUGS models

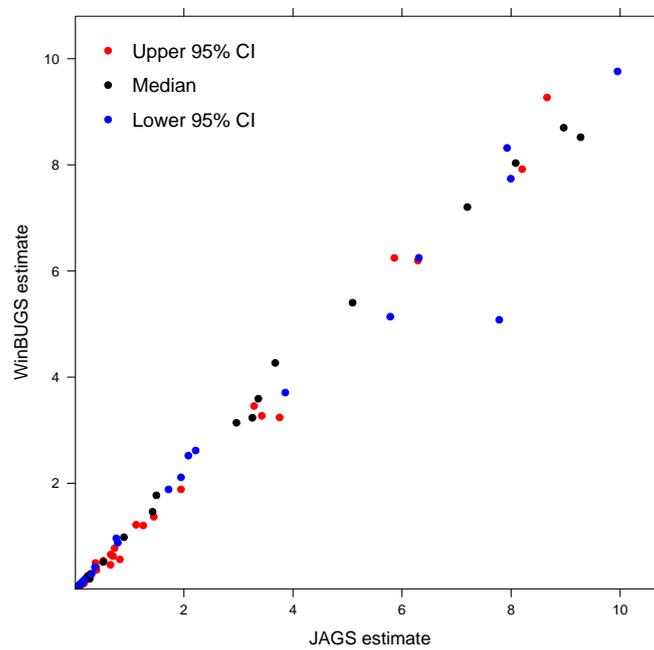


Figure 2.6: Comparative estimates of cv from the analysis of simulated data using equivalent JAGS and WinBUGS models

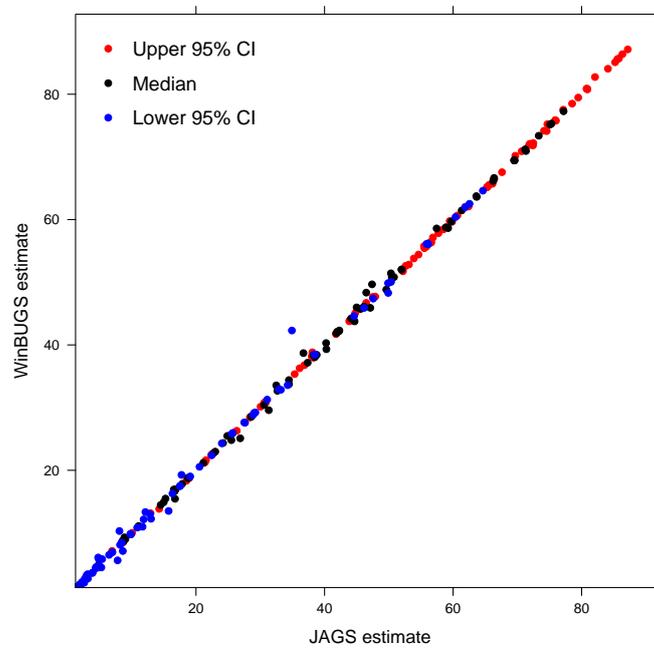


Figure 2.7: Comparative estimates of zero-inflation from the analysis of simulated data using equivalent JAGS and WinBUGS models

2.6.4 Discussion

Apart from relatively minor differences between the analyses, and given that the dataset was deliberately created to contain a small amount of data resulting in confidence intervals for several datasets being quite wide for both models, the results of the two models are almost identical. The minor differences between the results are most likely the result of either Monte Carlo error reflecting a need to take more samples to adequately describe the distributions, or imperfect convergence for some datasets with one or both models. It is therefore reasonable to conclude that use of JAGS and WinBUGS produce equivalent results, providing that the models used are identical.

2.7 Discussion

Inference made using MCMC will always depend on the use of the correct distribution to describe the data, and adequate attention to convergence and sample length diagnostics to ensure that any conclusions drawn are correct (Gilks et al., 1998). The choice of an appropriate prior distribution is also critical, and will have strong influence over the posterior distribution where information in the data concerning a parameter is sparse (Goodman, 2005). What is less obvious is that slight alterations in the model specification, using equivalent sampling limits for the same parameters but subtly different formulations or priors, can dramatically affect the degree of auto-correlation in the model and ultimately the inference made. Where novel or complex models are being used in the real world, validation using simulated data is important to ensure that posterior distributions of parameters are sensible.

For the model comparisons in this chapter, the priors used in each model were kept as similar as possible so as to minimally affect the behaviour of the model. The priors used were uninformative within limits of the allowed parameter space which was intended to be nonrestrictive in terms of real faecal egg count data. The prior for zero-inflation allowed true values to lie anywhere in the possible parameter space of between 0% and 100%, but the priors for mean and cv were limited to the extremes of the values likely to be encountered in real data. Mean egg counts of less than 0.001 are unlikely to demonstrate any eggs given the sample sizes typically encountered, and mean egg counts of over 1000 are unlikely to be observed. The cv was limited to values of between 0.01 and 10; these are reasonable values for faecal egg count data as a cv of below 0.01 would be almost indistinguishable from Poisson variation, and above 10 is much higher than reported values. Limiting the cv to these values should help prevent the MCMC algorithm exploring, and possibly encountering problems with, parameter values that are not practical. Although these priors were uninformative on the scale on which they were specified, the same priors become quite informative when transformed. This resulted in poor inference on cv for the ZIGP [Model 2.2](#) with smaller

sample sizes when cv was simulated on a different scale, compared to the results obtained from using a variability prior specified on the log scale in [Model 2.2a](#). The same effect was not observed with the inference on mean, which was also simulated on a different scale to the priors, reflecting the comparatively large amount of information in the data for this parameter compared to estimating the variability parameter.

Datasets consisting of all zero counts were discarded from all analyses presented in this chapter, as an alternative method of analysing these data would be more appropriate than attempting to fit a distribution with an absence of observations. As demonstrated in [Section 2.4](#), analysis of data with no observed counts in this way result in posterior confidence intervals for each parameter which are highly dependent on the prior distribution chosen for that parameter, as well as the prior distribution chosen for other parameters. This results in a correlation between the estimates for mean count and variability at each iteration ([Figure 2.4](#)). This phenomenon is observed because the likelihood of observing zero counts with a high mean and low cv is very small, therefore low estimates of cv (high estimates of the shape parameter) are only observed at iterations where the estimate for the mean is small. As there is so little information in the data the mean is informed by the prior which tends towards large values, resulting in high estimates for the cv in order to account for the high mean. The reason for the model allowing the prior for mean to distort the prior for cv , rather than *vice versa*, is not clear. It may be that this is a result of the Gibbs sampler updating the mean before the cv , resulting in a choice of value for cv that is dependent on the previously chosen value for the mean. However, use of an equivalently parametrised Metropolis-Hastings algorithm produces similar inference on cv , suggesting that this is not the case. An alternative explanation may be that the Poisson part of the mixture distribution results in zero observations being so much more likely with extremely low lambda values that the posterior for cv is pushed up to allow these extremely small lambda values to be sampled more often. Due to the large estimates of cv , the probability of observing very low lambda values would then be sufficiently high at any value of mean to allow this solution to be sampled from intensively. The same effect is not observed using the ZILP [Model 2.5](#) as the prior for mean count is on the log of the geometric mean, tending towards smaller rather than larger values.

2.8 Development of an R package to facilitate data analysis

Having developed model definitions to be used for the ZILP and ZIGP models, an R package was developed to aid the analysis of data using the (zero-inflated) Poisson, (zero-inflated) gamma Poisson or (zero-inflated) lognormal Poisson distributions. This R package entitled ‘bayescount’ is available to download from <http://cran.r-project.org/web/packages/bayescount> (see the list of publications on page [viii](#) for citation information). Because the choice of prior distribution for the inverse shape parameter has been shown to have an effect on inference

made, the ability to change this prior was built into the package. A function to calculate the likelihood at each iteration for each model is also included. The `runjags` package ([Denwood, 2008](#)) is used for lower-level functions to call JAGS, with convergence diagnostics based on the Gelman-Rubin statistic and run length diagnostics based on the Raftery and Lewis diagnostic. This software is used to implement the MCMC methods discussed in the remainder of this thesis.

CHAPTER 3

Distributional selection and model fit

Distributional selection and model fit

3.1 Introduction

The distribution of parasites within a group of animals provides an indication of the degree of parasitism present, as well as the dispersion of parasites across hosts (Wilson et al., 1996). Parasite distributions tend to be aggregated; some individuals are more heavily parasitised than others, which has an important influence on control of the parasite as the majority of pasture contamination is contributed by relatively few individuals (Barger, 1985; Sréter et al., 1994). Similarly, other factors such as the presence of uninfected animals within a group could also have implications for parasite control measures. Further benefits of understanding the distribution of parasites between individuals include the quantification of genetic components of nematode resistance in sheep and cattle, and the ability to obtain more accurate estimates of the true mean of the parasitised population in order to improve the detection of anthelmintic resistance. However, all of these potential benefits depend on the use of a statistical technique that is capable of extracting the most useful information from the available data. For parametric techniques, a parametric distribution which is well approximated to the process underlying the observed data is also required.

The most widely used method for implementing population modelling in parasitology is currently maximum likelihood estimation (McCullagh and Nelder, 1989; Wilson et al., 1996; Torgerson et al., 2005). One disadvantage of maximum likelihood estimation is that the distribution of error for each parameter is assumed to be normally distributed around the maximum likelihood estimate for the parameter. Although this asymptotic assumption is valid with large enough sample sizes, the number of data points required may be far in excess of the data available (Stear et al., 2004). Other more computationally complex techniques such as likelihood profiling or MCMC may produce more reliable confidence distributions for these types of data.

The Poisson distribution is a simple model that is capable of describing count data such as FEC. However, as FEC data are over-dispersed with respect to the Poisson, this distribution is usually adapted to allow the mean of the Poisson process to vary, according to another con-

tinuous distribution such as the gamma or lognormal (as detailed in [Section 1.2.1](#)). Another type of model which has been used to describe FEC data is the zero-inflated model ([Nødtvedt et al., 2002](#)). This is useful when describing FEC data which contains many zero counts, and introduces the idea that a subset of the population may be non-shedding because of recent anthelmintic treatment, lack of exposure to parasites, or resistance to infection. It is theoretically impossible to observe a positive count from these animals, so they are not part of the distribution which describes the infected individuals within the group. In effect, a zero-inflated model analyses the animals with a positive count to extract a lambda value which describes the infected herd, and then extrapolates from this the number of animals with a count of zero that would be expected to form part of the Poisson process. Any extra animals with a count of zero are then defined as being part of the zero-inflated population; the proportion of the group which are part of this population is therefore the prevalence of non-infection, or one minus the prevalence of infection, within the group. The natural extension of the zero-inflated Poisson model is to allow over-dispersion within the infected animals, using a zero-inflated negative binomial (ZINB) model, which is equivalent to a ZIGP, or a ZILP model. This model returns values for mean count and over-dispersion of the infected population, as well as an estimate of the proportion which are uninfected.

It is not clear which of these distributions and statistical methods is most likely to be useful for the analysis of FEC data. Distributional selection is usually made on the basis of empirical fit, taking into account the number of parameters in the model ([Akaike, 1973](#); [Schwarz, 1978](#); [Johnson and Omland, 2004](#)). Such penalised likelihood measures are widely used as the basis of model selection, but the penalty for increasing the number of parameters in the model differs between methods such as BIC, AIC and DIC. The underlying assumption is that the simplest possible explanation is to be preferred, a concept frequently known as Occam's razor and attributed to Occam ([Thorburn, 1918](#)). A Bayesian approach to model selection relies on biological plausibility as well as empirical fit as a basis for model selection. Such an approach may allow an over-specified model, such as the ZIGP, to be used in place of a simpler model, such as a gamma-Poisson, where there is a biological reason to believe that the additional parameters may be important, regardless of evidence for their existence based on empirical fit ([Miaou and Lord, 2003](#)). In the case of the ZIGP model, if there were sufficient evidence for the absence of appreciable zero-inflation or over-dispersion in the dataset, the estimates for these parameters would reduce until the model collapsed to the simpler form. This approach may yield more useful results than attempting to select the distribution based on a penalised likelihood, which is difficult to calculate as the number of degrees of freedom in a mixture model is not well defined.

Selection of the lognormal-Poisson or the gamma-Poisson must take a different approach because the two are not nested. The widespread adoption of the negative binomial appears to be based mainly on convenience; the alternative formulation to the gamma-Poisson that relieves the requirement for lengthy integration over lambda values of the Poisson distribution

is quite appealing. The distribution has also been shown to fit FEC data well, partly as a result of the flexibility of the distribution, and the use of an over-dispersed Poisson distribution such as the gamma-Poisson has been well justified (Crofton, 1962; Pacala and Dobson, 1988; Grenfell et al., 1995). However, little attempt has been made to justify the gamma distribution biologically compared to the potential reasoning behind the use of the lognormal distribution outlined in Section 1.2.1. In addition, with the advent of modern computing power the computational burden of having to integrate the lambda value of the compound distribution associated with using distributions such as the lognormal-Poisson become much easier to achieve, potentially reducing the reliance on the negative binomial.

In this chapter, the inference made using some of these distributions and the utility of model selection based on empirical fit are explored using simulated data.

3.2 Comparison of MCMC and ML for analysis of ZIGP data

3.2.1 Introduction

The current gold standard for estimating parameters of distributions in parasitology is maximum likelihood estimation (McCullagh and Nelder, 1989; Wilson et al., 1996; Torgerson et al., 2005), available using standard statistical software (SAS, 2003; R Development Core Team, 2009). Computationally intensive methods have been recently used as an alternative to maximum likelihood analysis, including likelihood profiling, bootstrapping and Bayesian analysis by means of MCMC methods using software such as WinBUGS (Lunn et al., 2000). The term MCMC used in this chapter specifically refers to Bayesian analysis by means of MCMC methods; although maximum likelihood estimates could also be achieved using MCMC methods. The major advantage of MCMC is that a combination of the available data and prior knowledge of the parameter can allow inferences to be made about the parameter, even in the absence of sufficiently informative data. MCMC methods also define the entire distribution of the parameters rather than just the maximum likelihood estimate and standard error, and allow models to be easily modified to represent the biological process (Goodman, 2005). Disadvantages of MCMC methods include the need to check for adequate convergence in order to avoid inappropriate conclusions being drawn (Brooks and Roberts, 1998; Toft et al., 2007), as well as the increased computer processing time required to analyse data in this way.

Both ML and MCMC methods can be used to fit a zero-inflated negative binomial model, although simple maximum likelihood estimates are the more widely used of the two. Given a large enough dataset each technique would be expected to produce equivalent results, although results are likely to differ between the two approaches with less informative datasets.

Comparison of the accuracy and reliability of the two techniques at different sample sizes is therefore of interest.

3.2.2 Materials and Methods

A total of 1000 datasets were simulated with mean egg count randomly drawn from between 0.1 and 100, and cv randomly drawn from between 0.01 and 5. Both of these parameters were drawn from log-Uniform distributions to maximise the number of datasets with small means and cv . Zero-inflation was sampled uniformly from between 0 and 75% for each dataset. For each of these 1000 combinations of parameters, data were simulated using a zero-inflated gamma-Poisson distribution in R with sample sizes of 10, 100 and 1000 counts. The resultant 3000 datasets were stored along with the corresponding true parameter values.

MCMC analysis was achieved using [Model 2.2a](#) as described in [Section 2.4](#). Datasets that failed to converge within a period of one hour and datasets where JAGS returned errors were discarded as before. ML analysis was done using the ‘zicounts’ package ([Mwalili](#)) in R. The ML analysis also failed to return results or did not converge for some datasets. These were removed as for the un-converged datasets from the MCMC analysis. In addition, for some returned datasets, the standard error of one or more parameter was returned as impossible to calculate or infinite. An ML estimate with an unknown 95% confidence interval is of little use, and these tended to result in inaccurate estimates for the other parameters, therefore these results were also removed. Any dataset that was removed for one of the techniques was removed for both in order to avoid biasing the results. The final outputs for each model as well as the simulation parameters were then input into R for comparative analysis. The 95% credible/confidence intervals output by the two types of model were compared to the simulation parameter for each dataset, and the relative RMSE of the MCMC median and ML estimates calculated for each combination of parameter and sample size.

3.2.3 Results

The percentage of data sets that were analysed successfully was higher for MCMC than ML ([Table 3.1](#)). [Table 3.2](#) shows the percentage of the simulated parameters that fell inside the 95% CI for each combination of model, sample size and parameter. In every case the 95% credible intervals produced by MCMC contained a greater proportion of the true values than the 95% confidence intervals produced by ML. The 95% confidence intervals produced by MCMC contained the true parameter at least 95% of the time for mean and zero-inflation, but less than this for cv . Of the 51 and 37 confidence intervals not containing the true cv parameter at sample sizes 100 and 1000 respectively, there were only 7 for sample size 100 and 10 for sample size 1000 for which the true value was below the confidence interval,

Table 3.1: Number of datasets returned by MCMC and ML techniques for each sample size from the analysis of simulated ZIGP data

	MCMC method	ML method
Total successfully analysed	778	538
Model returned an error	52	0
Incalculable 95% CI	0	290
Failed to converge	0	2
Dataset all 0	170	170
(a) Sample size 10		
Total successfully analysed	936	680
Model returned an error	60	0
Incalculable 95% CI	0	313
Failed to converge	0	3
Dataset all 0	4	4
(b) Sample size 100		
Total successfully analysed	829	723
Model returned an error	17	0
Incalculable 95% CI	0	274
Failed to converge	154	3
Dataset all 0	0	0
(c) Sample size 1000		

Table 3.2: Proportion of 95% confidence intervals produced by ML and MCMC techniques that contained the true parameter from the analysis of simulated ZIGP data

Sample size	Parameter	MCMC method	ML method
Sample size 10	Mean	0.978	0.919
	Variability	0.919	0.696
	Zero-inflation	0.988	0.958
Sample size 100	Mean	0.956	0.890
	Variability	0.925	0.678
	Zero-inflation	0.956	0.905
Sample size 1000	Mean	0.957	0.896
	Variability	0.931	0.715
	Zero-inflation	0.962	0.950

indicating a tendency for the MCMC method to under-estimate rather than over-estimate cv for the larger sample sizes. For the smallest sample size, 37 of the 67 confidence intervals not containing the true parameter had similarly under-estimated cv .

The relative RMSE for each method is shown in [Table 3.3](#). The MCMC method produced smaller relative RMSE for the largest sample size and cv with a sample size of 100, but had a large relative RMSE for the mean particularly at the smallest sample size. The relationship between the MCMC median estimate of mean and MCMC median estimate of cv for the sample size 10 data is shown in [Figure 3.1](#). There are a group of datasets with high estimated mean and corresponding high estimate of cv . The majority of the datasets with a high MCMC median estimate for mean had a small empirical mean ([Figure 3.2](#)).

3.2.4 Discussion

From the results presented here, it is apparent that analysis of data simulated with these parameter values presents some difficulties. The relative RMSE for zero-inflation was poor for both techniques, even with a sample size of 1000. The ML technique also produced poor relative RMSE for variability, although this was markedly improved with MCMC at the larger sample sizes. The ML analysis produced much smaller relative RMSE than the MCMC technique with a sample size of ten, which is likely to be a result of the prior distribution used for the mean for the MCMC technique. This prior distribution (Uniform(0.001, 1000)) would result in a tendency for the estimate of mean to be much higher than the distribution from which the mean was simulated. [Figure 3.1](#) shows a group of results with a median estimate for mean of ≈ 400 and corresponding high cv , indicating the possibility that these datasets were the result of a high mean with a large amount of variability. [Figure 3.2](#) demonstrates a small

Table 3.3: Relative RMSE produced by ML and MCMC techniques for each parameter and sample size from the analysis of simulated ZIGP data

Sample size	Parameter	MCMC method	ML method
Sample size 10	Mean	231.38	1.76
	Variability	8.12	1.05
	Zero-inflation	9.29	5.83
Sample size 100	Mean	1.07	0.98
	Variability	0.51	1.50
	Zero-inflation	7.10	5.97
Sample size 1000	Mean	0.38	0.44
	Variability	0.47	1.11
	Zero-inflation	5.73	6.00

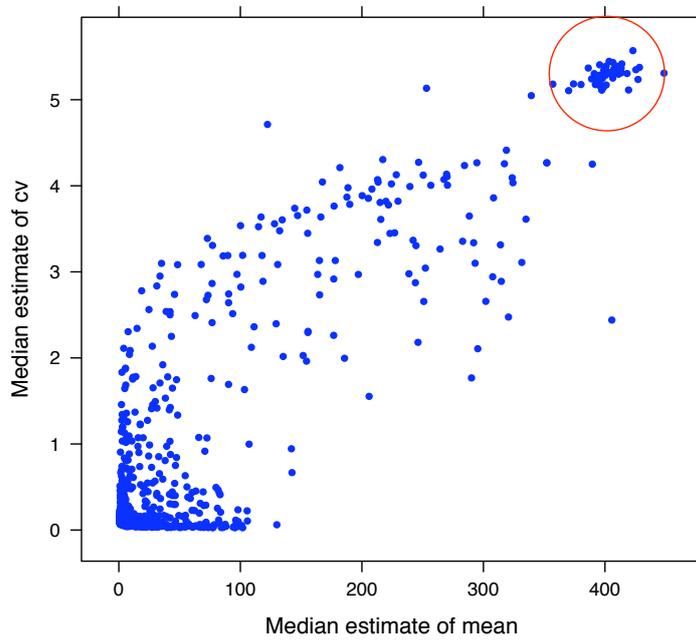


Figure 3.1: The relationship between the MCMC median estimates for mean and *cv* from simulated data with a sample size of 10 - group of results with high mean and corresponding high *cv* circled in red

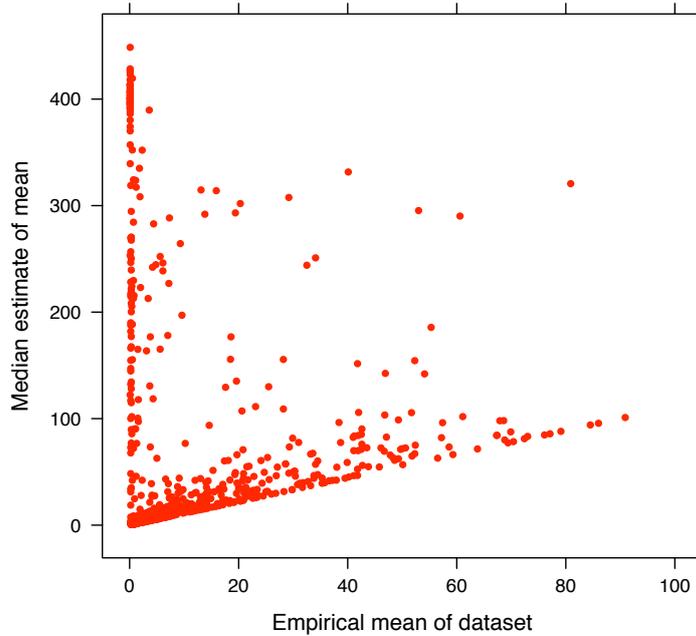


Figure 3.2: The relationship between the MCMC median estimate for mean and the empirical mean of the simulated dataset with a sample size of 10

empirical mean for these datasets, resulting in a similar effect to that discussed for datasets with all zero values discussed in [Section 2.7](#). The relative RMSE for each parameter appeared to be reduced with each increase in sample size for the MCMC method, but only with the mean parameter for the ML method. However, the values produced are not necessarily comparable between sample sizes as a result of the removal of a different sub-set of the datasets and parameter sets for each sample size.

In contrast to the relative RMSE, the MCMC technique consistently out-performed the ML technique in terms of the true confidence of the nominal 95% confidence intervals produced. The ML technique failed to identify the true interval as much as 32% of the time for variability and 10% for mean count, although the 95% CI for zero-inflation were closer to 95% confidence. The MCMC technique produced more reliable confidence intervals for all parameters, although confidence was slightly reduced for variability, possibly as a result of a combination of the lack of information regarding this parameter in the data and the prior tending towards lower values of cv than the distribution from which the parameter was simulated. Use of a prior distribution for cv and mean that was closer to the distribution from which these parameters were simulated would have greatly improved the apparent performance of the MCMC technique, but the intention here was to use priors that were minimally informative. However, it appears that the lack of information available in these types of data result in even minimally informative priors having a demonstrable effect on the posterior probability densities.

Another difference between the two techniques was the proportion of datasets that were successfully analysed. The MCMC technique returned usable parameter estimates for 90.0% of the total non-zero datasets, compared to only 68.7% for the ML technique. The ML technique produced more usable results, as a proportion of the total number of non-zero datasets, with increasing sample size. This could be as a result of datasets with higher cv tending towards being all zero, but is more likely to be a result of the increased information in the datasets with a larger sample size. A large proportion of the ML results indicated that convergence had been achieved, but were unable to determine confidence intervals for one or more parameters. The results from these models were extremely inaccurate, and highlight the importance of careful analysis of the outputs of generalised linear models before making inferences based on their results. Because of the way that MCMC prior distributions are defined, parameters cannot be returned with infinite or incalculable confidence intervals using MCMC. Using this method, the datasets that contain no information about one or more variables would either return the prior distribution or be removed due to poor convergence. Only 154 datasets failed to converge using the MCMC technique, and only with a sample size of 1000. This is a result of the time constraint on convergence; models were only allowed to run up to a maximum of one hour to allow chains to converge, therefore models that took longer to update as a result of more datapoints would take longer to achieve the same number of iterations and so would be less likely to converge within the same time period.

Chains were typically quite auto-correlated, resulting in a large number of required iterations calculated by the Raftery and Lewis diagnostic. Mean sample sizes of over 63,000, 100,000 and 167,000 iterations respectively were required for the 678 datasets that converged at all sample sizes of 10, 100 and 1000. The higher auto-correlation observed with larger sample sizes possibly reflects the greater impact of the likelihood on the posterior distribution compared to the smaller sample sizes, resulting in a reduced variance in the posterior for the variability parameter and consequently poorer mixing across the distribution.

It can be seen from the analysis presented that the 95% confidence intervals returned by ML were not truly of 95% confidence. The implementation used calculated these intervals assuming a normal distribution for the error of the parameters, which is likely to be partly responsible for this poor performance. It is possible to calculate 95% confidence intervals using ML without making this assumption, which would have improved the proportion of datasets where the true parameter was within the 95% CI. However, this increases the computational complexity of the ML technique which reduces the advantage over MCMC in this regard, and would not improve the relative RMSE associated with the results.

The comparatively poor performance of both methods with a sample size of ten indicates that very small datasets seldom produce accurate parameter estimates, and the relatively high error rate at sample size 100 for zero-inflation and over-dispersion is of concern because many biological samples are no larger than this. Given the difficulty in obtaining and studying very large sample sizes, computationally intensive methods such as Bayesian MCMC methods can and should be used to improve estimates of distribution parameters when analysing zero-inflated negative binomial data of smaller sample sizes, such as typically found with faecal egg counts.

3.3 Use of a ZIGP model to analyse GP and ZIP data

3.3.1 Introduction

It has previously been stated that parasite count data such as FEC would be expected to follow a Poisson distribution if parasites were mixed randomly in the population (Elliott, 1977), or a Poisson distribution with a further, possibly multi-modal, distribution describing the varying mean of the Poisson part if parasites are aggregated in the population. If the distribution describing the Poisson means is taken to be gamma, then the resultant compound distribution is the gamma-Poisson. This distribution is equivalent to the negative binomial (Venables and Ripley, 2002; Vose, 2004), and has been used in the analysis of countless parasite distributions where it usually provides a good empirical fit to the data as it allows for the data to be over-dispersed with respect to the Poisson (May and Anderson, 1978;

Wilson and Grenfell, 1997; Shaw et al., 1998). The gamma-Poisson accounts for several zero counts representing no observed parasites or eggs, which is a frequent feature of many datasets in parasitology, by producing a correspondingly large estimate for the cv of the distribution. One alternative to this approach is to use a zero-inflated Poisson distribution; a mixture distribution which allows the data to be separated into a group which may be positive or zero (the Poisson part) and a group which may only be zero (the “extra zeros”) (Lord et al., 2005). This allows the Poisson distribution to fit data with a larger number of zero counts than would be expected given the mean of the non-zero counts, and is the simplest form of multi-modal distribution that could be fitted to parasite burden or FEC data. The gamma-Poisson and zero-inflated Poisson distributions can be combined, resulting in a zero-inflated gamma Poisson distribution (Nødttvedt et al., 2002; Walker et al., 2009). This ZIGP distribution allows for large numbers of zero observations to be accounted for by either zero-inflation or a large cv in the infected population, or potentially a mixture of the two. The two possibilities are effectively separated based on the cv of the non-zero counts.

It is important to note that although the ZIGP model allows for the data to be both zero-inflated and over-dispersed, it is also capable of describing either a simple Poisson process, zero-inflated Poisson process or gamma-Poisson process by simply evaluating the zero-inflation and/or cv parameters as being close to zero. Although a frequentist approach requires that the simplest model that adequately fits the data should be used to preserve parsimony, a Bayesian approach is concerned more with the accuracy of the posterior distribution, and would therefore allow a ZIGP model to be used for data without zero-inflation or over-dispersion as long as the posterior value for zero-inflation or over-dispersion is close to zero. A similar approach has been used by Toft et al. (2006) to determine if use of the gamma-Poisson is justified in the analysis of micro-organism counts. The amount of zero-inflation and over-dispersion present in the data is rarely known prior to analysis, therefore being able to confidently use a single model to analyse all data types would represent a benefit to parasitologists in terms of both time spent evaluating the fit of several different models, and in eliminating errors in analysis caused by using inappropriate models. The difficulty in defining the degrees of freedom in a zero-inflated mixture model makes model selection based on penalised likelihoods even more difficult and controversial than with models with more easily defined effective parameters, although score tests for zero-inflated models have been used (Xiang et al., 2007). These studies have not examined the impact of using zero-inflated models on the accuracy of inference made from datasets where zero-inflation is not important, and conversely the impact of using non zero-inflated models in the analysis of datasets where zero-inflation is a feature. The accuracy of inference made using the ZIGP and simpler models when analysing datasets with and without over-dispersion and zero-inflation is therefore of interest, so that the consequences of incorrect model selection can be quantified.

3.3.2 Materials and methods

A total of 1,000 parameter combinations of mean count, cv and zero-inflation were generated using R, as presented in [Section 3.2](#). For each parameter combination, a ZIGP distribution was used to simulate a dataset with a sample size of 100. Zero-inflated Poisson (ZIP) and gamma-Poisson (GP) distributions were used to simulate further datasets of the same sample size, ignoring the cv and zero-inflation parameter respectively. Each of the 3,000 datasets were analysed using the ZIGP MCMC [Model 2.2a](#), as well as zero-inflated Poisson and gamma-Poisson models based on [Model 2.2a](#) but with the corresponding variability and zero-inflation components removed. Models were run until convergence for a maximum of one hour before calculating the necessary sample size as before. The results for each combination of model and dataset were saved and input into R for comparative analysis.

In order to directly compare the output for mean count between the zero-inflated and non-zero-inflated models, the mean count of the infected population calculated by the ZIGP and ZIP models was converted to the population mean as calculated by the Poisson and GP models using the following formula:

$$\begin{aligned} \text{Corrected mean count} &= \frac{(\text{Number infected} \times \text{infected mean}) + (\text{Number uninfected} \times 0)}{\text{Number of animals in the whole population}} \\ &= (1 - \text{zero inflation proportion}) \times \text{mean count} \end{aligned}$$

The same transformation was applied to the simulation parameters, and all references to the mean count here refer to this parameter.

The relative RMSE was calculated for each combination of parameter, model and dataset as described previously. In addition, the mean relative bias, calculated in a similar way to the relative RMSE with a negative value indicating on average a lower true value compared to the estimate, and mean relative size of the 95% confidence intervals, were compared between models. The latter was calculated using the following formula applied to each parameter applicable to the data type:

$$\text{relative range} = \frac{U - L}{M}$$

Where L and U are the lower and upper 95% confidence interval limits, and M is the median estimate.

3.3.3 Results

In Table 3.4, the number of models that returned an error or failed to converge for each model and data type are shown. Convergence was achieved for all datasets, except those that returned an error. The ZIGP and GP models returned errors for over twice as many ZIGP datasets than GP or ZIP datasets. The ZIP model returned errors for datasets that included a *cv* component. The relative RMSE, mean relative bias, proportion of true values that fell above and below the 95% confidence interval estimates, and mean relative CI size for each combination of parameter, model and dataset are shown in Table 3.5. The overspecified ZIGP model produced very similar statistics to the relevant simpler model for the mean count and zero-inflation parameters with the GP and ZIP data. The ZIGP model underestimated the *cv* present in 14% of the GP datasets, but produced a smaller relative RMSE and bias than the GP model. The GP model overestimated the *cv* present in 62% of the ZIGP datasets, resulting in a large bias and relative RMSE compared to the ZIGP model. Similarly, the ZIP model overestimated zero-inflation in 19% of the ZIGP datasets although the relative RMSE and bias was more similar to the ZIGP model. The relative RMSE for zero-inflation with the ZIGP model was much higher for the ZIGP data than the ZIP data. The relative RMSE and bias were similar between the ZIP and ZIGP model for all datasets, although the GP model produced larger relative RMSE and bias and larger CI sizes for the ZIP and ZIGP

Table 3.4: Number of datasets returned by each combination of GP, ZIP and ZIGP model with GP, ZIP and ZIGP simulated datasets

	GP model	ZIP model	ZIGP model
Total successfully analysed	980	955	985
Model returned an error	20	45	15
Failed to converge	0	0	0
Dataset all 0	0	0	0
(a) Gamma-Poisson data			
Total successfully analysed	967	999	974
Model returned an error	32	0	25
Failed to converge	0	0	0
Dataset all 0	1	1	1
(b) Zero-inflated Poisson data			
Total successfully analysed	920	951	934
Model returned an error	76	45	62
Failed to converge	0	0	0
Dataset all 0	4	4	4
(c) Zero-inflated gamma-Poisson data			

data compared to the ZIGP model.

Comparison of the median estimates for mean and cv or zero-inflation produced by the ZIGP and GP or ZIP models is shown in Figures 3.3 and 3.4. The median estimates for mean count were very similar between the two models for the GP and ZIP data, but small differences between the model estimates can be seen with the ZIGP data. The median estimate for cv was on average lower with the ZIGP model compared to the GP model for the GP data, but agreement was much poorer with the ZIGP data. There was good agreement for the median estimate for zero-inflation between ZIGP and ZIP models for the ZIP data, but not with the ZIGP data.

The empirical cumulative distribution function for the median and 95% CI estimates of cv produced by the ZIGP model for the ZIP data are shown in Figure 3.5. The lower 95% CI for cv was very close to the lower limit of the prior (below 0.0101) for 972 of the 974 datasets (99.8%), and below 0.1 for all datasets. The upper 95% confidence limit was below one for 788 (80.9%) of the datasets. Figure 3.6 shows the empirical CDF for the median and 95% CI estimates of zero-inflation produced by the ZIGP model for the GP data. The lower 95% CI was below 1% for 827 (84.0%) of the datasets, but as high as 40% for some datasets. Datasets resulting in a high lower 95% CI for zero-inflation tended to have a high simulated cv and low simulated mean count parameter (Figure 3.7). Use of Model 2.2 rather than Model 2.2a to analyse the same data increased the number of lower 95% CI estimates that were below

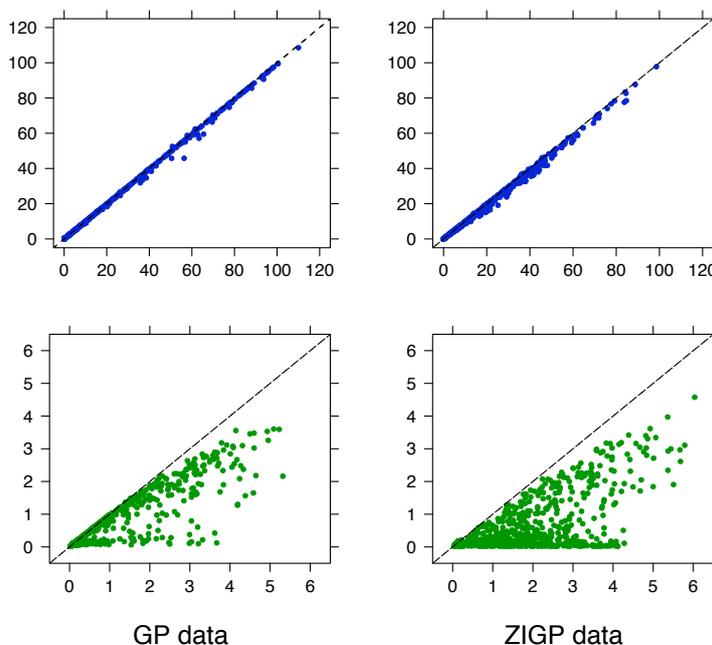


Figure 3.3: Comparison of median estimates produced by ZIGP model (y axis) and GP model (x axis) for GP and ZIGP simulated data - estimates for mean shown in blue and cv in green

Table 3.5: Comparison of the relative RMSE, bias, proportion of true values above and below 95% CI and mean relative 95% CI size returned by each combination of parameter and model/simulated dataset - bold entries highlight poor performance

Parameter	Statistic	GP model	ZIP model	ZIGP model
Mean	RMSE	0.36	0.30	0.33
	Bias	-0.03	0	-0.02
	Above CI	0.03	0.09	0.04
	Below CI	0.02	0.05	0.01
	CI size	0.56	0.43	0.54

Variability	RMSE	0.52		0.47
	Bias	-0.74		-0.31
	Above CI	0.04		0.14
	Below CI	0.01		0
	CI size	3.30		3.87

(a) Gamma-Poisson data				

Mean	RMSE	0.26	0.10	0.11
	Bias	-0.06	-0.02	-0.03
	Above CI	0.01	0.01	0.01
	Below CI	0.01	0	0
	CI size	0.79	0.57	0.61

Zero-inflation	RMSE		4.81	4.57
	Bias		-0.79	-0.73
	Above CI		0	0
	Below CI		0.01	0
	CI size		1.03	1.09

(b) Zero-inflated Poisson data				

Mean	RMSE	0.36	0.25	0.26
	Bias	-0.09	-0.02	-0.04
	Above CI	0.01	0.03	0.01
	Below CI	0.01	0.02	0.01
	CI size	0.98	0.61	0.75

Variability	RMSE	7.08		0.48
	Bias	-23.52		-0.69
	Above CI	0.01		0.06
	Below CI	0.62		0
	CI size	2.32		4.48

Zero-inflation	RMSE		11.51	9.64
	Bias		-2.22	-1.74
	Above CI		0	0
	Below CI		0.19	0.03
	CI size		0.84	1.10

(c) Zero-inflated gamma-Poisson data				

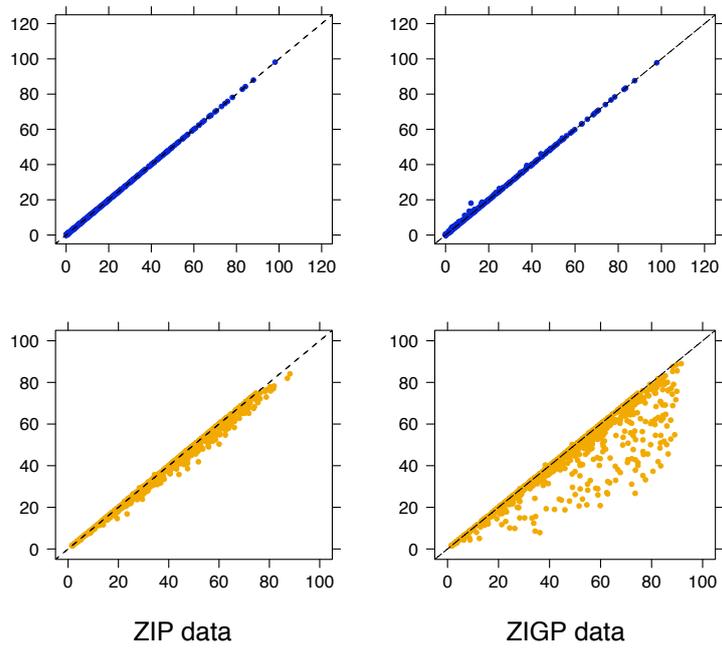


Figure 3.4: Comparison of median estimates produced by ZIGP model (y axis) and ZIP model (x axis) for ZIP and ZIGP simulated data - estimates for mean shown in blue and zero-inflation in orange

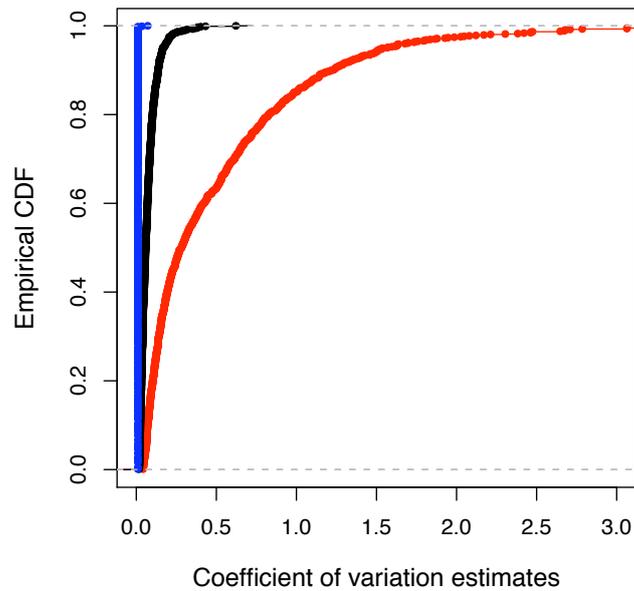


Figure 3.5: Empirical cumulative distribution function for the lower (blue) and upper (red) 95% CI and median (black) estimates of cv produced by the ZIGP model from analysis of the GP simulated data

1% to 952 (97.7%), and all but three datasets (99.7%) to below 5%. The upper 95% CI estimates were more similar between [Models 2.2a](#) and [2.2](#) with 337 (34.2%) and 344 (35.3%) of the estimates being below 5% for the respective models. Use of a ZILP model to analyse lognormal-Poisson data with the same parameters resulted in 928 (92.8%) of the lower 95% CI estimates below 1%.

3.3.4 Discussion

There was generally excellent agreement between the estimates of the ZIGP and corresponding simpler model when analysing the GP and ZIP data. The more complex ZIGP model produced all but identical median estimates, relative RMSE, mean relative bias and proportion outside the 95% confidence interval (95% CI) for both mean count and zero-inflation to the zero-inflated Poisson model when analysing zero-inflated Poisson data, indicating that the addition of the extra-Poisson variability parameter made no difference to the analysis in the absence of appreciable extra-Poisson variability in the data. The median estimates for mean count and cv between the ZIGP and GP models for the GP data also showed good agreement, although there was more disparity evident with cv . The bias and relative RMSE were similar between the models for both parameters, but a greater proportion of the 95% CI did not contain the true value for the ZIGP estimates. These datasets tended to have a higher estimate for zero-inflation than the datasets where the 95% CI contained the true

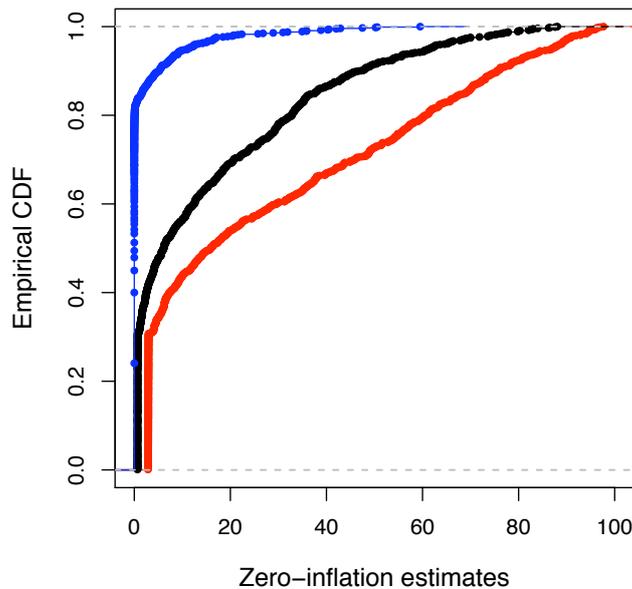


Figure 3.6: Empirical cumulative distribution function for the lower (blue) and upper (red) 95% CI and median (black) estimates of zero-inflation produced by the ZIGP model from analysis of the GP simulated data

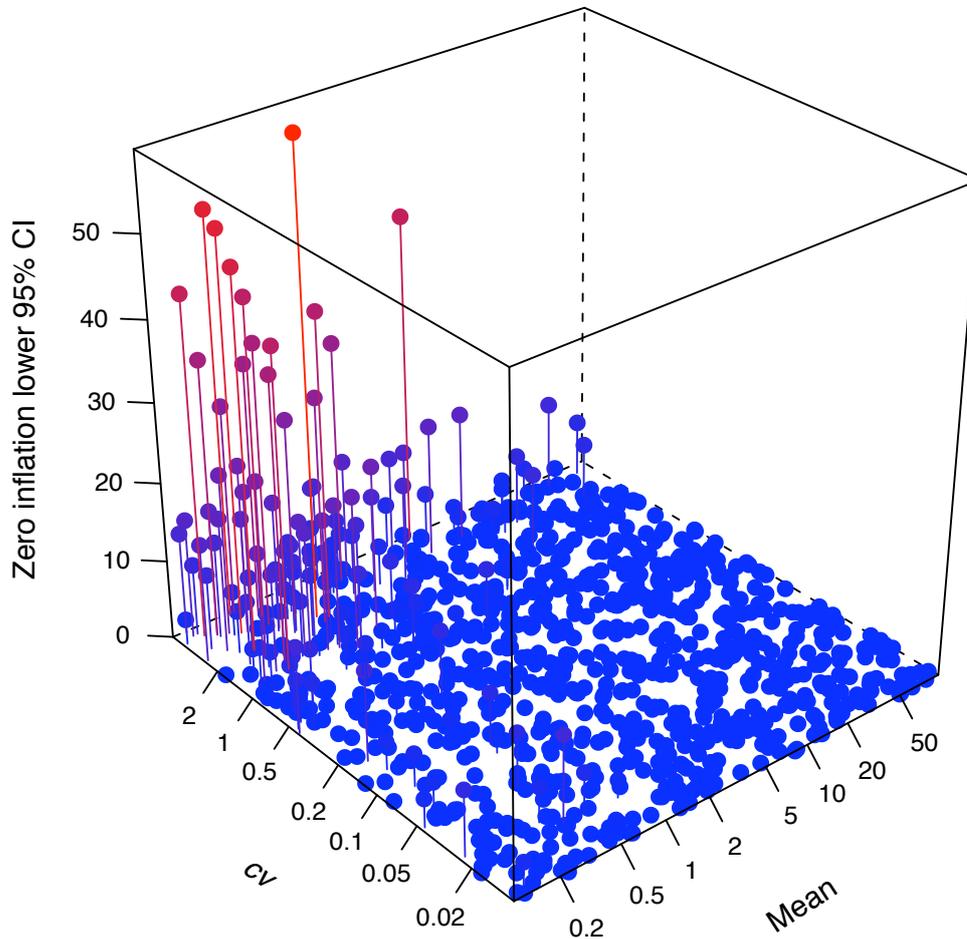


Figure 3.7: The relationship between the true mean, true cv and lower 95% CI produced by the ZIGP model from the analysis of the ZIP simulated data

value for cv , indicating the possibility that this tendency to under-estimate the cv may be as a result of the prior for cv tending towards smaller values, and accounting for extra zeros by increasing the zero-inflation parameter.

The ZIGP model produced very small lower 95% CI estimates for cv for nearly all ZIP datasets, preserving the possibility that the true cv was negligible, although the upper 95% CI was larger than 0.1 for the majority of the datasets indicating that the possibility of appreciable cv could not be ruled out. In these circumstances, the model is not reporting the presence of overdispersion, but rather the lack of proof for the absence, or existence, of over-dispersion in the data, which is an important distinction. The zero-inflation results for the GP datasets were similar, although the lower 95% CI for zero-inflation was relatively high

for some datasets. This was not seen when using a ZIGP specification with a prior tending towards larger values of cv , indicating that this was a result of the prior for cv tending towards low values of cv and consequently higher values of zero-inflation to explain the excess zeros observed.

In contrast to the high degree of agreement shown between the ZIGP and appropriate simpler models for the simpler forms of data, the gamma-Poisson and zero-inflated Poisson models produced results with large relative RMSE, bias and poor agreement with the ZIGP model for cv and zero-inflation respectively when analysing the ZIGP data. The large bias demonstrates that the ZIP model was accounting for over-dispersion by increasing the amount of zero-inflation, and similarly the GP model was accounting for zero-inflation by increasing the estimate for cv . This highlights the importance of using the appropriate or over-specified model to analyse data; based on these results use of a GP model to analyse real data with zero-inflation will detrimentally affect the analysis of the over-dispersion present in the data, to the extent that the 95% confidence interval will contain the true value less than 40% of the time. The relative RMSE and bias for mean count from the ZIGP data are relatively similar between the three models, with the ZIP model actually producing the most accurate estimates for mean. This may be as a result of the prior on cv with the GP and ZIGP models having an effect on the posterior estimate for mean count.

One disadvantage of using either the GP or ZIGP model appears to be an increased number of datasets where the model returned an error associated with sampling extreme values from the gamma distribution. The ZIP model also returned an error for some over-dispersed datasets, likely as a result of an extremely low probability that the observed counts could be represented by a single Poisson distribution, but successfully analysed all ZIP datasets. The ZIP model also tended to require less iterations before convergence as a result of a lower auto-correlation dependence.

The ZIGP model gave almost identical results to the ZIP model in the analysis of ZIP data, and superior results for other types of data, with the exception of the increased time to convergence and reduced number of datasets successfully returned. The ZIGP model also correctly identified the possibility that the true cv was small for nearly all of the ZIP datasets. These findings support the exclusive use of the zero-inflated gamma-Poisson model over the zero-inflated Poisson model when analysing data which is known to be zero-inflated and which has the possibility of being over-dispersed, ie. every faecal egg count dataset. The ZIGP model performed almost as well as the GP model in the analysis of GP data, and correctly identified the possibility that zero-inflation was minimal in the majority of these datasets. Use of a prior that tended towards larger values of cv improved the performance of the ZIGP model for the GP data, but may detrimentally affect the ability to analyse ZIP data appropriately. Analysis of datasets to conservatively quantify the amount of zero-inflation present would therefore be best using [Model 2.2](#) rather than [Model 2.2a](#) if the data were

gamma distributed. Use of the ZILP [Model 2.5](#) to analyse lognormal-Poisson data resulted in a lower 95% CI that was below 1% in 92.8% of the datasets, possibly as a result of a more moderate effective prior on cv than is used with [Models 2.2](#) or [2.2a](#).

The final conclusion from this analysis is that use of an under-specified distribution to analyse the data can lead to misleading inference. When analysing the ZIGP data, the GP model over-estimated the true cv in 62% of the datasets as a result of attempting to account for the number of zero counts. Similarly, the ZIP model overestimated zero-inflation in 20% of these datasets. The detrimental impact of using an under-specified model is therefore much greater than that of using an over-specified model, leading to the conclusion that use of the ZIGP model is associated with fewer type I errors than either the GP or ZIP models when the true distribution of data is unknown.

3.4 Analysis of empirical fit as a criterion for model selection

3.4.1 Introduction

Use of model fit as an aid to model selection in non-nested models was first suggested by [Cox and Hinkley \(1974\)](#), and has been fairly extensively used in many fields. However, the assumptions underlying the likelihood ratio statistic are only asymptotically valid, which is often not considered during the process of model selection. A model with more flexibility will always provide a better fit, so measures of fit such as AIC and BIC also take into account the number of degrees of freedom with which to evaluate the likelihood ratio. These model selection criteria are the basis for the use of the negative binomial to fit parasitology data, but the validity of using a likelihood based selection method and the impact of the selected distribution on the results obtained has not been demonstrated.

The lognormal and gamma distributions are both strictly positive continuous distributions, and so would therefore be expected to take a similar shape with equivalent parametrisation. Separation of these distributions on the basis of model fits would depend on finding this small difference. The differences between the distributions is exaggerated by increasing cv , but is unaffected by mean. This can be demonstrated by plotting the probability density function of each distribution with equivalent parametrisation, along with a truncated normal distribution for comparison ([Figure 3.8](#)). It is difficult to distinguish the lognormal and gamma distributions with a cv of 0.1 or 0.5. Adding Poisson ‘noise’ to these distributions to form gamma-Poisson and lognormal-Poisson distributions makes little difference when the mean is high, but makes it more difficult to distinguish the distributions with low means. In [Figure 3.9](#) this is demonstrated using the integrated probability mass function for each distribution and the truncated-normal-Poisson distributions. Even with a cv of 1 or 1.5, the

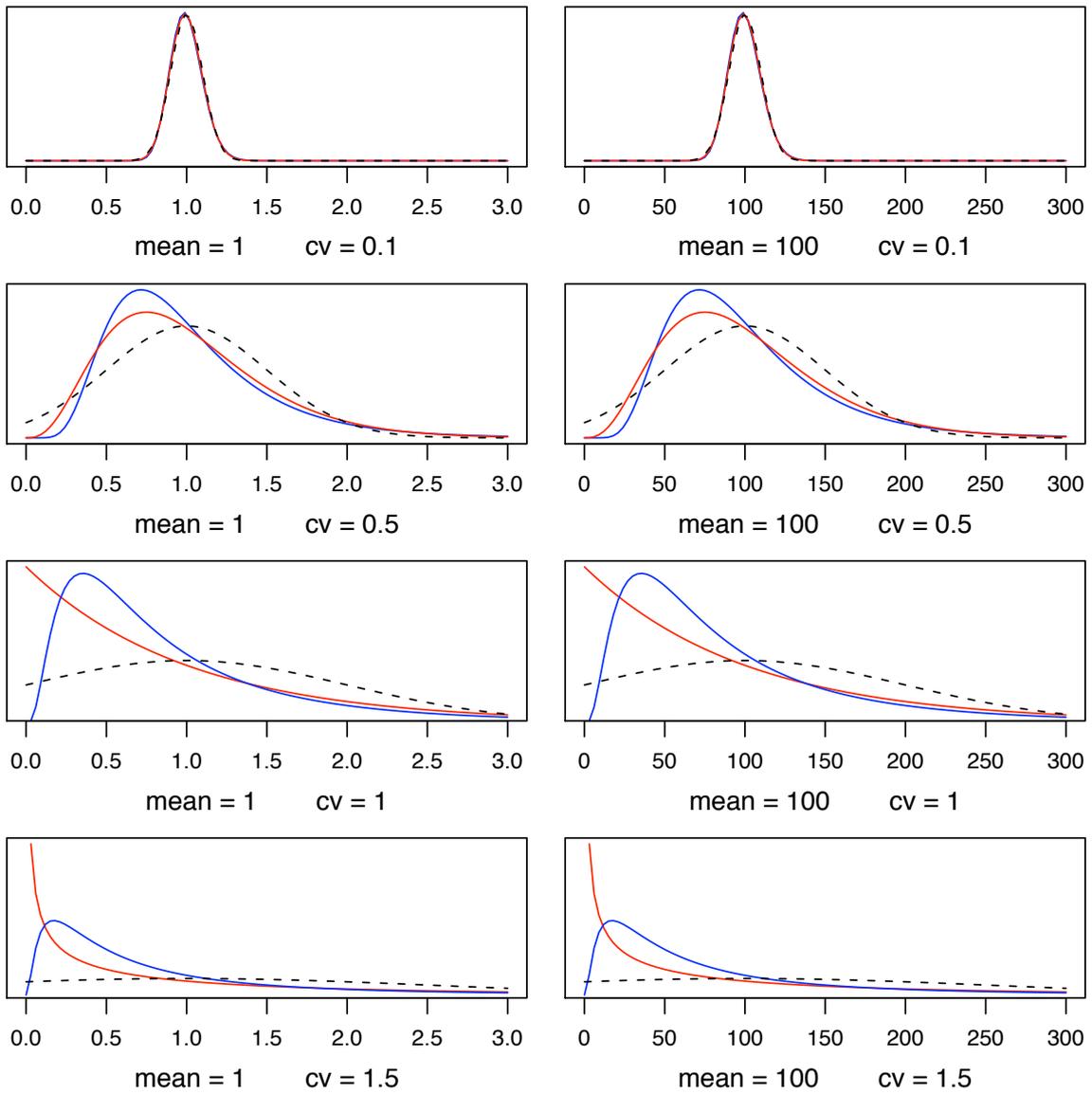


Figure 3.8: Comparison of an equivalent parametrisation of gamma (red), lognormal (blue) and truncated normal (black dashed) probability density functions

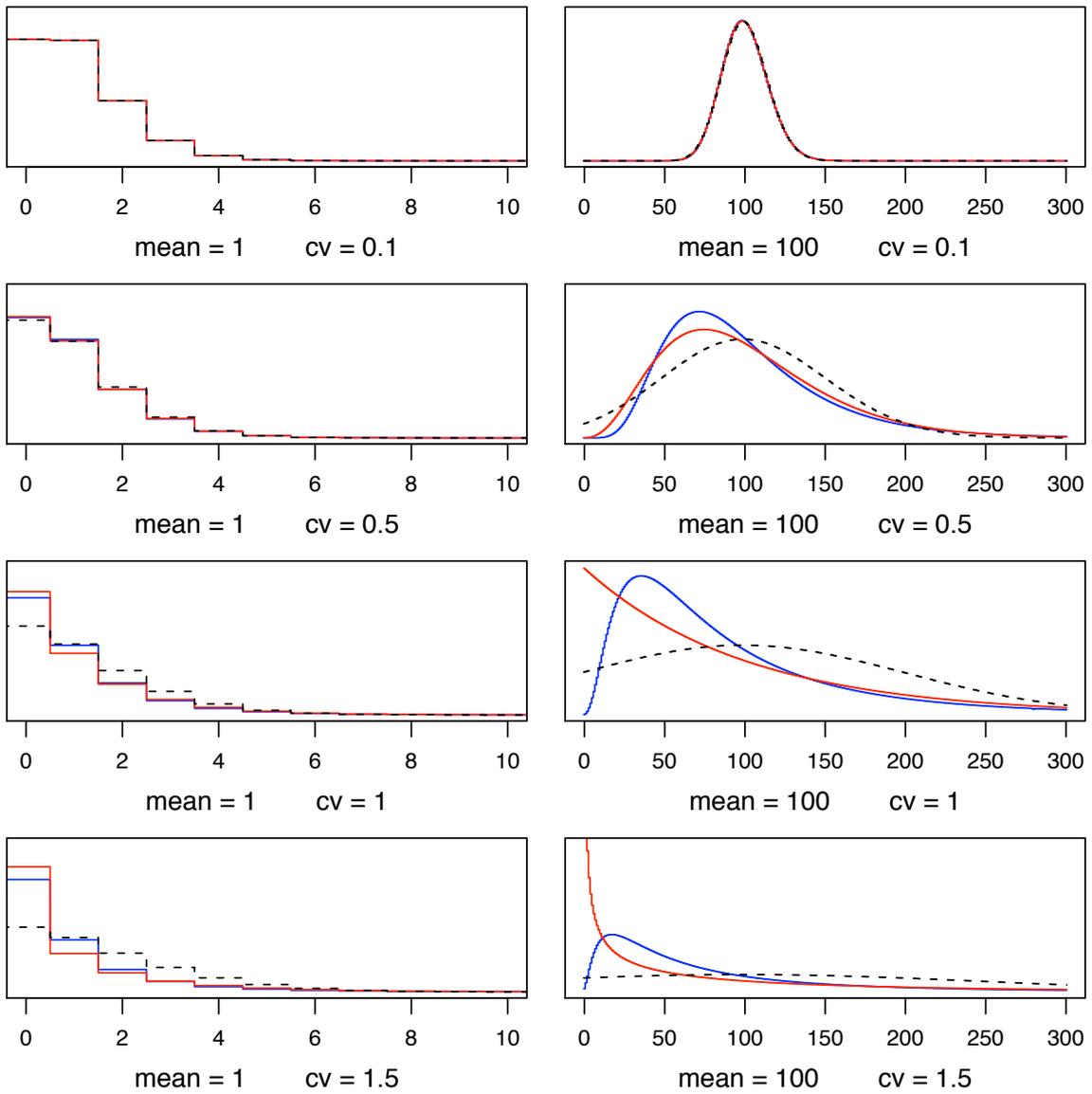


Figure 3.9: Comparison of an equivalent parametrisation of gamma (red), lognormal (blue) and truncated normal (black dashed) -Poisson probability mass functions (calculated using integration over lambda)

3.4 ANALYSIS OF EMPIRICAL FIT AS A CRITERION FOR MODEL SELECTION

only difference between the lognormal-Poisson and gamma-Poisson distributions is a slight increase in the number of one counts and decrease in the number of zero counts for the lognormal-Poisson compared to the gamma-Poisson, reflecting the exponential shape of the gamma distribution with these values of cv . It is therefore apparent that the ability to distinguish the true distribution of data based on empirical fit to these very similar distributions will depend on the mean and cv of the data, as well as the number of data points to fit. Due to the random nature of probability distributions there will always be a difference in penalised likelihood between gamma-Poisson and lognormal-Poisson fits to data, but based on these graphs alone it should be apparent that any difference in small datasets with low mean and cv is unlikely to be useful as a basis for model selection. A more detailed analysis of the effect of sample size, mean and cv on the proportion of likelihood ratios that would indicate the correct distribution follows.

3.4.2 Materials and methods

In order to test the usefulness of model selection based on likelihood ratios, data were simulated from gamma-Poisson and lognormal-Poisson distributions. Data were simulated with sample sizes drawn from the set $\{10, 20, 50, 100, 200, 500, 1000\}$, and with every combination of cv from the set $\{0.5, 1, 1.5\}$ and mean from the set $\{5, 20, 50\}$. For each combination, 100 datasets of each sample size were generated to give a total of 6,300 datasets. Each was analysed using a likelihood maximisation algorithm to find the maximum likelihood fit to the gamma-Poisson and lognormal-Poisson distributions. For each compound distribution, the mean and variability parameters were optimised to give the highest log likelihood fit to the data, integrating over all possible λ values of the Poisson part of the distribution. The maximum log-likelihood for each distribution was recorded for each dataset, and the true proportion of the 100 datasets where the maximum log-likelihoods indicated the distribution from which the dataset were simulated were calculated using a beta distribution.

This experiment was then repeated with zero-inflated models and data. The same combination of means and cv as used before were used to generate 100 datasets each with sample sizes drawn from the set $\{20, 40, 100, 200, 400, 1000, 2000\}$. A zero-inflation parameter of 50% was used for each dataset, so that each contained the same number of infected group observations as before. These data were analysed using likelihood maximisation to the ZIGP and ZILP distributions, to find the highest log-likelihood fit to the three parameters of mean, cv and zero-inflation.

3.4.3 Results

Figures 3.10 and 3.11 show the maximum log-likelihoods produced by the GP and LP models

3.4 ANALYSIS OF EMPIRICAL FIT AS A CRITERION FOR MODEL SELECTION

from analysis of the GP and LP data with sample sizes of 10 and 1000. For each graph, the maximum log-likelihood from the model corresponding to the distribution used to simulate the data is shown on the x-axis. Maximum log-likelihoods were similar between models for the same dataset, indicating a degree of similarity to the model fits. As would be expected, the log-likelihoods were lower with larger sample sizes, but also with larger means especially at the largest sample size. The difference in maximum log-likelihood between datasets with different values of cv was most obvious at the larger sample size, with datasets with a cv of 0.5 resulting in higher log likelihoods than datasets with a cv of 1.5, and the lowest log-likelihoods for the GP data being observed with a cv of 1.

The proportion of datasets where the maximum log-likelihood was greater using the distribution from which the data were simulated (y-axis) is shown in Figure 3.12. The proportion increased for both LP and GP data with increasing sample size, mean and cv . The proportions were on average lower for the LP data than the GP data, especially with the smallest sample sizes where the maximum log-likelihood was greater for the correct model as little as 50% of the time. The median estimate of the proportion of LP datasets that produced a better fit with the LP model was less than 95% for all combinations of mean and cv with sample sizes of less than 200, although the upper confidence limit was above 95% for sample sizes of 100 and means of 20 or 50 with a cv of 1.5. The proportion of GP datasets that produced a better fit to the GP model was on average slightly higher with a cv of 1 or above, but similar with a cv of 0.5.

The effect of zero-inflation on the median estimates for the proportion of maximum likelihoods

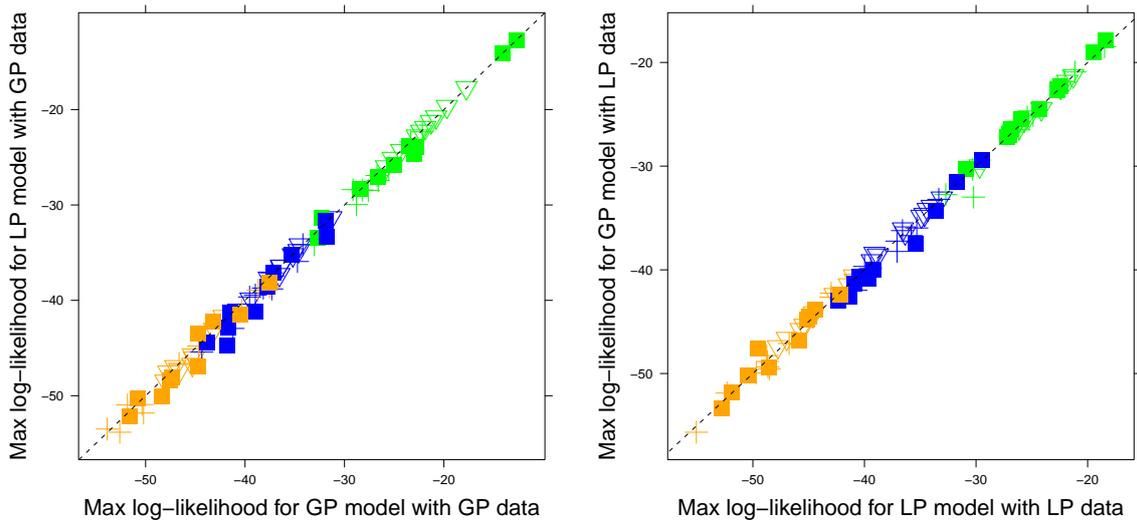


Figure 3.10: Comparison of the maximum log-likelihood produced by GP and LP models for GP and LP simulated data with a sample size of 10 - means of 5 (green), 20 (blue) and 50 (orange) with cv of 0.5 (triangles), 1 (crosses) and 1.5 (squares). A subset of the data, with 90 data-points only, is shown for clarity.

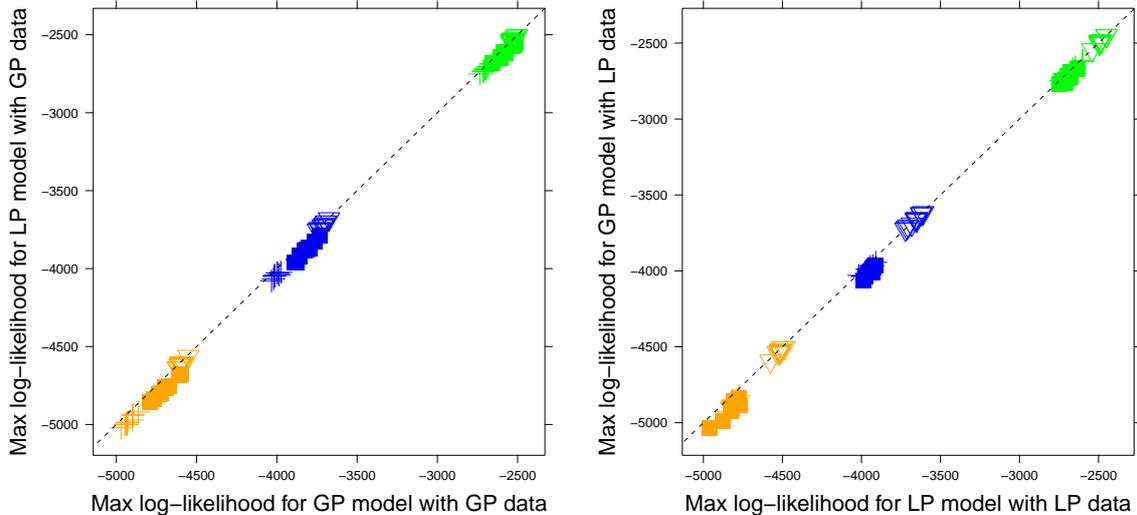


Figure 3.11: Comparison of the maximum log-likelihood produced by GP and LP models for GP and LP simulated data with a sample size of 1000 - means of 5 (green), 20 (blue) and 50 (orange) with cv of 0.5 (triangles), 1 (crosses) and 1.5 (squares). A subset of the data (90 data-points only) shown for clarity.

indicating the true distribution is shown in [Figure 3.13](#). The zero-inflated datasets produced a significantly smaller proportion ($p < 0.001$ and $p = 0.002$ for gamma and lognormal data respectively) than the datasets without zero-inflation. For three of the ZILP datasets with the smallest sample size, the upper 95% CI for the true proportion producing a better fit to the ZILP distribution was less than 0.5, indicating that the ZIGP model produces a higher maximum likelihood for these data than the ZILP model.

3.4.4 Discussion

The agreement between the maximum log-likelihoods produced by the GP and LP models was striking, and suggests that equivalently parametrised LP and GP distributions produce very similar fits to these kinds of data. The tendency for a better fit to data with a higher mean and cv of one was interesting, but of little practical implication because the effect was mirrored between the models. The proportion of the maximum likelihoods that were higher for the distribution from which the data were generated is a measure of the usefulness of using empirical fit as a criterion for model selection. The LP and GP models have the same number of parameters and degrees of freedom when analysing the same data, therefore the use of any penalised likelihood such as AIC or BIC would have given the same result as comparing the maximum likelihoods directly. The value of cv had a large impact on the usefulness of empirical fit, with datasets with smaller cv proving more difficult to correctly identify the underlying distribution. A sample size of at least 500 with the GP data and 1000 with the LP data was required to correctly identify the true distribution 95% of the time with a cv of

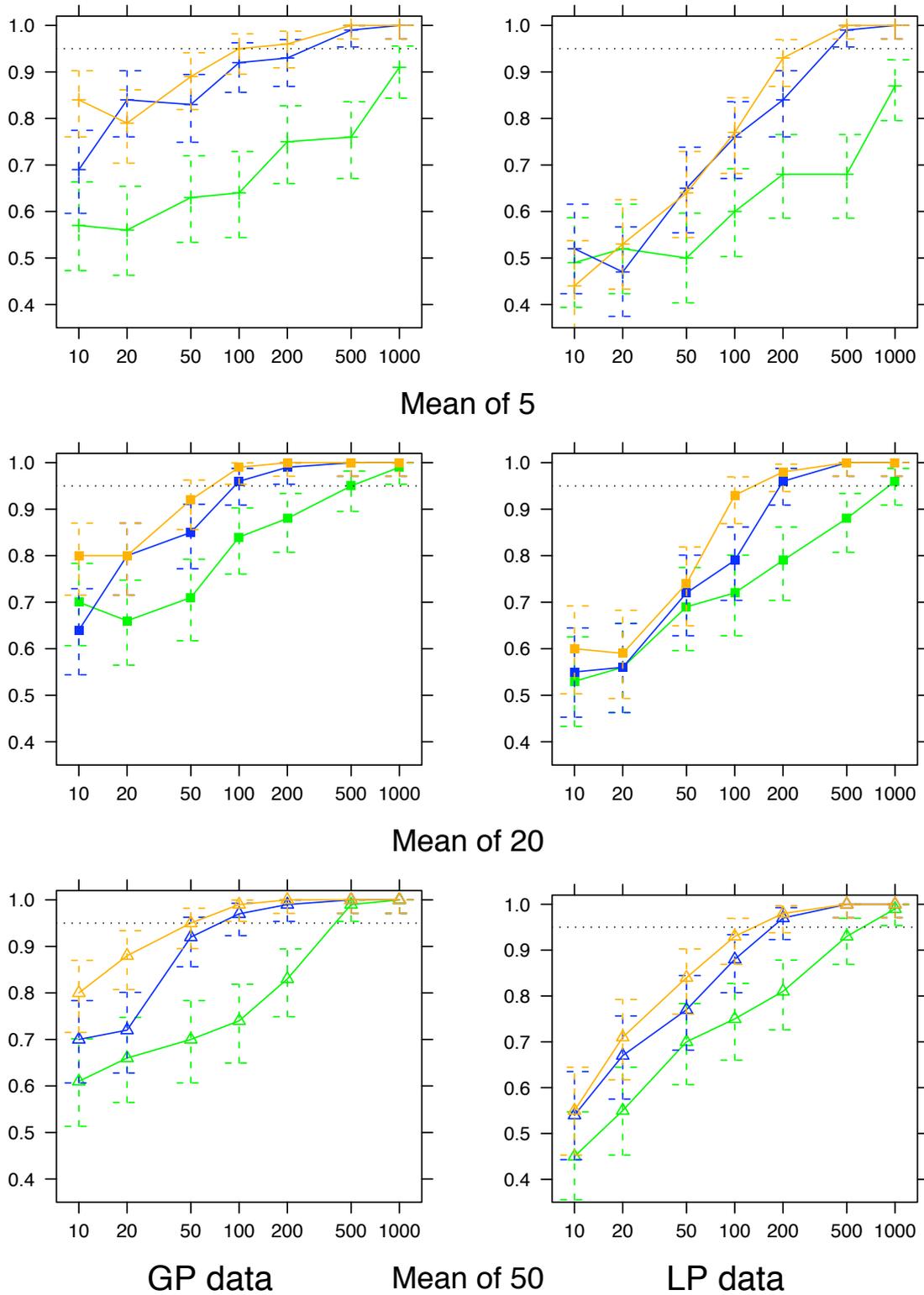


Figure 3.12: The relationship between the proportion of maximum likelihoods produced by GP and LP models that indicated a better fit to the distribution from which the data were simulated (y axis - 95% CI shown dotted) and the sample size of the data (x axis) - means of 5 (top), 20 (middle) and 50 (bottom) shown with cv of 0.5 (green), 1 (blue) and 1.5 (orange). GP data shown on left, LP data on right.

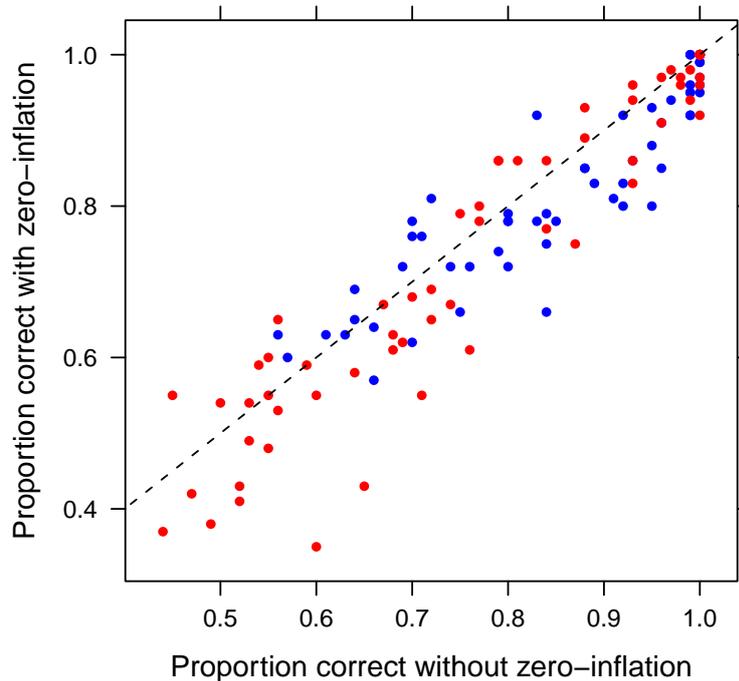


Figure 3.13: Comparison of the proportion of maximum likelihoods giving a better fit to the distribution from which the data were simulated with and without zero-inflation in the GP and LP simulated data and models - (ZI)GP data in blue, (ZI)LP data in red

0.5, and a sample size of more than 1000 was required when the mean was very small. The proportion correctly identifying the GP data was always higher than that for the LP data except with the smallest cv , possibly indicating a slight increase in flexibility to the gamma distribution with a cv of one or more due to the change in shape to exponential decay curve. A sample size of 100, or 50 with a large mean, was required to correctly identify GP data with a higher cv , but this rose to a sample size of 200, or 500 with a small mean, for the LP data. Repeating the experiment with the use of zero-inflated data and models yielded even lower proportions of datasets where empirical fit correctly identified the true distribution. The number of degrees of freedom in mixture models such as zero-inflated models is not defined, so that penalised likelihoods cannot normally be calculated. However, the number of degrees of freedom between the ZILP and ZIGP models would be the same if the estimate for zero-inflation is the same. The slight decrease in proportion of datasets correctly identified using the zero-inflated models may be as a result of a discrepancy in the estimate for zero-inflation slightly altering the degrees of freedom between the models.

Based on these results, it is apparent that model selection based on empirical fit of GP and LP models may be difficult. Sample sizes of more than the typical parasitology dataset are required to be sure of making the correct choice of underlying distribution more than 95%

of the time with single datasets. Crucially, it appears that datasets that are closer to having a lognormal distribution are more likely to fit a gamma distribution better than *vice versa*. Model selection on biological plausibility is therefore to be recommended in preference to model selection on the basis of empirical fit for these kinds of data, especially where there is a biological reason to prefer the lognormal over the gamma distribution. Alternatively, more sophisticated inferential techniques such as reversible jump MCMC that take into account the uncertainty in the true distribution of the data could be employed.

3.5 Comparative analysis of ZILP and ZIGP simulated data

3.5.1 Introduction

The results presented in [Section 3.4](#) suggest that it may be difficult to distinguish data that are closer in shape to a lognormal-Poisson compound distribution from data that are closer to a gamma-Poisson compound distribution, and that differentiation of their zero-inflated mixture equivalents may be even more difficult. This is likely to be a result of the asymptotic assumptions necessary for the use of likelihood ratio tests for non-nested models being flawed with these kinds of data. As a result, likelihood based model selection may result in the use of an incorrect model. A comparison of the results obtained using each model given both ZIGP and ZILP data is therefore warranted to quantify the resultant inaccuracies that are likely to be made in parameter inference.

3.5.2 Materials and methods

The 1000 parameter combinations of mean count, cv and zero-inflation discussed in [Section 3.2](#) were used to generate data from ZILP and ZIGP distributions. Sample sizes of 10, 100 and 1000 were used with both distributions to give a total of 6000 datasets. Each of these was analysed using a ZILP MCMC model and ZIGP MCMC model as discussed previously. The median and 95% confidence intervals produced by each model were compared to the simulation parameters and the estimates compared between models.

3.5.3 Results

The total number of datasets that were successfully returned for each type of data with each model are shown in [Table 3.6](#). The ZILP model successfully returned results for all datasets for sample sizes 10 and 100. The ZIGP model returned an error for some datasets, but failed

3.5 COMPARATIVE ANALYSIS OF ZILP AND ZIGP SIMULATED DATA

to converge within the maximum allowed one hour time period less frequently than the ZILP model for the sample size 1000 data. The ZIGP model returned fewer errors for the ZILP datasets than the ZIGP datasets at the larger sample sizes.

The proportion of 95% confidence estimates produced by each model that contained the simulation parameter for each combination of data type and sample size is given in [Table 3.7](#). The 95% CI produced by the ZILP model in analysis of the ZILP data contained the true value approximately 95% of the time in all cases, although the proportion for variability at the largest sample size was slightly lower. Similarly, the 95% CI produced by the ZIGP model in analysis of the ZIGP data contained the true value approximately 95% of the time in all cases, with the exception of the *cv* which had slightly lower confidence. In contrast, the true confidence of the notional 95% CI for each parameter in analysis of data generated using the contrasting distribution was markedly lower at the larger sample size for both models, for zero-inflation at sample size 100 using the ZILP model for ZIGP data, and for *cv* at all sample sizes using the ZIGP model for ZILP data. This was due to a tendency for the ZILP model to over-estimate the mean, *cv* and zero-inflation for ZIGP data, and for the ZIGP model to under-estimate the mean, *cv* and zero-inflation for ZILP data ([Figure 3.14](#)). This effect was

Table 3.6: Number of datasets returned by the ZILP and ZIGP models in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions

Data	ZIGP data		ZILP data	
Model	ZIGP	ZILP	ZIGP	ZILP
Total successfully analysed	778	830	813	867
Model returned an error	52	0	54	0
Failed to converge	0	0	0	0
Dataset all 0	170	170	133	133
(a) Sample size 10				
Total successfully analysed	936	996	968	999
Model returned an error	60	0	31	0
Failed to converge	0	0	0	0
Dataset all 0	4	4	1	1
(b) Sample size 100				
Total successfully analysed	829	762	855	795
Model returned an error	17	0	4	0
Failed to converge	154	238	141	205
Dataset all 0	0	0	0	0
(c) Sample size 1000				

Table 3.7: Proportion of 95% CI returned by the ZILP and ZIGP models that contained the true parameter value in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions - bold entries highlight notional 95% CI with markedly lower true confidence

Sample size	Parameter	ZIGP data		ZILP data	
		ZIGP	ZILP	ZIGP	ZILP
Sample size 10	Mean	0.970	0.989	0.970	0.984
	Variability	0.914	0.977	0.886	0.986
	Zero-inflation	0.986	0.970	0.994	0.996
Sample size 100	Mean	0.953	0.939	0.939	0.954
	Variability	0.929	0.920	0.810	0.958
	Zero-inflation	0.956	0.886	0.951	0.973
Sample size 1000	Mean	0.958	0.702	0.823	0.940
	Variability	0.933	0.777	0.712	0.921
	Zero-inflation	0.971	0.750	0.796	0.968

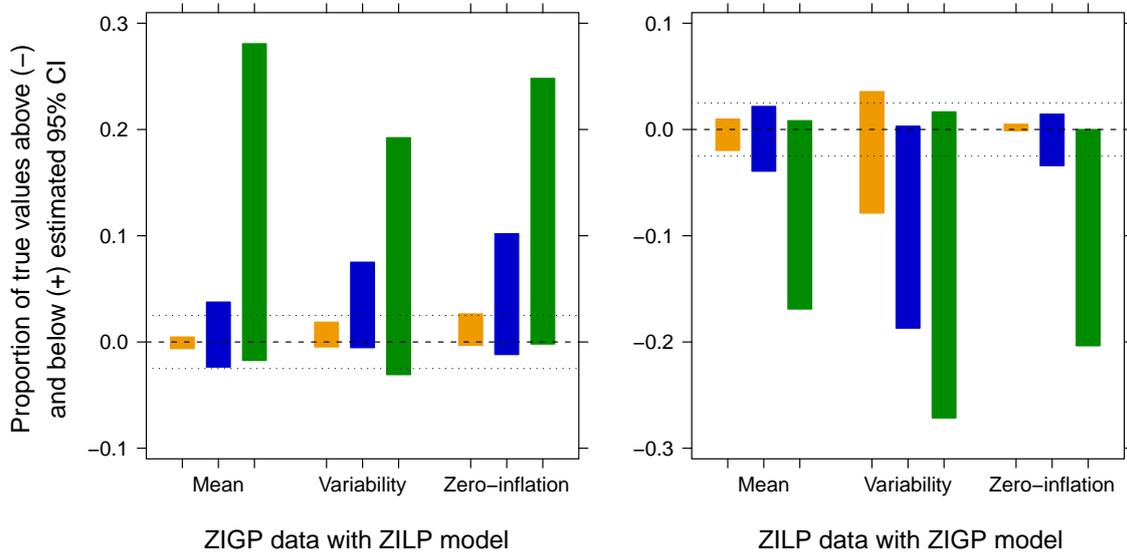


Figure 3.14: Proportion of true parameters that fell above (negative value) and below (positive value) the 95% CI returned by the ZILP and ZIGP models in analysis of data simulated with sample sizes of 10, 100 and 1000 using ZIGP and ZILP distributions - sample sizes of 10, 100 and 1000 are shown in orange, blue and green, respectively

most pronounced at the largest sample size, although the estimation of the cv parameter with the ZIGP model and the cv and zero-inflation parameters with the ZILP model were also affected at sample size 100.

In addition to the differences in 95% CI produced by the models, there were also discrepancies between the median estimates produced by the two models. The ZILP model produced on average larger estimates for cv than the ZIGP model for both data types, with the exception of a cluster of points with a sample size of ten and median estimate of between one and two using the ZILP model and between three and six using the ZIGP model (Figure 3.15). A similar pattern was less obvious with the median estimates of zero-inflation, although a slight tendency for the ZILP model to produce higher estimates than the ZIGP model with sample size 1000 ZILP data may be present (Figure 3.16). However, there was a tendency towards the ZILP model producing higher lower 95% CI estimates for zero-inflation than the ZIGP model, especially with data generated from a ZILP distribution with sample sizes of 100 and 1000 (Figure 3.17). Analysis of the same ZILP datasets yielded lower 95% CI of less than 5% with the ZIGP model and over 75% with the ZILP model. The median estimates produced for mean count were more similar, with the exception of some datasets that produced a large estimate for mean with the ZIGP model and a very low estimate for mean with the ZILP model, although a tendency towards higher estimates for some datasets with the ZILP model and ZIGP data is evident (Figure 3.18).

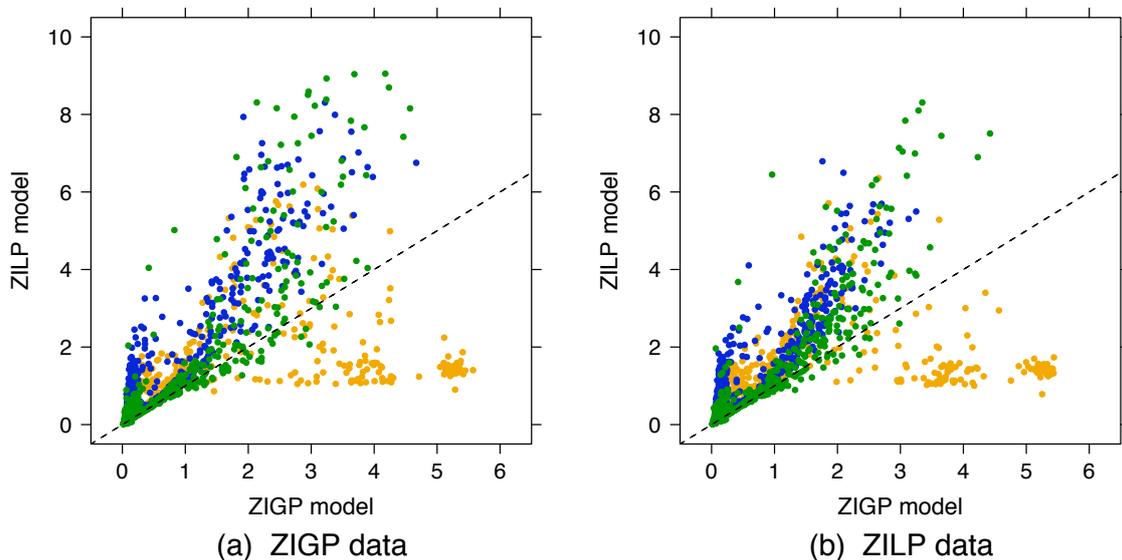


Figure 3.15: Comparison of median estimates for cv produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data - ZIGP data shown on left, ZILP data on right, sample sizes of 10, 100 and 1000 shown in orange, blue and green respectively

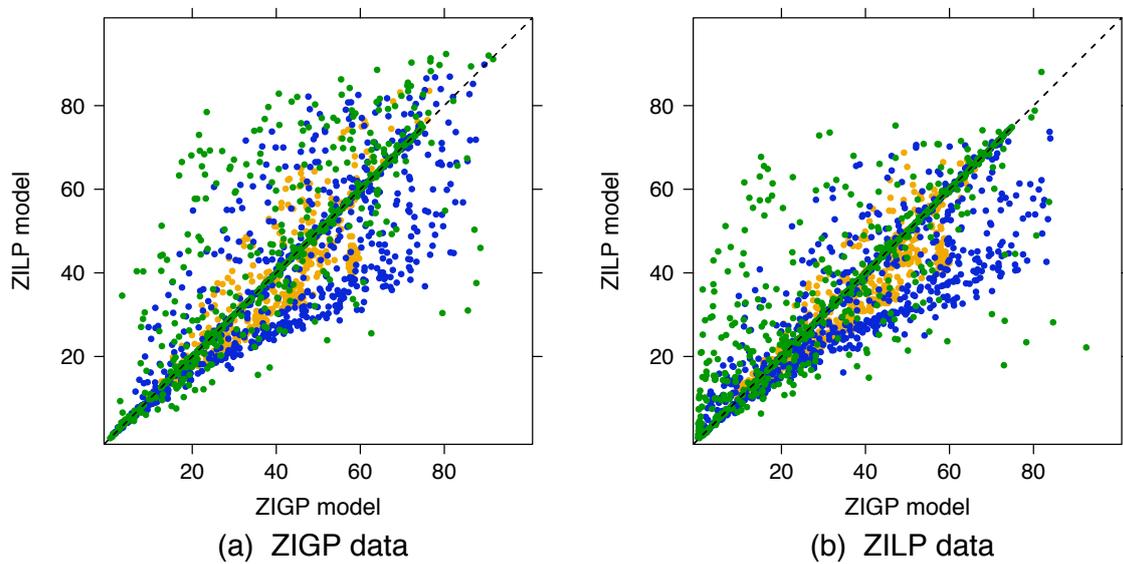


Figure 3.16: Comparison of median estimates for zero-inflation produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data - ZIGP data shown on left, ZILP data on right, sample sizes of 10, 100 and 1000 shown in orange, blue and green respectively

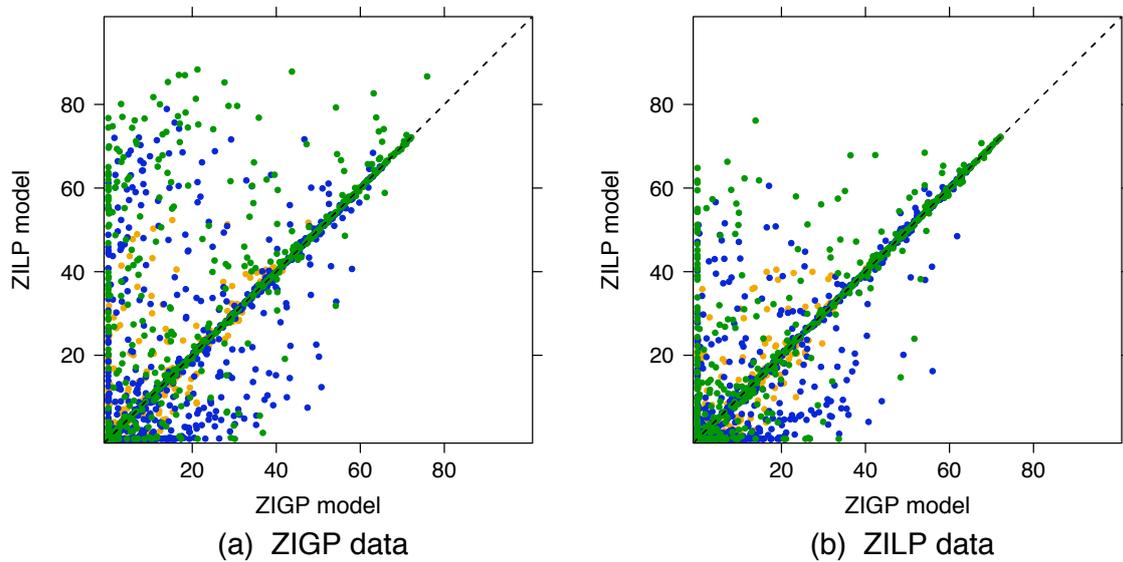


Figure 3.17: Comparison of lower 95% CI estimates for zero-inflation produced by the ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data - ZIGP data shown on left, ZILP data on right, sample sizes of 10, 100 and 1000 shown in orange, blue and green respectively

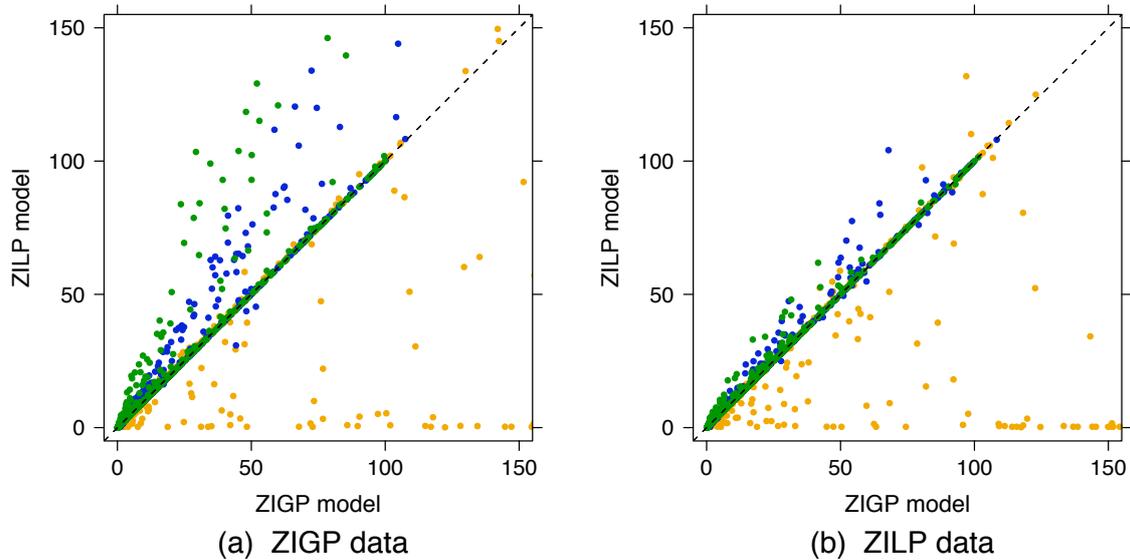


Figure 3.18: Comparison of median estimates for mean produced by ZILP and ZIGP models from analysis of ZIGP and ZILP simulated data - ZIGP data shown on left, ZILP data on right, sample sizes of 10, 100 and 1000 shown in orange, blue and green respectively

3.5.4 Discussion

These results demonstrate that use of ZILP and ZIGP models can provide very different inference from the same data. A large proportion of the 95% CI for the ZIGP datasets using the ZILP model and *vice versa* did not contain the simulation parameter for cv and zero-inflation, potentially leading to erroneous inference on the degree of aggregation and infection prevalence in the data. The performance of these 95% CI was worse with larger sample sizes as 95% CI tended to be large for small data, and the difference between the distributions is not so apparent with smaller sample sizes. The mean count parameter was subject to less error, probably as a result of the relatively large amount of information in the data for this parameter, although the larger estimates of zero-inflation from the ZILP model may be responsible for a slight increase in the estimate of mean using this model. The difference in median estimates is also worthy of note as this represents the central estimate of the parameter and is the most often quoted. The ZILP model consistently produced larger estimates of cv than the ZIGP model, except for the cluster of points with a sample size of ten with large estimates produced by ZIGP. These points are likely to represent instances where the ZIGP model produced large estimates for mean count and cv as a result of a paucity of information in the data, and also explain the few high median estimates for mean from the ZIGP model with very low estimates from the ZILP model at the same sample size. This phenomenon is explained more fully in [Section 2.7](#). As with the 95% CI, the differences between median estimates was more apparent at larger sample sizes.

These findings indicate that the variability parameter is most susceptible to choice of distribution, with the zero-inflation parameter being affected to a lesser extent and the mean count parameter less still. The practical implication of this is that for datasets of the sample size typically encountered in parasitology, the inference on mean should not be affected by the choice of distribution regardless of the true underlying distribution of the data. Where inference is made on the variability and zero-inflation, more attention should be paid to the distributional selection. The main difference between the gamma-Poisson and lognormal-Poisson distributions is the shape of the distribution with a cv of greater than one, when the gamma distribution takes on an exponential decay shape with a mode of approximately 0, or more correctly $\frac{1}{\infty}$ because the gamma distribution is strictly positive. By contrast, the lognormal distribution retains a skewed bell curve shape, and consequently has a higher mode than the gamma distribution with the same combination of mean and cv . The result of this is that the gamma-Poisson distribution would be expected to have a greater weight on zero, which can be seen in [Figure 3.9](#). A similar finding was made by [Joe and Zhu \(2005\)](#) when comparing the fit of the zero-inflated generalised Poisson and the zero-inflated gamma-Poisson, formulated as a zero-inflated negative binomial. This appears to result in a lower estimate for cv and zero-inflation for the ZIGP model than the ZILP model, as the ZILP model requires a greater degree of over-dispersion and/or zero-inflation to explain the same number of zero counts.

3.6 Discussion

In this chapter, several factors relating to the analysis of count data such as parasite burden and FEC data have been examined. Due to the difficulty inherent in fitting mixture models such as the ZIGP, and the skewed distributions of parameter estimates that typically result from fitting the model to data similar to that likely to be encountered in parasitology datasets, parameter estimation by simple maximum likelihood methods appears to be insufficient. The disadvantage of MCMC is the increased computational effort required to implement the technique ([Gilks et al., 1998](#)), although automated tools for data analysis and increasing availability of faster computers make this less important. Another disadvantage of using the gamma distribution is the proportion of datasets that returned an error, which appeared to be associated with extreme values sampled from the gamma distribution. Further detailed analysis of the model output identified that the errors occurred when very small values were sampled, which tended to happen more frequently when the posterior for cv was high. Truncating the gamma distribution with a lower bound of 10^{-200} stops this problem occurring, and should not affect the posterior estimates because 10^{-200} is less than the 1st percentile for the most extreme value of cv allowed under the prior distributions used. However, even taking into account these errors the MCMC technique still produced more useful analyses than the ML technique. More computationally intensive parameter estimation methods such

as MCMC should therefore be used to avoid incorrect confidence interval estimation of the variability and zero-inflation parameters particularly. The issue of model selection was also addressed, both in terms of the selection of nested models including a mixture model, and the selection of non-nested models. Although likelihood ratio tests are frequently used to select nested models based on the additional number of parameters (Johnson and Omland, 2004), the application of these tests to mixture models such as the ZIGP is not straightforward (Millar, 2009). Based on the analyses presented, the over-specified model appeared to produce equivalent results to the zero-inflated Poisson distribution and almost equivalent results to the gamma-Poisson distribution for the analysis of data generated from the simpler distributions. In contrast, neither the ZIP or GP models produced accurate parameter estimates in analysis of the ZIGP data. The ZIGP model should therefore be used exclusively with data that are suspected to contain some zero-inflation, and use of a prior tending towards larger values of cv , such as the one implemented in Model 2.2, should provide reliable 95% CI on the true degree of zero-inflation in the data, provided that the observations in the infected animals truly follow a gamma-Poisson distribution.

Comparative analysis of the maximum likelihood fits of non-nested models identified a potential problem with the use of empirical fit as the basis of model selection. The correct model was identified as little as 60% of the time when the data were simulated using a GP distribution, and as little as 50% of the time when the data were simulated using an LP distribution. These figures did not improve to 95% until the sample size was increased to 100 with the GP data and 200 with the LP data, depending on the mean and cv . The larger sample size required when the data is LP distributed reflects the fact that the maximum likelihoods were on average slightly higher for the GP than the LP model, possibly indicating that the gamma distribution is more flexible than the lognormal for these combinations of parameters. Comparison of the maximum likelihood fits of the zero-inflated equivalents produced even less satisfactory results, despite the fact that the sample size was doubled to account for the 50% zero-inflation parameter used. This is possibly as a result of a slight difference between the effective number of parameters between the two mixture distributions, reflecting the ability of the gamma-Poisson distribution to account for more zero counts than a similarly parametrised lognormal-Poisson distribution. This finding highlights the difficulties of model selection on the basis of penalised likelihoods using mixture distributions. On the basis of this, model selection between the ZILP and ZIGP distributions based on likelihood ratio tests should be considered unreliable for the majority of FEC data because the sample size, taking into account any zero-inflation present, is not likely to be sufficient. When viewed alongside the results presented in Section 3.5, the potential for misleading inference becomes apparent. If data are truly ZIGP distributed and a ZILP model is used for analysis, the true cv and zero-inflation present in the data is likely to be over-estimated. Conversely, if the data are truly ZILP distributed and a ZIGP model is used for analysis, the true cv and zero-inflation present in the data is likely to be under-estimated. Given the tendency for higher maximum likelihoods with the GP compared to LP distributions, and the current almost ubiquitous

use of the negative binomial in parasitology (May and Anderson, 1978; Wilson and Grenfell, 1997; Shaw et al., 1998), the true cv and zero-inflation present in the data are therefore likely to be under-estimated in existing parasitological data analyses. In addition, it would also be expected that comparison of GP and LP distribution likelihoods would indicate a superior fit to the GP distribution for a reasonable proportion of truly LP distributed data because typical parasitological datasets rarely exceed 100 observations.

In this chapter, the practical implications of various distributional assumptions have been assessed using simulated data with parameters likely to be observed in FEC data. In the following chapter, data obtained from sheep and horses are analysed using the same methods as used here to compare the inference made under each distributional assumption, and the results used to infer the possible true underlying distribution of the data.

CHAPTER 4

Analysis of field faecal egg count data

Analysis of field faecal egg count data

4.1 Introduction

Methods of analysing parasite count and FEC data using Bayesian MCMC methods have been developed and validated using simulated data, and the potential consequences of making various distributional assumptions have been explored. In the work reported in this chapter, the models are applied to field parasitology data in an attempt to demonstrate the use of the MCMC method in a real world setting, and to examine the underlying distribution of such data.

4.1.1 Nematodirus infections of sheep

Nematodirus battus is one of the most pathogenic organisms that infect sheep. Infected lambs develop acute enteritis with watery diarrhoea accompanied by inappetence and weight loss, and will die if the disease is not controlled by anthelmintic treatment ([Armour and Coop, 1991](#)). The life-history of this species is unusual compared with other nematodes in that the larvae develop to the infective third stage within the egg, and these eggs only hatch when the temperature exceeds 10 °C following a cold spell. It is this temperate control that causes simultaneous hatching of large numbers of infective larvae, which poses a severe disease risk. Historically, disease occurs in late spring in those years with suitable weather, although more recently, warmer and more variable weather suggests that outbreaks could also occur in late autumn.

Despite its clinical importance, *N. battus* has received surprisingly little attention in terms of veterinary parasitological research. In particular, the distribution of parasites and egg output among hosts following natural infection is unknown, although the distribution of *Nematodirus* spp. (predominantly *Nematodirus spathiger*) in 104 3-month old Australian Merino lambs was similar to a negative binomial distribution ([Barger, 1985](#)). Mixed model analyses indicated that the heritability of faecal egg count following natural infection with *Nematodirus* spp. was similar to the heritability of egg counts due to other strongyle species in

New Zealand Coopworth and Romney sheep but higher for *Nematodirus* spp. in Scottish Blackface sheep (Bishop et al., 2004). In all cases there were strong positive genetic correlations between *Nematodirus* spp. egg counts and those of other nematode species, suggesting that similar mechanisms underlay either resistance to or control of all sheep nematodes, and that sheep could be simultaneously selected for resistance to *Nematodirus* spp. and other species. Such genetic analyses depend on the use of an appropriate statistical distribution to describe the data, so comparative analysis of *N. battus* and *Nematodirus* spp. FEC and worm burden data using both ZILP and ZIGP models would therefore be of benefit.

4.1.2 Nematodes of the domestic equid

The clinically most significant nematode parasites of horses in the western hemisphere based on their prevalence and pathogenicity are those of the super-family Strongyloidea (Warnick, 1992; Love, 1995; Osterman Lind et al., 1999). The life cycle of equine strongyles involves the production of eggs from sexually mature female worms fertilised by male worms inside a final host, followed by an obligate stage of development of at least two weeks during favourable conditions in the external environment (Urquhart et al., 1996). The strongyle eggs undergo hatching to L₁ followed by a moult to L₂ and finally L₃ larvae, at which point they are infective via ingestion by the final equid host (Urquhart et al., 1996). The location and speed of development of the ingested L₃ larvae to adults in the large intestinal lumen of the host is dependent on the genera, as outlined below.

The genus *Strongylus* contains three species of importance to equine parasitology; *S. vulgaris*, *S. edentatus* and *S. equinus*. All are relatively large adult worms of approximately 2.5 cm in length, and all undergo migration in the final host during the development from L₃ larvae to adults (Urquhart et al., 1996). The primary pathogenesis of these large strongyles is as a result of the adults' feeding habits in the large intestine: the adults have large buccal capsules which can cause significant damage to blood vessels during the process of feeding on the mucosal surface. This causes weight loss and ill-thrift as part of the 'mixed strongyle infection' syndrome (Love, 1995). The migrating larvae of *Strongylus vulgaris* are unique among the strongyles in that their predilection for development within the arterial blood supply of the gut, especially the cranial mesenteric artery, results in a very different type of pathology. Damage to the endothelium of the blood vessels caused by the developing larvae leads to thrombus formation and marked inflammation of the arterial wall, and in some cases aneurysms accompanied by dilatation of the arteries can occur (Urquhart et al., 1996). The resultant reduction of the blood supply to areas of gut can result in the development of intestinal infarction and colic, with potentially life-threatening consequences (Love, 1995).

The small strongyles or cyathostomins encompass worms of more than one genera and over 40 species (Lichtenfels, 1975), all of which share behavioural and morphological characteristics.

All are non-migratory strongyles, and develop from the ingested L₃ larvae to L₄ larvae in the large intestinal mucosa. They then emerge into the gut lumen to complete their development into adults (Urquhart et al., 1996). This emergence of L₄ larvae can cause significant damage to the mucosa if many larvae emerge ‘en mass’ - a clinical syndrome known as larval cyathostomiasis (Uhlinger, 1991; Mercier et al., 2001). The L₃ larvae can also undergo a period of hypobiosis, typically of several months and sometimes several years, before emerging from the mucosa (Love, 1995; Duncan et al., 2002). During the period of hypobiosis larvae are naturally less susceptible to anthelmintics. This behaviour is therefore of importance to the control of the parasite (Love, 1995; Mercier et al., 2001).

Cyathostomins are relatively small worms of only around 1cm in length, but are often present in large numbers (Uhlinger, 1991). They can contribute to the mixed strongyle infection syndrome, and have also been implicated as a major cause of colic (Uhlinger, 1990; Barrett et al., 2004). However, many clinically healthy horses also carry cyathostomins (Uhlinger, 1993; Love, 1995), and represent an important source of pasture contamination.

The remaining strongyle species of relevance to equine medicine are those of the genus *Triodontophorus* (Lichtenfels, 1975). These non-migratory large strongyles are relatively rare, and are similar to the other large strongyles in that the adults’ feeding habits contribute to the deleterious effects of a mixed strongyle infection (Urquhart et al., 1996).

Before the use of anthelmintics became a widespread feature of horse management, large strongyles were highly prevalent in the developed world (Duncan, 1975; Reinemeyer et al., 1984). The extensive use of anthelmintics has historically been primarily to control these worms, not least to prevent ischaemic colic associated with *Strongylus vulgaris*, and the simultaneous control of the small strongyles was little more than coincidence. More recently, the prevalence of the large strongyles has become markedly reduced (Osterman Lind et al., 1999; Young et al., 1999), while the cyathostomins have surpassed them both in prevalence (Reinemeyer et al., 1984, 1986; Herd, 1986; Uhlinger, 1991), and in clinical significance (Love and Duncan, 1992; Uhlinger, 1993; Love, 1995). The small strongyles are now considered to be the principle parasitic pathogens of horses (Herd, 1990; Love, 1992, 1995; Love et al., 1999). Accordingly, anthelmintic dosing and worm control in horses in the developed world is now almost always primarily directed against these parasites (Love, 1995; Nielsen et al., 2006b).

4.2 Sheep FEC data

4.2.1 Introduction

The epidemiology of *N. battus* infection in sheep is primarily driven by infection of young animals, with a large seasonal challenge of simultaneously hatching larvae and rapid development of age-related immunity (Urquhart et al., 1996; Winter, 2002). This somewhat unusual life cycle might be expected to result in an altered distribution of FEC and worm burdens between animals compared to other nematodes, although the distribution of *N. battus* in lambs has not been well studied beyond the observation that zero count observations are relatively common. However, it has not been established if these zero counts are a result of the heavy skewness of the distribution of positive counts, or if a proportion of the zero count observations are the result of bi-modality in the distribution of parasites between hosts. As such, comparative analysis of *N. battus* FEC and worm burden data using both ZILP and ZIGP models would be of benefit.

4.2.2 Materials and methods

FEC and necropsy worm burden counts were obtained from straight-bred Scottish Blackface sheep from a commercial upland farm in southwest Strathclyde. All husbandry procedures followed usual commercial practise and have been described previously (Stear et al., 1998). All lambs were kept on the same field after weaning at three or four months of age until the end of the study. To maintain health and productivity, all lambs were given the anthelmintic albendazole sulphoxide at the dose rate recommended by the manufacturer every 28 days until six or seven weeks before necropsy. Cohorts of 200 lambs were studied for five consecutive years to avoid bias due to unusually heavy or light infections in specific years. The lambs were the offspring of 39 sires and 496 dams.

Of the 1000 lambs, 530 were examined post-mortem at 6.5 months of age, in late October and early November each year. Standard procedures were used to estimate the concentration of nematode eggs per gram of faeces and to enumerate adult nematodes in the abomasum and small intestine (Armour et al., 1966; Stear et al., 1998). *N. battus* eggs were recorded separately from Nematodirus spp. eggs, although adult *N. battus* adult worms were not recorded separately from other Nematodirus spp.. The minimum egg detection threshold used for the McMasters technique was $\frac{1}{100}$ EPG in the first year, $\frac{1}{50}$ EPG in the second year, and $\frac{1}{25}$ EPG in the remaining years.

The data were then analysed as described previously using the ZILP [Model 2.5](#) along with the ZIGP [Models 2.2a](#) and [2.2](#) to ensure inference made using the ZIGP model was robust

to the choice of prior for cv . Datasets with all zero observations were excluded from the analysis. Each dataset was allowed to run for a maximum of five hours in order to achieve convergence.

4.2.3 Results

Shown in [Figure 4.1](#) is the agreement between the median estimates of mean egg count in eggs per gram (EPG), cv and zero-inflation produced by the ZILP and ZIGP model with log-uniform prior. There was good agreement for mean count, although the ZIGP estimates for some of the lowest values for the *Nematodirus* spp. datasets were slightly higher than those produced by the ZILP model. Agreement for cv and zero-inflation was less good. The ZILP median estimates of cv were higher for all *N. battus* datasets, although five of the *Nematodirus* spp. datasets had a much higher estimate with the ZIGP model. These datasets had relatively few observed eggs, indicating that the high estimate of cv and mean count produced by the ZIGP model for these datasets is likely to be as a result of the prior for mean count and lack of information in the data as explained in [Section 2.7](#). The ZIGP model produced higher estimates of zero-inflation for the *Nematodirus* spp. datasets, but more similar estimates to the ZILP model for the *N. battus* datasets. The agreement between the ZILP median estimates and those produced using the ZIGP model with uniform prior on cv is shown in [Figure 4.2](#). The agreement line is similar for mean count, although the datasets with a small number of observed counts had a much higher mean using the second ZIGP model. The median estimate for cv was slightly higher for all datasets using the second ZIGP model, but the estimate for the five *Nematodirus* spp. datasets with few observed eggs was markedly increased.

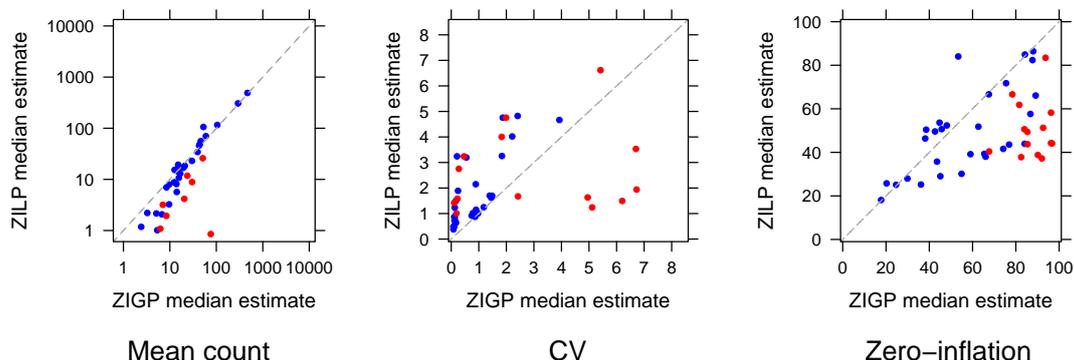


Figure 4.1: Comparison of the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with log-uniform prior on cv , from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - *N. battus* datasets shown in blue, *Nematodirus* spp. in red.

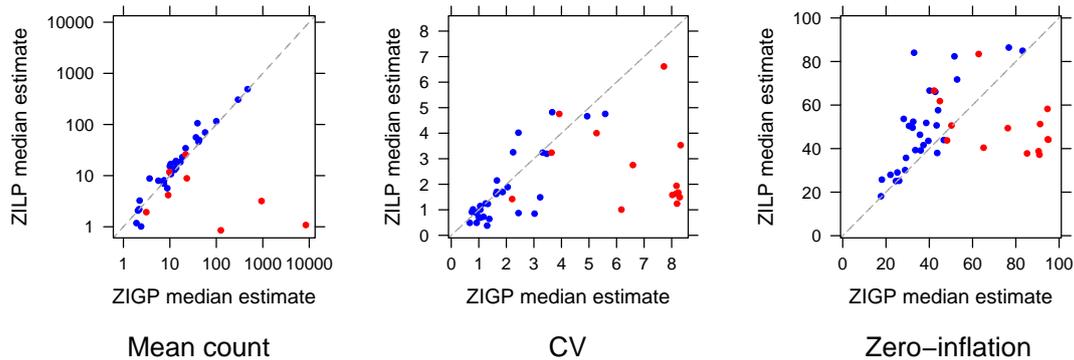


Figure 4.2: Comparison of the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with uniform prior on cv , from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - *N. battus* datasets shown in blue, *Nematodirus* spp. in red.

The relationship between median estimates of mean count, cv and zero-inflation for the ZILP model and two ZIGP models is shown in Figures 4.3 to 4.5. There was no significant relationship between any of the estimates using the ZILP model, although there was a relationship between mean and cv , and cv and zero-inflation for both of the ZIGP models. These relationships were stronger for the model with the uniform prior on cv , and there was also a relationship between mean and zero-inflation using this model.

The 95% CI estimates for zero-inflation produced by each of the models are shown in Tables 4.1a and 4.1b. The lower 95% CI produced by the ZILP model was above 10% for 15 of the 29 *N. battus* datasets, two of the 14 *Nematodirus* spp. datasets and all four worm burden datasets. The ZIGP model with log-uniform prior on cv produced a lower 95% CI estimate for zero-inflation of greater than 10% for 18 of the *N. battus* datasets, although only four of these were also above 10% using the ZIGP model with uniform prior on cv indicating that the lower 95% CI estimate for zero-inflation was quite dependent on the model and prior distribution choice for these data. All but one of the lower 95% CI for *Nematodirus* spp. datasets produced by the ZIGP model with log-uniform prior on cv , and all but four of those produced using the uniform prior on cv , were above 10%. Five of these had a higher lower 95% CI estimate using the model with uniform prior on cv . The lower 95% CI for zero-inflation was above 10% for all adult worm burden datasets using the ZILP model and ZIGP model with log-uniform prior on cv , and all but one dataset using the uniform prior on cv . There was evidence that zero-inflation was above 10% using all three models for the worm burden data in 1992, 1993 and 1995, *Nematodirus* spp. data from post-mortem 1994 only, and *N. battus* data from June 1994, May 1995, June 1995 and June 1996. Of these, the lower 95% CI estimates were highest for the May 1995 data, a histogram of which is given in Figure 4.6.

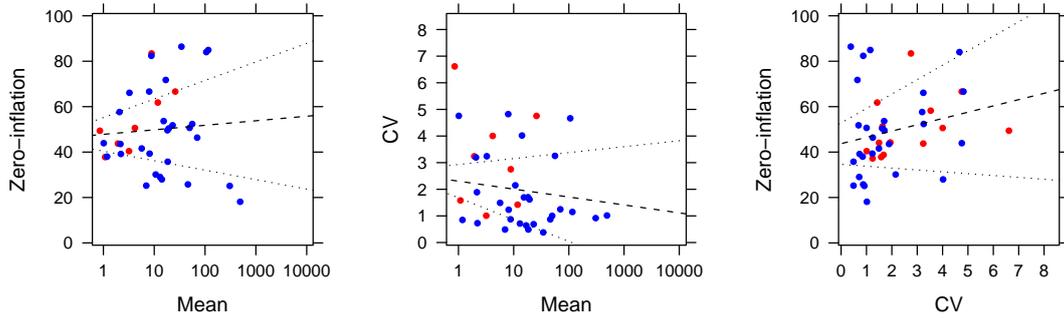


Figure 4.3: The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZILP model from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - Line of best fit shown dashed, 95% CI dotted. *N. battus* datasets shown in blue, *Nematodirus* spp. in red.

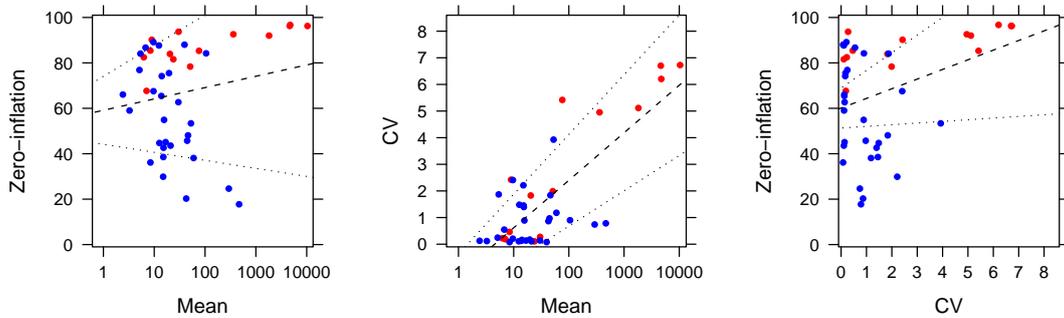


Figure 4.4: The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZIGP model, with log-uniform prior on cv , from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - Line of best fit shown dashed, 95% CI dotted. *N. battus* datasets shown in blue, *Nematodirus* spp. in red.

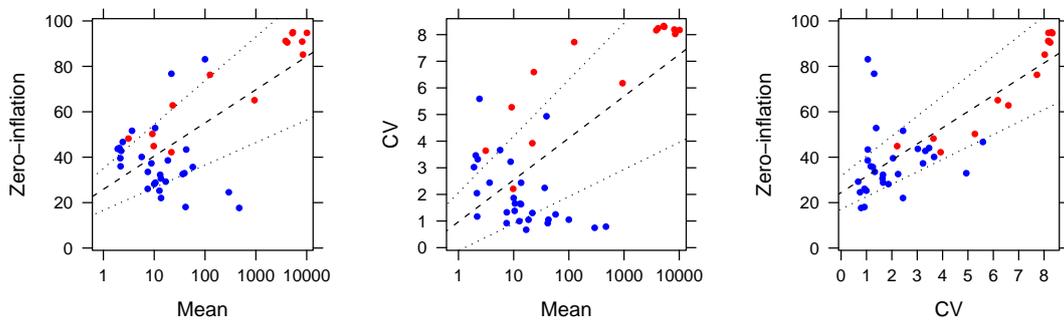


Figure 4.5: The relationship between the median estimates of mean FEC (EPG), cv and zero-inflation produced by the ZIGP model, with uniform prior on cv , from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - Line of best fit shown dashed, 95% CI dotted. *N. battus* datasets shown in blue, *Nematodirus* spp. in red.

Table 4.1a: The lower and upper 95% CI estimates of zero-inflation produced for each dataset by the ZILP model and ZIGP model, with log-uniform prior (ZIGP model 1) and uniform prior (ZIGP model 2) on cv , from analysis of the distribution of *N. battus* and *Nematodirus* spp. between five cohorts of 200 lambs - *N. battus* FEC data shown. Datasets with lower 95% CI estimates for zero-inflation of above 10% using all models shown in bold.

	Lower 95% CI		Upper 95% CI	
	ZILP model	ZIGP model 1	ZIGP model 2	ZIGP model 1
May 1992	8.3	5.2	0	78.1
June 1992	9.1	28.8	0.1	75.8
August 1992	0.1	4.9	0	59.7
September 1992	0	0	0	61.4
Post-mortem 1992	0	13.6	0	80.2
May 1993	0	0	0	58.4
June 1993	4.8	16.0	0.1	56.7
August 1993	0	21.8	0	70.8
September 1993	0	8.5	0.1	82.1
Post-mortem 1993	0	5.3	0	67.7
May 1994	74.7	5.0	0	90.8
June 1994	41.4	27.2	20.6	59.3
August 1994	0	11.6	0	72.3
September 1994	0.1	17.2	0.1	78.9
October 1994	11.8	35.6	0.1	94.0
Post-mortem 1994	72.1	80.6	7.1	94.1
May 1995	79.6	77.2	68.0	90.0
June 1995	12.9	12.6	12.3	23.5
July 1995	16.2	4.9	2.4	34.9
August 1995	13.8	1.9	0	70.0
September 1995	0	23.6	0	78.7
October 1995	31.0	54.5	3.2	94.8
November 1995	15.4	3.3	0	74.3
Post-mortem 1995	10.6	2.9	0.1	72.0
June 1996	19.3	18.8	18.6	31.1
July 1996	37.2	16.0	8.7	55.1
August 1996	41.2	47.5	4.2	84.9
September 1996	16.0	11.1	0	91.5
October 1996	8.5	25.3	0.1	95.5
				78.2
				80.9
				72.3
				71.6
				95.7
				64.1
				61.0
				83.9
				96.8
				81.7
				86.1
				58.1
				86.0
				98.4
				97.1
				94.0
				90.5
				23.2
				32.0
				64.2
				93.9
				95.6
				70.5
				68.6
				30.8
				52.6
				85.7
				92.6
				97.4
				66.9
				68.7
				59.0
				63.4
				82.1
				56.7
				50.3
				69.8
				85.3
				66.4
				74.3
				57.3
				73.2
				86.6
				86.3
				93.0
				23.0
				30.0
				59.2
				80.1
				90.1
				61.1
				61.9
				30.5
				51.7
				79.4
				79.9
				86.7

The only deaths due to *Nematodirus* spp. infections on this farm occurred in May and June 1995 during the period of peak intensity, which corresponds to the highest numbers of observed eggs. Consequently, the 95% CI for zero-inflation were relatively small and it was possible to demonstrate a reduction in zero-inflation from a lower 95% CI of between 68% and 80%, depending on the model, in May to an upper 95% CI of between 23% and 24% in June.

4.2.4 Discussion

There was agreement between the median estimates for mean count produced by the ZILP and ZIGP models, but less for zero-inflation and cv . The ZILP model produced higher median estimates of cv than the ZIGP model with log-uniform prior on cv , but not compared to the ZIGP model with uniform prior on cv . The estimates for cv derived using the ZIGP model are therefore influenced quite strongly by the prior, indicating a lack of information in some datasets. The most extreme examples of this lack of information are for the five *Nematodirus* spp. datasets that produced high estimates of cv from both ZIGP models relative to the ZILP models; these datasets also produced relatively high estimates for mean count using the ZIGP models and had relatively low numbers of observed eggs, and are therefore likely to be influenced by the prior for mean as described in [Section 2.7](#). The estimates for zero-inflation produced by the ZIGP model were also quite dependent on the prior for

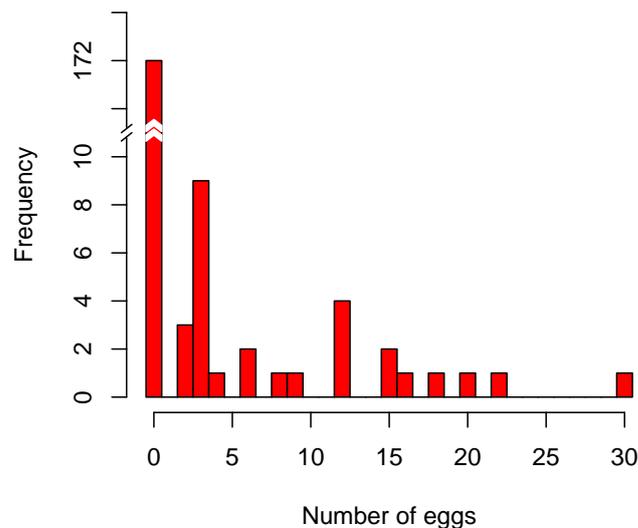


Figure 4.6: Histogram of observed *Nematodirus battus* eggs enumerated in the faeces of 200 lambs in May 1995 - 1 egg represents 25 EPG

cv , indicating a cross-dependence between these parameters. This could partly explain the correlation between median estimates of cv and zero-inflation seen in [Figures 4.4 and 4.5](#). The estimate for zero-inflation was lower using the uniform prior on cv for all but the five *Nematodirus* spp. datasets with low numbers of observed eggs, likely as a result of the model tending towards high estimates for cv allowing for higher estimates for mean count and therefore increasing the estimate for zero-inflation to help account for the observed zeros. The lower 95% confidence interval for zero-inflation was also quite variable between ZILP and ZIGP models; the figure for *N. battus* in May 1994 (75% compared to 5% and 0%) being an extreme example. This highlights the practical implication of the discrepancy between LP and GP distributions with certain combinations of parameter values. There appeared to be more consistency between models for the worm burden data, which most likely reflects the difference in parameter values between the worm burden and FEC data giving the gamma-Poisson and lognormal-Poisson distributions a more similar shape for the worm burden data, but could potentially be the result of a true difference between the underlying distributions of worm burden and FEC.

There was no discernible relationship between the median estimates of mean count, zero-inflation and cv produced by the ZILP model, but there was evidence for such a relationship using both ZIGP models. This indicates that the estimates of cv and zero-inflation produced by the ZIGP model are more closely related to each other than when using the ZILP model. This is likely to be a result of reduced identifiability of the ZIGP model compared to the ZILP model, either due to a possible greater flexibility of the gamma distribution compared to the lognormal distribution, or a weaker fit to the ZIGP model. However, it is also possible that the data truly exhibit a correlation between these factors which is not detected using the ZILP model because of a poorer fit to the data.

There were on average more observed *N. battus* eggs than *Nematodirus* spp. eggs in the faeces of the lambs. Several of the *Nematodirus* spp. datasets had no observed eggs at all, and the remainder produced on average higher estimates for cv and zero-inflation and lower estimates for mean count than the *N. battus* datasets. The 95% CI for zero-inflation were very wide for the *Nematodirus* spp. datasets, indicating the possibility that these datasets had a low mean and large cv rather than appreciable zero-inflation. For the *N. battus* data, there was sufficient evidence to show that the distribution was closer to a ZIGP than GP and closer to ZILP than LP for four of the datasets, one of which contained sufficient evidence to be sure that the zero-inflation component was greater than 65% of the total. The histogram for this dataset shows all positive counts within the range of 2-30, with the majority contained between 3 and 20. On the basis of these 28 counts, there is sufficient evidence to be sure that the cv is not very high if the data are fit to a gamma-Poisson or lognormal-Poisson distribution, although the large proportion of zero counts is not consistent with this. The only way that these zero counts can be explained is by attributing the majority of them to a distribution which is separate from the gamma or lognormal distribution describing the

mean of the other counts. One biological explanation for these excess zero counts is that the animals in the ‘extra zeros’ group have a mean FEC of zero, i.e. do not contain any adult *N. battus* worms. The prevalence of *N. battus* is known to be lower than that of many other nematodes, potentially justifying this explanation. However, it is also possible that the infected animals are described by a distribution with more than one positive mode, or by a uni-modal distribution with higher weight at zero given similar mean and variance than would be expected from the gamma or lognormal distributions.

The demonstrated reduction in zero-inflation between May and June 1995 is consistent with the view that zero-inflation in *N. battus* is as a result of a lack of adult worms in some animals - with zero-inflation being higher earlier in the grazing season due to fewer lambs having mature worms in the gut. There was no significant relationship between the mean count and zero-inflation using the ZILP model and ZIGP model with log-uniform prior on cv , which would suggest that an empirical rise in FEC could be explained by either increased mean FEC of the infected group or increased prevalence independently of each other.

4.3 Equine FEC data

4.3.1 Introduction

The epidemiology of infection with cyathostomins is quite different to that of *N. battus*, both in terms of the biology of the parasite and in terms of the usual host characteristics and husbandry. These factors might be expected to lead to a quite different fit to the ZILP and ZIGP models compared to that observed with sheep data in [Section 4.2](#). Comparative analysis of equine FEC data using both ZILP and ZIGP models were undertaken on field data obtained from a variety of sources in Denmark and the UK.

4.3.2 Materials and methods

FECRT data from 64 herds of horses in Denmark, referred to henceforth as ‘Danish FECRT’, were obtained. The data were collected as voluntary submission of faeces from groups of horses to monitor for anthelmintic resistance. Only pre-treatment data from full groups was used; post-treatment data were discarded. A further two FECRT datasets denoted ‘UK FECRT’ were obtained from a UK property involved in the manufacture of biological products from a large number of long-term resident horses; post-treatment data were discarded as before. Further equine data denoted ‘UK FEC’ were obtained from voluntary FEC submissions to a commercial parasitology laboratory (Diagnosteq Liverpool), from private practises

around the UK. Submissions were made from veterinary surgeons rather than individual establishments, so all data submitted from the same practise on the same day was deemed to have originated from the same group of horses. Of the 3115 FEC datasets available from Liverpool, 2592 were excluded from analysis due to a sample size of less than ten, and 79 of the remaining datasets were excluded on the basis of all observations being zero. A total of 444 UK FEC submission datasets were included in the final analysis. None of the strongyle eggs were speciated, but the majority of the observed counts are likely to be cyathostomin eggs.

4.3.3 Results

Shown in [Figure 4.7](#) are the agreement between the median estimates of mean egg count in eggs per gram (EPG), cv and zero-inflation produced by the ZILP and ZIGP model with log-uniform prior. As with the sheep data, there was generally good agreement for mean count for the majority of the datasets, although some datasets produced much higher estimates for mean count using the ZIGP model compared to the ZILP model. Agreement for zero-inflation was better than observed for the sheep data, although the agreement for cv was still poor. The ZILP median estimates of cv were higher for all Danish FECRT datasets, both of the UK FECRT datasets and over half of the FEC submission datasets. The remainder of the FEC submission datasets produced much higher estimates of cv with the ZIGP model, probably corresponding to those with a relatively high estimate for mean as a result of low numbers of observed eggs. The agreement between the ZILP median estimates and those produced using the ZIGP model with uniform prior on cv is shown in [Figure 4.8](#). The agreement line is similar for all parameters for the Danish FECRT and UK FECRT datasets, although those UK FEC submission datasets with low numbers of observed counts produced higher estimates

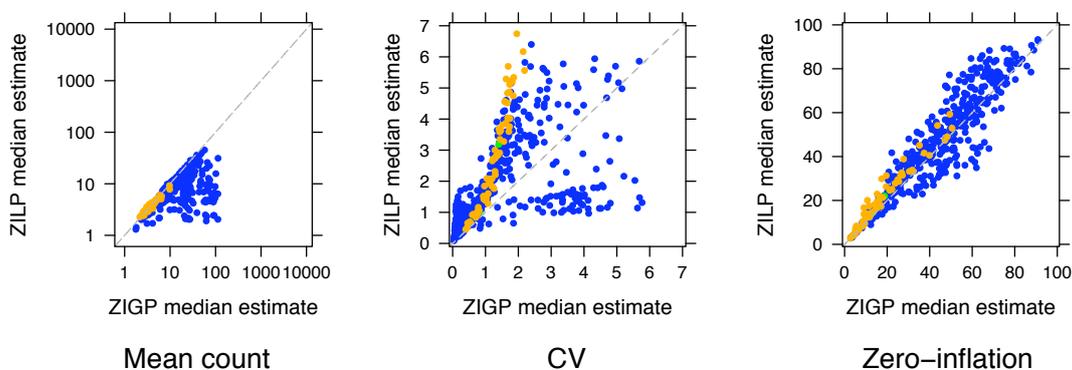


Figure 4.7: Comparison of the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with log-uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC data shown in blue, Danish FECRT data in orange, and UK FECRT data in green.

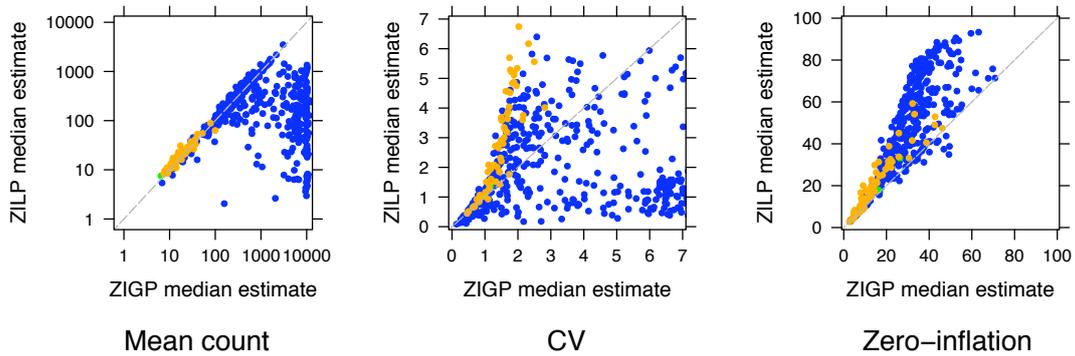


Figure 4.8: Comparison of the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model and ZIGP model, with uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC data shown in blue, Danish FECRT data in orange, and UK FECRT data in green.

for mean count and cv and lower estimates for zero-inflation using the ZIGP model with the uniform prior compared to with the log-uniform prior. There was no discernible relationship between any of the median estimates of mean count, cv and zero-inflation parameters for the ZILP model (Figure 4.9). The median estimates for cv and zi , and possibly mean and cv , showed evidence of correlation using both ZIGP models (Figures 4.10 and 4.11) as with the sheep data.

Shown in Figure 4.12 are the empirical cumulative distribution function for the lower (lines to top left of each plot) and upper (lines to bottom right of each plot) 95% CI for zero-inflation produced by the three models (shown coloured) for the UK FEC submission data (left hand plot) and combined Danish and UK FECRT data (right hand plot). There was evidence for zero-inflation using the ZIGP model with uniform prior on cv for 12% of the UK FEC submission data and less than 10% of the combined FECRT data, although the lower 95%

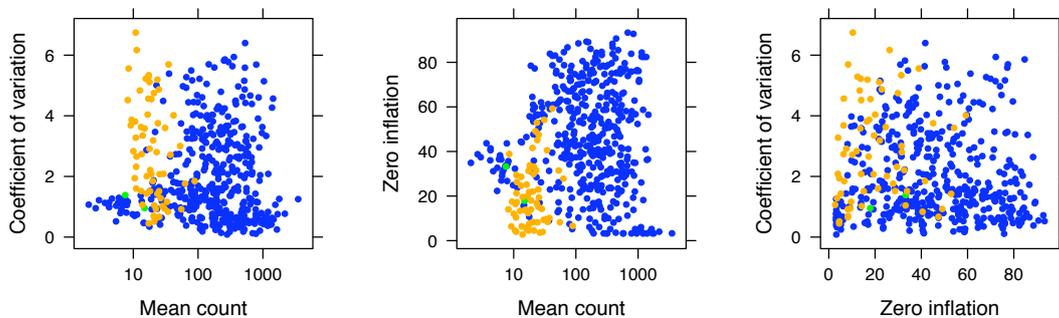


Figure 4.9: The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZILP model for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC data shown in blue, Danish FECRT data in orange, and UK FECRT data in green.

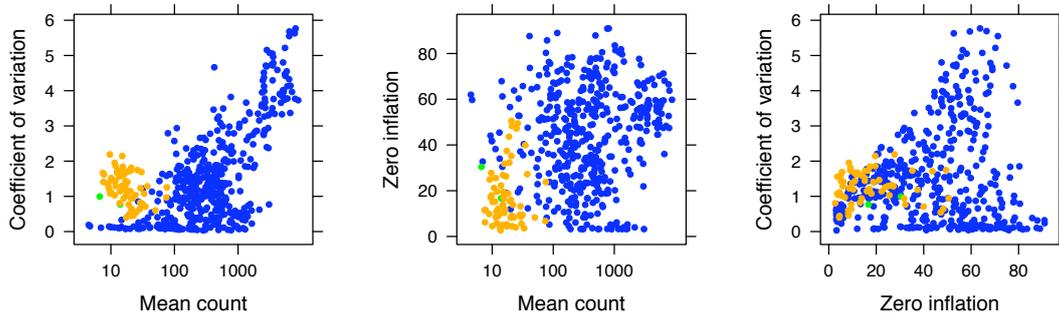


Figure 4.10: The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZIGP model, with log-uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC data shown in blue, Danish FECRT data in orange, and UK FECRT data in green.

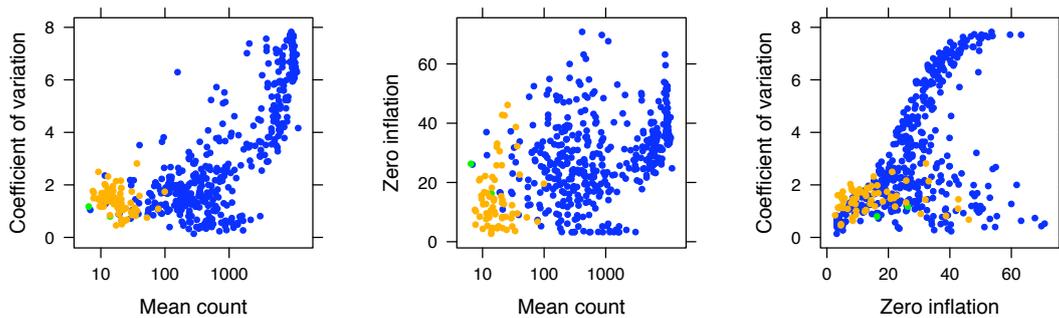


Figure 4.11: The relationship between the median estimates of mean strongyle FEC (EPG), cv and zero-inflation produced by the ZIGP model, with uniform prior on cv , for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC data shown in blue, Danish FECRT data in orange, and UK FECRT data in green.

CI were higher using the ZILP model. The upper 95% CI was above 10% for all datasets, and as high as 100% for the FEC data and 80% for the FECRT data.

4.3.4 Discussion

The median estimates of mean count and cv for the Danish FECRT data and a subset of the UK FEC submission data were unaffected by the choice of prior for the ZIGP model, indicating that the results obtained from these datasets are predominantly data driven. The subset of UK FEC submission datasets for which the inference was affected by the choice of prior gave relatively high median estimates for mean count and cv using the ZIGP model compared to the ZILP model, indicating a lack of information in the data due to low numbers

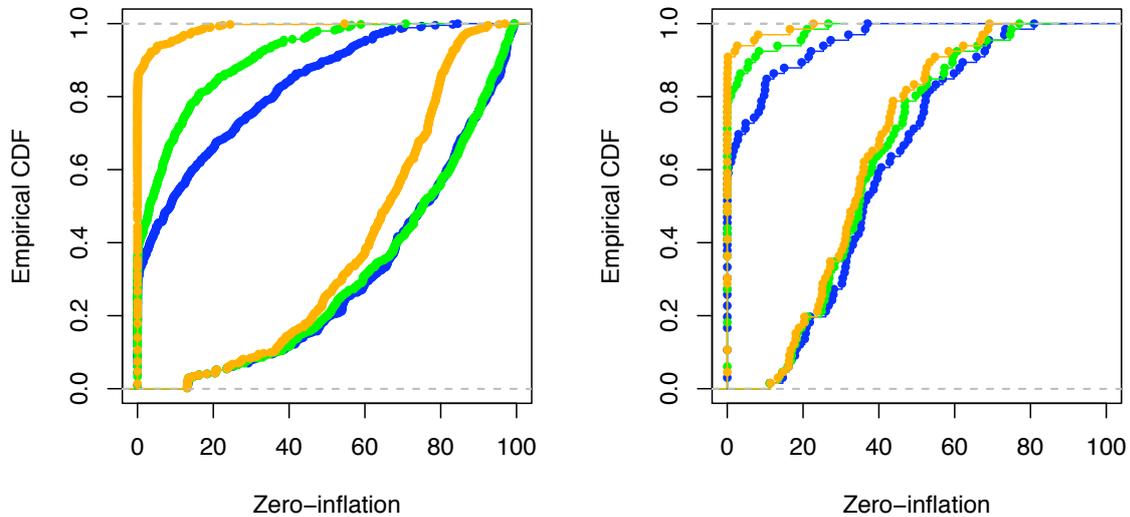


Figure 4.12: Empirical cumulative distribution function of the lower and upper 95% CI estimates for zero-inflation produced by the ZILP model (blue), ZIGP model with log-uniform prior on cv (green) and ZIGP model with uniform prior on cv (orange) for 444 UK FEC, 64 Danish FECRT and 2 UK FECRT datasets - UK FEC submission data shown on the left hand plot, combined Danish and UK FECRT data shown on the right hand plot.

of observed eggs in these datasets as discussed in [Section 2.7](#). These datasets are the group of UK FEC datasets that have a higher estimate of cv and mean count produced using the ZIGP model than the ZILP model, and appear to be a distinct group compared to the remaining datasets shown in [Figure 4.8](#). The median estimate for zero-inflation was lower for all datasets using the ZIGP model with the uniform prior on cv , which suggests that this parameter is affected by the choice of prior for cv . With the exception of those UK FEC submission datasets which were substantially prior driven, agreement between the ZILP and ZIGP models was generally good for mean count and zero-inflation, although the median estimate for cv was larger for the ZILP model than the ZIGP models when both estimates had a value of greater than one. This apparent change in relationship with a cv of less than one compared to greater than one corresponds with the change in shape of the gamma distribution between skewed and exponential shapes, and is the most likely explanation for the phenomenon. The ZILP model also produced higher estimates for the lower 95% CI of zero-inflation than either ZIGP models, although the upper 95% CI estimates for zero-inflation were similar to those produced by the ZIGP model with log-uniform prior on cv . These findings are consistent with those produced using simulated data ([Section 3.5](#)). As with the sheep data ([Section 4.2](#)), there was no discernible relationship between the median estimates of each parameter produced using the ZILP model.

The median estimates of cv and zero-inflation were similarly distributed between the UK FEC submission datasets and the Danish FECRT data, although there were proportionately more UK datasets with a low mean of ≈ 10 EPG. The UK dataset also included proportionately more smaller sample sizes than the Danish data, leading to the inference for a greater proportion of the UK FEC submission datasets being heavily influenced by the prior for cv . The discrepancy in sample sizes is likely to be a result of the different reasons for submission of the data; FEC data are often submitted for individual or small groups of animals as part of long term parasite burden surveillance. However, a FECRT is unlikely to be performed on such a small sample size of animals. The discrepancy in mean count may also be a result of this, but could also be influenced by the more strict Danish legislation regarding use of anthelmintics decreasing their frequency of use. It is also important to note that the UK FEC submission data were retrieved from a general database of laboratory results, which included only the address of the veterinary surgeon from which the data were submitted and a date of submission. Although data submitted from the same place on the same date were assumed to be from the same group of animals, it is possible that this may not be the case for some datasets. Veterinary practises may have submitted samples from a number of premises in the same batch, which would have inadvertently been combined for this analysis, although the potential for this is somewhat limited as samples would have to be submitted to the laboratory shortly after collection. It is also impossible to identify groups of animals that were grazed as a group as opposed to being grazed separately using this data, which may also affect the distribution of the data. This may partly explain why the relationship between the median estimates for cv produced by the ZILP and ZIGP models is different for a group of the UK FEC data than the trend for the Danish and UK FECRT data, although this is more likely to be a consequence of the issue of small numbers of observed counts and the prior for mean count as discussed above.

As with the sheep data ([Section 4.2](#)), there was evidence of appreciable zero-inflation in some of the equine datasets using all three models. There were more UK FEC submission datasets with lower and upper 95% CI for zero-inflation of above 10%, which is possibly a result of analysing some data comprising animals grazed in more than one group as a single dataset. However, the lower 95% CI for approximately 30% of the Danish FECRT data indicated zero-inflation of more than 10% using the ZILP model, although this figure was lower using the ZIGP models. It is possible that the estimates for zero-inflation truly represent instances where animals are uninfected, although the prevalence of cyathostomins was 100% in a study in the Ukraine ([Kuzmina and Kharchenko, 2008](#)), and is anecdotally thought to be very high in most groups of horses, even those that undergo frequent anthelmintic treatment. The indication of a zero-inflated group is therefore more likely to represent instances where the observed data did not easily fit a GP or LP distribution due to a greater weight at zero than expected given the mean and variance of the remaining counts. The greater flexibility of the gamma distribution compared to the lognormal distribution explains the inflated estimates of zero-inflation using the ZILP model compared to ZIGP models. This finding suggests that

the true underlying distribution of the data is closer to either a multi-modal distribution, or to a uni-modal distribution with greater weight at zero than provided by the LP or GP distributions, than to the LP or GP distributions.

4.4 Empirical fit

4.4.1 Introduction

The inference made from the data presented in this chapter was quite different between the ZILP and ZIGP models. The lower 95% CI for zero-inflation and median estimate for cv was particularly disparate for the equine data. Based on the work presented in [Section 3.4](#) it is clear that differentiation of ZIGP data from ZILP data on empirical fit is difficult, and that the parameter values as well as sample size effect the usefulness of comparison of model fit based on a likelihood ratio. However, model selection using penalised likelihoods is common practice, and may still provide an indication of which analysis is more likely to be accurate. Based on the results presented in [Section 3.4](#) the likelihood ratio is dependent on the sample size and parameter values of the data as well as the true underlying distribution of the data. A generalised linear model is used here to examine the relationship between maximum likelihood fits of each model and the parameter values and sample size of the data, and observations are related to those made using simulated data of known distributional origin in [Section 3.4](#).

4.4.2 Materials and methods

The procedure described in [Section 3.4](#) was used to provide maximum likelihood fits to the data presented in [Sections 4.2](#) and [4.3](#) using the GP, LP, ZIGP and ZILP distributions. The maximum log likelihood for each combination of distribution and dataset was then input into R for comparative analysis. The proportion of each type of dataset that produced a better fit to the LP compared to GP or ZILP compared to ZIGP model was calculated, using a beta distribution with minimally informative prior to generate 95% CI. A generalised linear model was then used to examine the relationship between the likelihood ratios for the zero-inflated models and the sample size and median estimates for mean count, cv and zero-inflation produced by the ZILP MCMC model. The maximum likelihood estimates for zero-inflation using the ZILP and ZIGP models were also recorded.

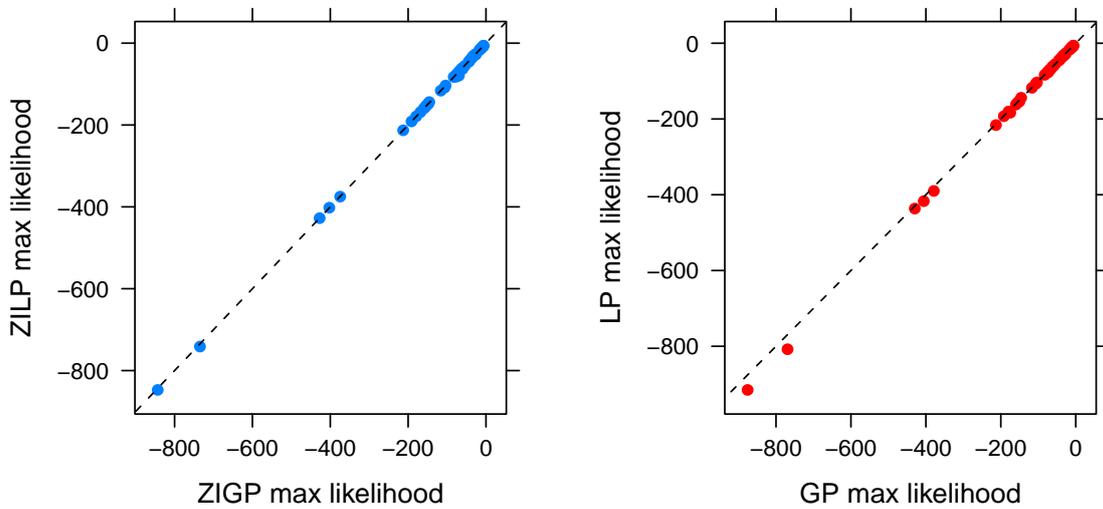


Figure 4.13: Maximum log likelihood produced using the ZIGP compared to ZILP, and GP compared to LP models, from analysis of all *N. battus* and *Nematodirus* spp. worm burdens and FEC in five cohorts of 200 lambs - dotted line represents equivalence of the two models

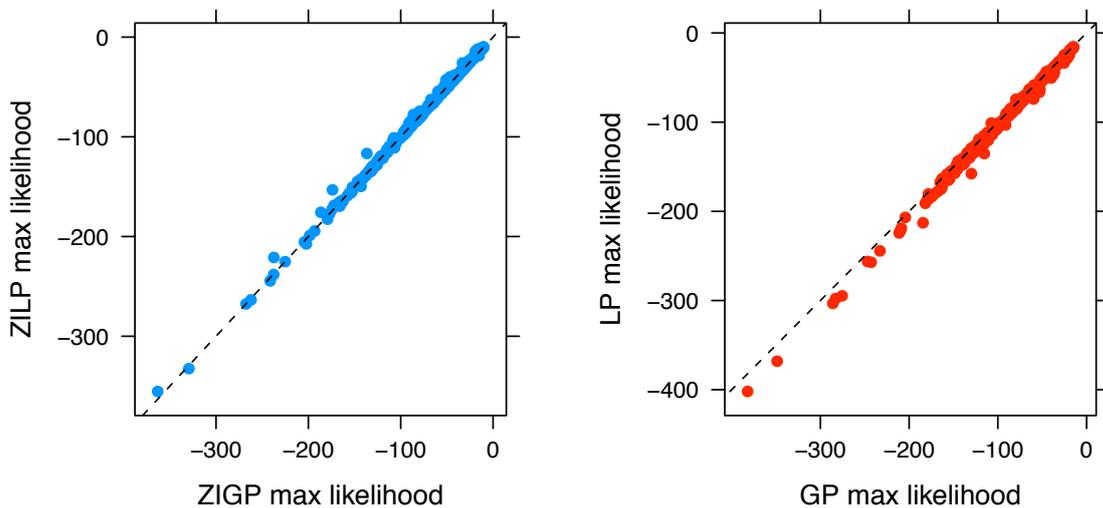


Figure 4.14: Maximum log likelihood produced using the ZIGP compared to ZILP, and GP compared to LP models, from analysis of all 444 UK FEC, 64 Danish FECRT and 2 UK FECRT equine datasets

4.4.3 Results

A comparison of the maximum log likelihoods produced by the GP compared to LP and ZIGP compared to ZILP distributions is shown in Figures 4.13 and 4.14 for the sheep and horse data respectively. The maximum log likelihoods were very similar between the ZIGP and ZILP models, but consistently less for the LP compared to GP model. Table 4.2 shows the proportion of datasets of each type that produced a higher maximum log likelihood using the LP compared to GP or ZILP compared to ZIGP models. The observed proportion of datasets that produced a better fit to the LP model was below 50% in all but the *Nematodirus* spp. FEC dataset, and the upper 95% CI was below 50% for four of the six types of data excluding the two combined datasets. The *Nematodirus* spp. FEC data that produced a better fit for the GP model were those from May 1992, PM 1993, PM 1994 and August 1996 - all of which had relatively high lower 95% confidence intervals for zero-inflation (Table 4.1b). The same four *Nematodirus* spp. FEC datasets also produced a better fit to the ZIGP model. The 95% CI for the true proportion of sheep data that produced a better fit to the ZILP model included 50% in every case, whilst this was true for only one set of equine data. The lower 95% CI was above 50% for the UK FEC submission data, indicating a better fit to the ZILP model, and the upper 95% CI was below 50% for the Danish FECRT data, indicating a better fit to the ZIGP model.

The estimates for coefficients obtained using the GLM with the log of the ratio of ZILP maximum likelihood to ZIGP maximum likelihood are shown in Tables 4.3 and 4.4. There was no evidence that the likelihood ratio is related to the mean count, zero-inflation or sample size for the sheep data, but there was potentially a positive relationship between cv and likelihood ratio for the *N. battus* data. However, this relationship was not significant, even when all sheep FEC data were pooled. There was also no evidence for a relationship

Table 4.2: The proportion of maximum likelihoods that were higher for the LP compared to GP model (left) and for the ZILP compared to ZIGP model (right) for all equine and sheep data - 95% CI calculated using a beta distribution.

	Better fit to LP			Better fit to ZILP		
	Lower	Median	Upper	Lower	Median	Upper
<i>N. battus</i> FEC	0.07	0.17	0.33	0.22	0.38	0.56
<i>Nematodirus</i> spp. FEC	0.47	0.71	0.90	0.47	0.71	0.90
All <i>Nematodirus</i> FEC	0.22	0.35	0.49	0.35	0.49	0.63
<i>Nematodirus</i> burden	0.00	0.00	0.45	0.15	0.50	0.85
UK FECRT	0.00	0.00	0.63	0.00	0.00	0.63
Danish FECRT	0.17	0.27	0.38	0.21	0.31	0.43
UK FEC	0.01	0.02	0.03	0.51	0.56	0.60
All equine	0.03	0.05	0.07	0.48	0.53	0.57

Table 4.3: Coefficients for mean, *cv*, zero-inflation and sample size in a GLM fit to the (log) likelihood ratios for *N. battus* and Nematodirus spp. FEC obtained from five cohorts of 200 lambs, fitted separately and together

		Lower 95% CI	Mean estimate	Upper 95% CI
Mean	<i>N. battus</i>	-0.04	0.00	0.04
	Nematodirus spp.	0.00	0.00	0.00
	All Nematodirus	-0.01	0.00	0.01
CV	<i>N. battus</i>	-0.18	0.43	1.04
	Nematodirus spp.	-0.33	-0.08	0.17
	All Nematodirus	-0.14	0.31	0.76
ZI	<i>N. battus</i>	-0.06	0.01	0.07
	Nematodirus spp.	-0.02	0.00	0.02
	All Nematodirus	-0.04	0.00	0.04
Size	<i>N. battus</i>	-0.01	0.00	0.01
	Nematodirus spp.	0.00	0.00	0.01
	All Nematodirus	-0.01	0.00	0.01

Table 4.4: Coefficients for mean, *cv*, zero-inflation and sample size in a GLM fit to the (log) likelihood ratios for Danish FECRT and UK FEC equine data, fitted separately and together

		Lower 95% CI	Mean estimate	Upper 95% CI
Mean	Danish FECRT	-0.03	-0.01	0.02
	UK FECRT	0.00	0.00	0.00
	All equine	0.00	0.00	0.00
CV	Danish FECRT	-0.48	-0.24	0.00
	UK FECRT	-0.02	0.12	0.26
	All equine	-0.08	0.04	0.17
ZI	Danish FECRT	-0.06	-0.04	-0.02
	UK FECRT	0.01	0.01	0.02
	All equine	0.01	0.02	0.02
Size	Danish FECRT	-0.03	0.01	0.04
	UK FECRT	0.02	0.03	0.05
	All equine	0.01	0.02	0.04

between mean count and likelihood ratio for the equine data, although there was evidence for a weak positive effect of zero-inflation and sample size on likelihood ratio for the UK FEC submission data. There was also evidence for an effect of zero-inflation on likelihood ratio for the Danish FECRT data, although this effect was opposite to that seen in the UK FEC data. There was some evidence for a negative effect of cv on likelihood ratio for the Danish FECRT data and positive effect of cv on likelihood ratio for the UK FEC submission data. Plotting the likelihood ratios against median estimate of cv for both the sheep and horse data (Figure 4.15), and against the median estimate of zero-inflation and sample size for the horse data (Figure 4.16) shows no obvious pattern.

The ML zero-inflation estimate was also recorded for each dataset for the ZIGP and ZILP models. The ML estimates for zero-inflation were significantly higher for the ZILP model than ZIGP model ($p < 0.001$, calculated using the paired Wilcoxon signed rank test). There was also a significant Spearman's rank correlation between the log maximum likelihood and corresponding zero-inflation estimate for both the ZILP ($p < 0.001$) and ZIGP ($p < 0.001$) models. The slope and intercept of this correlation, shown in Figure 4.17, was almost identical between the models.

4.4.4 Discussion

The close association of the maximum log likelihoods produced by the ZILP and ZIGP, and to a lesser extent the LP and GP models, shows the relative similarity between the underlying

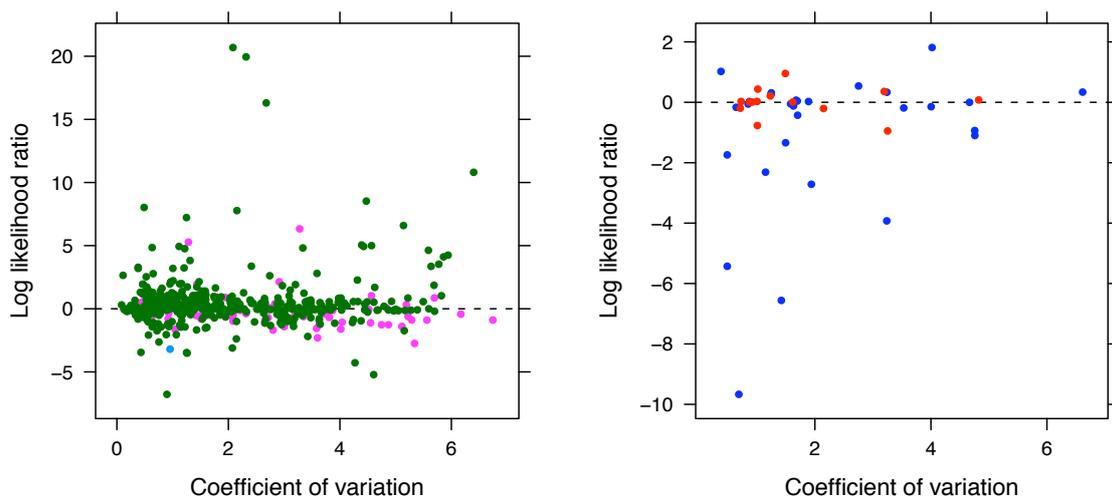


Figure 4.15: The relationship between the median estimate for cv produced by the ZILP model and the log of the likelihood ratio, for all equine (left) and all sheep (right) data - UK FEC data shown in green, Danish FECRT data in pink, and UK FECRT data in light blue. *N. battus* datasets shown in royal blue, *Nematodirus* spp. in red.

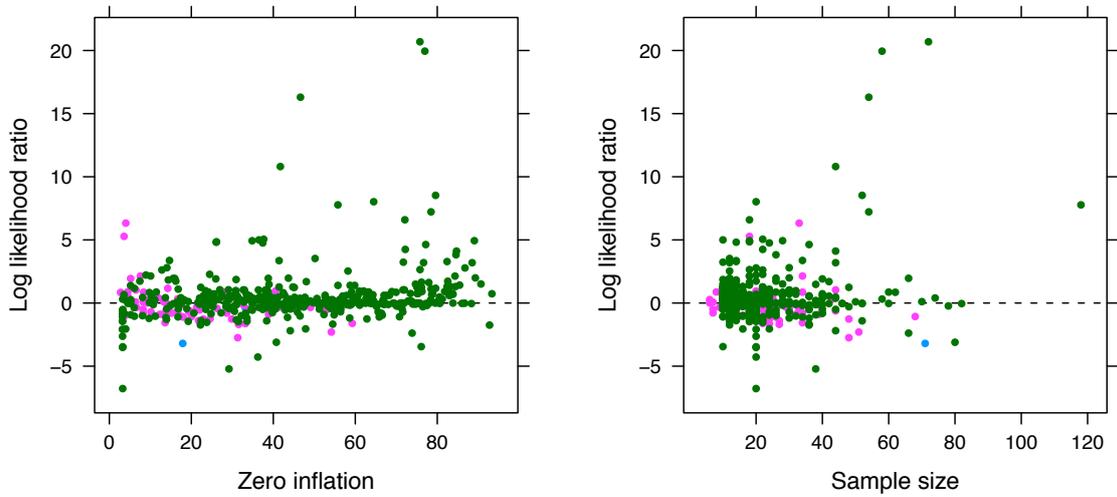


Figure 4.16: The relationship between the log of the likelihood ratio and the median estimate for *zero – inflation* produced by the ZILP model (left) or sample size (right) - UK laboratory FEC submission data shown in green, Danish FECRT submission data in pink, and blood farm data in light blue.

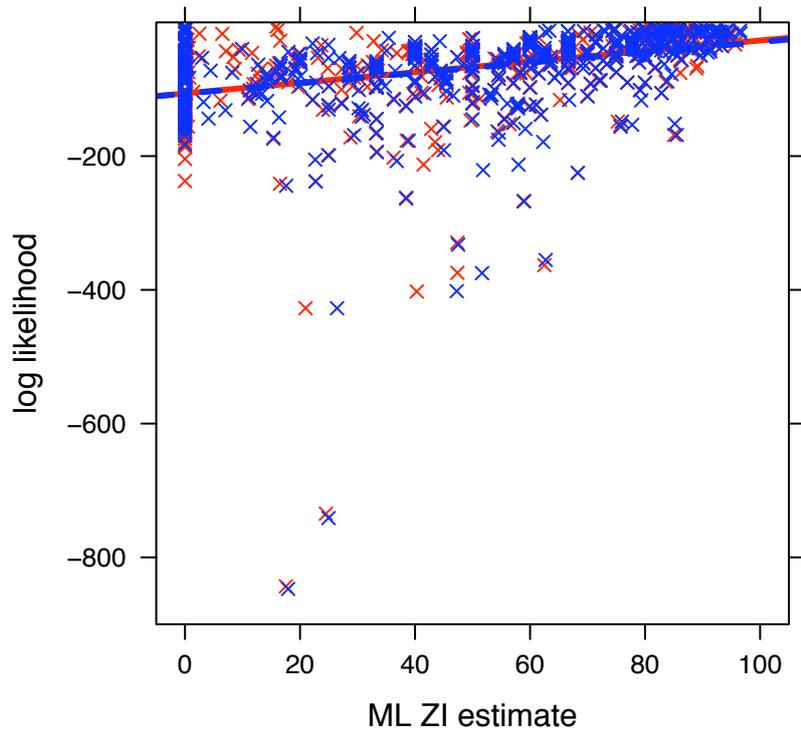


Figure 4.17: The relationship between (log) maximum likelihood and ML estimate for zero-inflation produced by the ZIGP (red) and ZILP (blue) models for all equine and sheep data - correlation line shown in solid red for ZIGP model, and superimposed dashed blue for ZILP model

gamma and lognormal distributions. However, there were differences between the maximum likelihood fits produced. With the exception of the *Nematodirus* spp. data all of the datasets examined provided a better fit to the GP than the LP model, and the four *Nematodirus* spp. datasets with demonstrable zero-inflation also produced a better fit to the GP model. This is likely to be as a result of the greater flexibility to account for zero observations using the gamma distribution compared to lognormal distribution. The UK FEC submission data provided a particularly poor fit to the LP data, possibly as a result of a proportionately larger number of zero counts. There was no evidence to suggest that the sheep data produced a better fit to either the ZILP or ZIGP distributions; although a greater proportion of the *Nematodirus* spp. data produced a better fit to the ZILP model and a greater proportion of the *N. battus* data produced a better fit to the ZIGP model, the 95% CI for the true proportion included 50% in both instances. Of the equine data, a greater proportion of the Danish FECRT data gave a better fit to the ZIGP model, whereas a greater proportion of the UK FEC submission data gave a better fit to the ZILP model. The apparent discrepancy in model fit could potentially be explained by the differences in husbandry practises between the UK and Denmark, but it is also possible that the true distribution of equine FEC data follows neither of the distributions examined.

The results presented in [Section 3.4](#) suggest that the true proportion of datasets that produce a better fit to the ‘true’ underlying distribution of the data is effected by the sample size, mean, and *cv* of the data. This simulated data did not include a zero-inflated component, but it is also likely that zero-inflation will have an effect by reducing the effective sample size of infected individuals. A GLM using these explanatory variables to describe the log of the likelihood ratio, the ratio of the ZILP maximum likelihood to ZIGP maximum likelihood, indicated that the mean count, zero-inflation and sample size had little effect on the likelihood ratio for the sheep data. There was possibly an effect of *cv* on the likelihood ratio, with a greater *cv* corresponding to a relatively better fit to the ZILP model for the *N. battus* data and a greater *cv* corresponding to a relatively better fit to the ZIGP model for the *Nematodirus* spp. data, although the effect was not significant in either case. There was also no evidence for an effect of mean count on the likelihood ratio using the equine data, although a weak positive effect of zero-inflation and sample size was evident for the UK FEC submission data and a weak negative effect of zero-inflation was evident for the Danish FECRT data. These findings are consistent with the apparent better fit of the Danish FECRT data to the ZIGP model and *vice versa*. There was a stronger but less significant relationship between the median estimate of *cv* and the likelihood ratio, again suggesting a better fit to the ZIGP model for the Danish FECRT data and to the ZILP model for the UK FEC submission data. However, graphical representation of these effects demonstrates the weakness of these relationships relative to the similar effect of sample size, mean count and *cv* on the proportion of simulated datasets that produced a better fit to the ‘true’ distribution ([Section 3.4](#)). This may be as a result of using median estimates of each parameter rather than the simulation parameter, but also suggests that the underlying distributions are not fully represented by either distribution.

The median estimates used were produced using the ZILP model to reduce the impact of the large estimates for mean count and cv produced using the ZIGP models for datasets with low numbers of observed counts. However, similar results could have been obtained using the ZIGP median estimates. It would also have been possible to approach the likelihood comparisons by examining a distribution of likelihood ratios with data from each type of source combined together to increase the power to differentiate the models for the combined data. However, this method would not have facilitated the relationship between likelihood ratio and parameter value, which was the main objective here.

There was a significant relationship between the ML estimate for zero-inflation and the maximum log likelihood produced by each model. This is probably a result of the variable and ill-defined degrees of freedom inherent in mixture models, and demonstrates the futility of penalised likelihood ratio tests where one or more of the alternative models is a mixture models. This prevents the use of penalised likelihoods to compare a GP with ZIGP or LP with ZILP model, although the relationship between zero-inflation and likelihood was virtually identical between the ZILP and ZIGP models, indicating that these models do have the same number of effective parameters for the same value of zero-inflation. Comparison of likelihoods between the models is therefore valid as long as the estimate for zero-inflation is similar. However, the tendency towards higher estimates of zero-inflation with the ZILP model were also observed with ML estimates, which may be expected to artificially improve the apparent fit of the ZILP model compared to the ZIGP model as a result of the associated increased effective degrees of freedom.

4.5 Discussion

In this chapter, the inference made from sheep and equine FEC data using the ZILP and ZIGP models has been examined. The two models produced similar median estimates of mean count and zero-inflation from datasets that were predominantly likelihood driven, but the ZILP model tended to produce higher estimates for cv . The 95% CI estimates for zero-inflation also differed between the ZILP and ZIGP models, with the ZILP model producing lower 95% CI estimates that indicated a greater degree of zero-inflation for the equine data. Although the two distributions are very similar with values of cv less than one, it appears from these results that the disparity between the distributions with greater cv is sufficient to affect the inference made, especially for cv and zero-inflation. The GP distribution was able to account for a larger number of zero observations than the similarly parametrised LP model as a result of the shape of the distribution with a cv equal to or greater than one. The higher mode of the LP distribution results in a reduced weight at zero compared to the GP with equal mean and variance, so in order to explain the observed zero counts the estimate for cv and zero-inflation is increased for the ZILP model relative to the ZIGP model. This

inconsistency between the two models introduces difficulties to the analysis of real datasets where the true underlying distribution of the data is unknown, and means that estimates of the variability within a dataset cannot be directly compared between ZIGP and ZILP models. This is further compounded by the fact that the effective degrees of freedom in the zero-inflated models is linked to the estimate for zero-inflation, which tends to be higher for the ZILP compared to ZIGP model. Such differences are likely to be small between the zero-inflated models, and are unlikely to have affected these results as all inference was compared to that made using simulated data which was subject to the same effect of zero-inflation estimate. However, this does illustrate the difficulty of defining the effective degrees of freedom in a mixture model and prevents comparison with the simpler GP or LP models based on penalised likelihoods.

An appreciable degree of zero-inflation was found in both the sheep and equine data. Zero-inflation in *N. battus* FEC data is consistent with the current belief that this parasite is not present in all animals within a population, although this is not the case for existing studies of either *Nematodirus* spp. parasites (Barger, 1985) or cyathostomins (Kuzmina and Kharchenko, 2008). In the strict interpretation of a zero-inflated model, a proportion of the observed zero counts are derived from animals with an infection rate of zero; i.e. worm eggs would never be identified in the faeces of these animals. This could be as a result of differences in grazing, for example the uninfected animals are grazed on pasture that does not contain infective larvae, or any other factor affecting those animals differently to the remainder of the group and which prevents them from containing mature adult worms. In terms of *N. battus* in lambs, a zero-inflated distribution early in the grazing season could be a result of one group of older lambs containing mature worms and another group of younger lambs that are not yet old enough for the life cycle of the parasite to be completed. However, it is difficult to imagine a similar effect in the distribution of cyathostomin eggs between adult horses that have been grazed and treated with anthelmintic as a group. In this situation, it is more likely that the zero-inflated distribution is actually describing a multi-modal positive distribution of underlying mean counts, one of which has a relatively low mean and contributes mostly zero observations to the overall distribution. In this situation, the use of a zero-inflated distribution produces similar inference to a bi-modal distribution, but any low positive counts that may arise from the low mean group would be falsely attributed to the high mean group using a zero-inflated distribution. Each sub-group of the multi-modal distribution would still be expected to follow a lognormal-Poisson distribution, but a separate lognormal distribution is needed to describe each group of animals with different means. This effect potentially explains the inconsistency in the fit to the ZILP and ZIGP distributions; as the gamma is more flexible than the lognormal the fit of the ZIGP to any multi-modal distribution might be expected to be superior. However, the ZILP would produce a greater maximum likelihood when the vast majority of the data is described by either a single lognormal distribution or a lognormal distribution with other distributions contributing nearly all zero observations.

A multi-modal distribution may also be expected to arise as a result of differences between homozygotes and heterozygotes for a single gene that is important in the control of parasitism, with the continuous distributions surrounding these major effects representing the relatively minor effects of many other genes and environmental influences. A multi-modal distribution of multi-species nematode eggs may also be expected to arise as a combination of distributions describing the distribution of eggs of each species, although this explanation cannot be used with a single species of worms as with the observed *N. battus* data unless more than one population of worms was present. An alternative explanation is that the observed distribution of FEC is quite different to the distribution of worm burdens, resulting in an observed distribution of eggs that has a different shape to the underlying lognormal distribution of worm burdens. However, the four worm burden datasets available all showed evidence of zero-inflation using both models, and two of each produced a better fit to the ZILP and ZIGP distributions suggesting that similar effects are seen in worm burdens and FEC.

The suggestion of multi-modality in the data and inconsistent preferential fit between the applied distributions prevents any firm conclusions regarding the most appropriate distribution to use for the analysis of real data. The GP distribution is more flexible than the LP, and is better able to account for a large number of zero counts, leading to a better fit to the data observed here. However, this advantage is reduced when using the zero-inflated equivalent distributions and as a result neither distribution consistently fits the data better. One advantage of the ZILP is an apparent better identifiability, with less correlation between median estimates of cv and zero-inflation suggesting that the ZILP is better able to separate these effects than the ZIGP distribution. This may be a chance observation or a result of the model specification, but it is consistent with the greater flexibility of the gamma distribution. Ultimately, it may be the case that the underlying means of observed FEC data are not well described by either the gamma or lognormal distributions, either as a result of multi-modality, or the non-linear relationship between worm burden and mean FEC. In this case, inference made using either of the distributions presented may be inherently inaccurate.

The work presented in this thesis has so far been concerned with the application of novel statistical methods to FEC data obtained from a group of animals at a single point in time. However, FEC data are also frequently collected as paired samples taken approximately one week apart to assess the effect on parasite burden of dosing with anthelmintic. More specialised adaptations of the models presented here are required to analyse such data, and will be explored in the next chapter.

CHAPTER 5

Applying Bayesian MCMC to the Faecal Egg Count Reduction Test

Applying Bayesian MCMC to the Faecal Egg Count Reduction Test

5.1 Introduction

The FECRT is the most widely used method of assessing the *in vivo* efficacy of anthelmintics against parasitic nematodes of horses, sheep and cattle (Kaplan, 2002; Coles et al., 2006), and is an essential tool in the process of monitoring the increasing prevalence of anthelmintic resistance. Several methods based on arithmetic, geometric or harmonic empirical mean reductions calculated with or without the use of a control group have been proposed to analyse such data (Torgerson et al., 2005). Most of these work well in sheep, goats and cattle where large sample sizes, high means and low variability reduce the impact of statistical oversimplifications, but none work well for the small, highly variable datasets often encountered with equine data. These factors introduce an increased variability into equine FEC data (Uhlinger, 1993), leading to a relatively inconsistent FECRT result (Miller et al., 2006). Given the worldwide importance of anthelmintic resistance, there is an urgent need to improve and standardise the statistical method used to analyse equine FECRT data (Kaplan, 2002; Coles et al., 2006).

5.1.1 Control of equine nematodes

Control of the Strongyloid equine parasites is primarily achieved in the developed world via the use of interval anthelmintic dosing (Mercier et al., 2001), given to a group of horses at a frequency designed to reduce the number of infective larvae on the pasture (Craven et al., 1998), and therefore to limit the rate of infection of the herd. The principle involves dosing at a frequency sufficient to remove most of the egg producing adults, but to avoid wasting resources and precipitating the development of resistance (Mercier et al., 2001) by treating animals that are unlikely to be harbouring many adult worms (Uhlinger, 1991). It is also undesirable to render a horse completely free of worms because of the beneficial effects due to the buildup of immunity. However, total elimination of the parasite is frequently the goal for many owners (Uhlinger, 1993).

The pre-patent period of the parasite is a measure of the time it takes for an ingested L₃ larvae to develop to an adult and start producing eggs (Urquhart et al., 1996), and is important to consider when designing efficient dosing strategies. The effects of worm parasitism have been demonstrated to show seasonal variation, and it is known that in some parts of the world the density of infective larvae on pasture is greater at some times of year than others (Ogbourne, 1976; Herd, 1986; Langrová, 1998; Baudena et al., 2000).

The susceptibility of individual animals to worms is not uniform; it is well understood that younger and older horses are more frequently heavily parasitised than adult horses (Uhlinger, 1993; Döpfer et al., 2004), and some individuals appear to show greater susceptibility to infection than others (Duncan and Love, 1991). There is some evidence to show that this may be at least partly due to the effects of immunity (Duncan and Love, 1991; Klei and Chapman, 1999). Some authors advocate the specific targeting of the persistently high shedding animals for anthelmintic treatment and worm burden monitoring. This could potentially reduce the costs of parasite control and slow the development of resistance to anthelmintics (Döpfer et al., 2004; Nielsen et al., 2006a). Such an approach has also been suggested in other species such as dairy goats (Hoste et al., 2002). More strategic attempts to control parasitism based on the known seasonal epidemiology of equine parasites have also been suggested (Herd et al., 1985). The optimum dosing interval therefore depends on many factors and as a result is unique to each group of animals (Uhlinger, 1991). This has resulted in the use of varying control strategies between premises (Lendal et al., 1998; O'Meara and Mulcahy, 2002). Many of the strongyle control programs currently practised fail to safeguard the health and welfare of the horses they intended to protect; either because of inappropriate timing, drug choice (Herd, 1986), an inability to identify the truly high shedding animals, or due to the increasing problem of anthelmintic resistance (Mercier et al., 2001).

The broad spectrum anthelmintics currently used in equine medicine all originate from one of the following three classes of drug; the benzimidazoles, pyrantel/piperazine, and the avermectins/milbemycins (O'Meara and Mulcahy, 2002). The efficacy of these drugs are known to differ (Pook et al., 2002), with the avermectins/milbemycins and benzimidazoles generally acknowledged to have an efficacy of greater than 95% (Klei and Torbert, 1980; Malan et al., 1981), whereas that of pyrantel has been estimated to be in the range of 90-92% (Lyons et al., 1974; Dargatz et al., 2000). Efficacy of anthelmintics against the pre-adult tissue-dwelling stages of parasite is also dependent on the anthelmintic; both a five-day consecutive treatment with oxibendazole (Dargatz et al., 2000) and, in susceptible worm populations, treatment with fenbendazole (McBeath et al., 1978) are known to be more efficacious than the avermectins/milbemycins and pyrantel/piperazine against larval stages (Bauer et al., 1986).

Alternative management techniques for the control of equine nematodes such as removal of faeces from pasture to reduce grazing contamination are also effective (Uhlinger, 1991).

Twice weekly removal of faeces from pasture has even been shown to control parasite burden more effectively than regular anthelmintic dosing (Herd, 1986), and is often advocated as a worthwhile control measure (Coles, 2002; Comer et al., 2006). While faeces removal would not necessarily be expected to remove all larvae from pasture, the density of infective larvae on pasture and therefore larval challenge to animals should be greatly reduced. In addition, the pasture contamination from recently dosed animals, which would be more likely to contain larvae showing resistance to anthelmintic, is also reduced. Complementary veterinary medicine has also been used to attempt to control equine parasites, although the beneficial effects of such treatments remain unproven (Lloyd and Martin, 2006).

5.1.2 Development of anthelmintic resistance in equine parasites

Over the last two decades, an increase in resistance to commonly used anthelmintics has been observed both in the number of host species affected and the number of drugs where resistance has been reported (Sangster, 1999b). Resistance of parasites to chemical substances is a natural phenomenon that in some cases pre-dates the use of the chemical as a therapeutic agent (Dargatz et al., 2000), and the effects can be manifest in one of several ways. Parasites are naturally resistant to many therapeutic agents, either as an inherent lack of susceptibility during larval or encysted stages, differences in susceptibility between different species of parasite, or as a tolerance to the drug at lower concentrations (Sangster, 1999a). However, resistance also represents the ability of a parasite to evolve when faced with a selection pressure; changing its own physiology to resist the effects of the therapeutic agent. Separating these two phenomena in order to assess the development of anthelmintic resistance within a parasite population is not straightforward (Sangster, 1999a).

The development of anthelmintic resistance is a gradual process which often progresses in stages (Dargatz et al., 2000), ultimately leading to the situation where an anthelmintic drug is no longer as effective in the same concentration as it was when first used in the population. The rate of development depends on both the genetics of the nematode, and the selection pressure for resistance which is affected by the efficacy and intensity of use of the drug (Bauer et al., 1986; Dargatz et al., 2000), and the proportion of the parasites that are ‘in refugia’ (Matthews, 2008). It has been shown that there is often no reversion to susceptibility once resistance has developed (Lind et al., 2007). Definitions of some concepts relevant to anthelmintic resistance can be found in Table 5.1 (taken from Dargatz et al. (2000)). Prichard et al. (1980) define resistance in a similar way:

“Resistance is present when there is a greater frequency of individuals within a population able to tolerate doses of a compound than in a normal population of the same species and is heritable”

Side-, cross- & multiple-resistance, and reversion are also defined in a similar way to those given in [Table 5.1](#).

In equine parasites, anthelmintic resistance is almost exclusively seen in cyathostomins ([Coles et al., 1992](#); [von Samson-Himmelstjerna et al., 2002](#)). Resistance to the benzimidazoles was first reported in cyathostomins in 1965 ([Pook et al., 2002](#)), and has resulted in a gradual decline in the effectiveness of equine parasite control ([Mercier et al., 2001](#)). Benzimidazole resistance is now common in the domestic horse population ([Ihler and Bjorn, 1996](#); [Chandler and Love, 2002](#); [Kaplan et al., 2004](#); [Traversa et al., 2009](#)), although susceptible cyathostomins can still be found in feral animals ([Young et al., 1999](#)). There are few reports of anthelmintic resistance within the other strongyle nematodes of horses, although there have been recent reports of apparent lack of efficacy of ivermectin against ascarid worms in young horses ([Schougaard and Nielsen, 2007](#); [Lind and Christensson, 2009](#); [Reinemeyer, 2009](#)). However, it is not entirely clear whether this was attributable to anthelmintic resistance *per se* or other factors associated with ascarid infections in young horses.

Recently, the apparent reduction of efficacy of pyrantel in horses has been reported, firstly in the USA ([Tarigo-Martinie et al., 2001](#); [Kaplan, 2002](#)) but later also in Europe ([Ihler, 1995](#); [Comer et al., 2006](#); [Lind et al., 2007](#); [Traversa et al., 2009](#)). The issue of whether pyrantel resistance has been found is somewhat controversial, especially given the lack of standardisation between tests for anthelmintic resistance ([Craven et al., 1998](#)) and the reportedly low initial efficacy of pyrantel ([Lyons et al., 1974](#); [Dargatz et al., 2000](#)). However, apparent resistance to pyrantel has been reported ([Traversa et al., 2009](#)). Given the high dependence on Pyrantel within the equine community, this is of great concern. To date, little evidence for resistance to the avermectins/milbemycins has been reported in horses ([Ihler, 1995](#); [Coles, 2002](#)), although one case of suspected Moxidectin resistance has been reported in donkeys ([Trawford et al., 2005](#)), and one case of reduced efficacy of ivermectin in horses has been reported in Germany ([Traversa et al., 2009](#)). [Traversa et al. \(2009\)](#) also found several cases of multiple anthelmintic resistance in several countries.

Anthelmintic resistance has had a major economic impact in both equine and farm animal medicine ([Waller, 1997](#)). It will also become a potential welfare issue as resistance inevitably develops to the currently relied-upon classes of drug ([Lloyd, 1998](#); [Nielsen et al., 2006a](#)). Emerging resistance in all available anthelmintic classes has led to calls for the development of novel anthelmintics ([Kaplan, 2004](#); [Besier, 2007](#)), with a new class of drugs isolated recently ([Kaminsky et al., 2008](#)) appearing to be feasible for use in animals ([Mason et al., 2009](#)). Anthelmintic resistance appears to be increasing steadily in prevalence ([Waller, 1997](#); [Kaplan et al., 2004](#)), despite increasing awareness of the problem. Extensive efforts are therefore required to reverse this trend by using good management practises, more sparing use of anthelmintics ([Pascoe et al., 1999](#); [Coles, 2002](#)), appropriate drug selection and adequate monitoring for the emergence of resistance ([Lloyd, 1998](#); [Comer et al., 2006](#)).

Table 5.1: Terms and definitions used in describing anthelmintic resistance, taken from Dargatz et al. (2000)

Term	Definition
Resistance	“... occurs when a greater frequency of individuals in a population of parasites, usually affected by a dose of concentration of compound, are no longer affected. Resistance is inherited”. Includes cases where higher doses or concentrations of drug are required to remove or kill a certain proportion of nematodes.
Side resistance	Parasites resistant to one drug of a chemical class are resistant to other drugs in the same class.
Multiple resistance	Parasites develop resistance to several anthelmintic classes independently.
Cross resistance	Describes resistance between chemical classes and has not been described for equine parasites.
Reversion	Loss of anthelmintic resistance. There are no reports of full reversion in anthelmintic resistance.
Clinical resistance (or field resistance)	Recommended doses of drug fail to remove parasites
Treatment failure	Clinical resistance is one form of treatment failure, but other reasons for apparent failure include diseases with the same signs, use of the wrong drug, under-dosing, and rapid reinfection.
Dose response	Concept of more drug killing more worms in a population. A plot of response (% dead) versus log drug concentration (in vitro) or log drug does (in vivo) is sigmoidal. The concentration, on average, that kills 50% of worms is the EC_{50} . The concentration at which 95% are killed is the EC_{95} .
Resistance factor	The ratio of EC_{50} between resistant and susceptible worms.

5.1.3 Methods for detecting resistance

In order to monitor the progression of anthelmintic resistance, it is necessary to test the susceptibility of worms to the anthelmintic in question. There are several techniques that can accomplish this *in vitro* such as the larval development assay (Pook et al., 2002), tubulin binding assay (Martin et al., 1989) and the egg hatch test (Martin et al., 1989; Coles et al., 1992), although these techniques are technically quite difficult and general availability for field studies is not very widespread. Currently the most widely used test is the *in vivo* FECRT (Cabaret and Berrag, 2004; Vidyashankar and Kaplan, 2005), the principle of which is to compare the faecal egg excretion rates of horses before and after dosing with anthelmintic (Martin et al., 1989; McKenna, 2006a). The degree of correlation between this test and the *in vitro* tests has been shown to be quite poor (Craven et al., 1999; von Samson-Himmelstjerna et al., 2002; Konigova et al., 2003), possibly as a result of the different tests evaluating different aspects of resistance (Craven et al., 1999).

To perform a FECRT, faecal samples are collected from individuals within a group of horses

and the number of eggs per gram of faeces are calculated using the McMaster method (described by [Coles et al. \(1992\)](#) and in detail by [CSIRO and Australian Agricultural Council \(1989\)](#)). Either the whole group or part of the group, leaving untreated controls, is then dosed with anthelmintic before re-sampling the faeces of all animals 10-14 days later. The faecal egg count reduction percentage, usually along with a lower confidence interval, is then produced from the data. This test has several limitations, including a low sensitivity ([Martin et al., 1989](#); [Johansen, 1989](#); [Coles et al., 1992](#)), the dependence of faecal egg count on factors such as faecal consistency ([Uhlinger, 1993](#)), the variation in fecundity between female worms ([Miller, 1953](#); [Uhlinger, 1993](#)), and the inability to speciate strongyle eggs ([Urquhart et al., 1996](#)). This is a potential cause of error when performing a FECRT, because the egg output of some horses has been shown to consist of as much as 2-8% large strongyle ova ([Ihler, 1995](#)), which will in effect dilute the egg output of resistant cyathostomins at the pre-treatment FEC and therefore increase the apparent efficacy of the drug against cyathostomins. However, the assumption that the detected faecal eggs are cyathostomin eggs is probably sufficiently accurate because of the dominance of cyathostomins in most equine premises where anthelmintics have been used in the past ([Uhlinger, 1991](#); [Warnick, 1992](#); [Fisher et al., 1992](#); [Ihler, 1995](#); [Martin-Downum et al., 2001](#); [Pook et al., 2002](#)). The repeatability of FEC has been assessed in both sheep ([Stear et al., 1995b](#); [Miller et al., 2006](#)) and horses ([Warnick, 1992](#)), and found to be poor. The variability due to the daily change in egg output and variations in laboratory technique ([Döpfer et al., 2004](#)) may be low enough that a single sample can be used to decide if an animal should be dosed, but is more important when considering changes in faecal egg count after treatment ([Warnick, 1992](#)). Faecal egg output is also not a direct measure of parasitism ([McKenna, 1981](#)); the fact that pre-adult parasites do not produce eggs and the considerable variation in fecundity between worms, means that the number of parasites within an animal cannot be determined from the faecal egg shedding rate ([Uhlinger, 1993](#); [Coles, 2002](#)). The faecal egg count test is also an extremely poor indicator of disease due to parasitism ([Uhlinger, 1993](#)), although the correlation between faecal egg count reduction and parasite reduction has not been studied.

There are also many different factors which affect the parasitic burden or egg shedding rate of the host, and complicate a faecal egg count reduction test. Faecal consistency, and therefore the volume and egg concentration of faeces ([Uhlinger, 1993](#)), is affected by diet and concurrent disease, which can also alter the parasite burden directly via the effects of stress ([Prichard et al., 1980](#)). Seasonal variation in egg output ([Herd et al., 1985](#); [Reinemeyer et al., 1986](#); [Uhlinger, 1993](#)) can also affect the results of the test, although this can also be compensated for by the use of control animals within the group. However, disadvantages associated with the necessary reduction of animals in the treated group may more than outweigh this benefit ([Coles et al., 1992](#)). Variation in egg output between animals due to age or differing grazing habits is also important ([Uhlinger, 1993](#); [Osterman Lind et al., 1999](#)), and is a strong argument for using pre- and post-treatment counts in each animal. Diurnal variation in faecal egg output was traditionally thought to occur ([Uhlinger, 1993](#)), although [Bennett \(1990\)](#) did not

find a significant diurnal variation in a study of ten horses. However, performing pre- and post- treatment sampling at the same time of day would be a sensible precaution. Despite all this, FEC are an excellent index of the contamination rate of pasture, the reduction of which is the main objective of anthelmintic dosing (Craven et al., 1998; Mercier et al., 2001), and are sufficiently easy to perform to provide an appropriate method for the detection of resistance (Martin et al., 1989; Fisher et al., 1992).

Although most authors agree on the laboratory method of the FECRT, the calculation of the percentage reduction and when to diagnose resistance in equine parasites has been the subject of much debate. Several different techniques have been used to calculate the reduction, all of which can give different results when applied to the same data (Maingi et al., 1996; Craven et al., 1998; Cabaret and Berrag, 2004; Torgerson et al., 2005; Miller et al., 2006, and demonstrated in Table 5.2). The cutoff value to distinguish a faecal egg count reduction from a resistant worm burden from that of a susceptible worm burden also varies, with some authors stipulating that a FEC reduction of less than 80% indicates resistance, while some authors prefer to use 90% (CSIRO and Australian Agricultural Council, 1989), and yet others use the 95% figure that is used in other species. Unlike the situation in sheep parasitology, where the cutoff figure of 95% has been demonstrated to be appropriate given a large enough flock size and a drug of efficacy > 99% (Coles et al., 2006), the values used in equine parasitology are somewhat arbitrary (Bauer et al., 1986; Martin et al., 1989; Kaplan, 2002; Vidyashankar and Kaplan, 2005).

Coles et al. (1992), in the World Association for the Advancement of Veterinary Parasitology guidelines, stipulate that the arithmetic mean be used according to Equation 5.1.

$$FECR = 100 \times \left(1 - \frac{t}{c}\right) \quad (5.1)$$

Where t is the arithmetic mean of the dosed group and c is the arithmetic mean of the control group 10-14 days after dosing. The advantages to using the arithmetic mean are listed by

Table 5.2: A comparison of the FEC reduction (with confidence intervals) as calculated using four different techniques for three equine FECRT datasets, adapted from Torgerson et al. (2005)

Technique	Menbendazole	Ivermectin	Moxidectin
FECR% (Coles et al., 1992)	60.9	39.8	97.6
95% CIs (Coles et al., 1992)	-13.686.5	-65.678.1	92.199.3
FECR% (Dash et al., 1988)	46.4	50.4	90.7
95% CIs (maximum likelihood)	-105.787.1	-40.281.0	21.598.2
FECR% (Presidente, 1985)	18.4	55.7	94.9
FECR% (maximum likelihood)	60.8	39.8	97.6
95% CIs (maximum likelihood)	8.981.8	-36.171.1	81.399.3

Coles et al. (1992) as:

1. easier to calculate
2. provides a better estimate of the worm egg output
3. provides a more conservative measure of anthelmintic efficacy than using the geometric mean

These reasons are arguably in the case of (1) irrelevant, and (3) not statistically valid as the arithmetic and geometric means are distinct and the two measures should have a different interpretation. The arithmetic mean of the sample is also more heavily affected by single large value than is the geometric mean. However, it is true that the population arithmetic mean provides a better estimate of the worm egg output than the population geometric mean.

The 95% confidence intervals are then calculated using Equation 5.2 (formula given in appendix of Coles et al., 1992).

$$\text{Confidence limit} = 100 \times \left(1 - \frac{t}{c} \times \exp(\pm 2.048) \times \sqrt{Y^2} \right) \quad (5.2)$$

Where Y^2 is the variance of reduction, calculated as:

$$Y^2 = \frac{\bar{\sigma}_2^2}{N * \bar{x}_2^2} + \frac{\bar{\sigma}_1^2}{N * \bar{x}_1^2}$$

Where \bar{x}_1 is the empirical mean pre-treatment FEC, \bar{x}_2 is the empirical mean post-treatment FEC, $\bar{\sigma}_1^2$ is the empirical variance in FEC pre-treatment, $\bar{\sigma}_2^2$ is the empirical variance in FEC post-treatment, and N is the number of animals.

The origin of the number 2.048 used to calculate the upper and lower confidence limits is not given by Coles et al. (1992), although it does correspond to the value taken from a t-distribution with 28 degrees of freedom. However, the potential justification for choosing 28 degrees of freedom are equally unclear. According to the guidelines, if both the lower confidence limit is below 90% and the FECR is below 95% then resistance is declared, and if either condition is met then resistance is suspected. These conditions have also been advocated by other authors (CSIRO and Australian Agricultural Council, 1989; Ihler, 1995).

An alternative, more complicated method is to use Equation 5.3 as advocated by Presidente (1985).

$$FECR = 100 \times \left(1 - \frac{t_2 \cdot c_1}{t_1 \cdot c_2} \right) \quad (5.3)$$

Where the subscript $_1$ denotes mean of the dosed or control group pre-dosing, and the subscript $_2$ that of the group post-dosing. However, this method requires sampling on two separate occasions, and has been shown to give similar results to the less complicated method above (Vizard and Wallace, 1987; McKenna, 2006b).

The third commonly used method is to use Equation 5.4 as advocated by McKenna (1990, 2006a).

$$FECR = 100 \times \left(1 - \frac{t_2}{t_1}\right) \quad (5.4)$$

Where all animals are included in the dosed group and single pre- and post-dosing counts are used. This method has the advantage that all animals are dosed, effectively increasing the sample size for the test (Coles et al., 1992; Pook et al., 2002) or decreasing the cost of the test by requiring fewer animals for the same diagnostic test sensitivity/specificity (McKenna, 2006b). Each of these three techniques can then be modified by using the geometric means in the calculations, for example Martin et al. (1989) advocate Equations 5.1 and 5.2 using the geometric mean. Egg counts from animals with a zero count pre-dosing (Martin-Downum et al., 2001), or in some cases from animals with a count of less than 150 eggs per gram (EPG) (Craven et al., 1998; Coles et al., 2006), are typically discarded before analysing the remaining data.

The faecal egg count reduction test has been standardised in sheep (Vercruyse et al., 2001; Vidyashankar and Kaplan, 2005), where large sample sizes increase the usefulness of the test and minimise the impact of using some animals as controls. The distribution of parasites within sheep is also likely to have a higher mean and reduced variability compared to that in horses, which improves the accuracy of the faecal egg count test (McKenna, 1981). In order to be able to apply parametric techniques, the data are often log transformed to improve the normality (Bauer et al., 1986; Osterman Lind et al., 1999), although even this often fails to normalise the distribution. In addition, it has been shown that the use of log-transformed data is more likely to produce both Type I errors, or incorrectly rejecting the null hypothesis, as well as Type II errors, or incorrectly accepting the null hypothesis, both for FEC data (Wilson et al., 1996) and FECRT data (Vidyashankar and Kaplan, 2005). The possibility of generating Type II errors is particularly concerning, as it could potentially lead to the false rejection of anthelmintic resistance from data. It is also more difficult to find suitable groups of horses to demonstrate anthelmintic efficacy in a controlled environment; many horse owners insist on dosing of their horses even if a FEC of zero has been demonstrated (Kaplan, 2002), and it is much more difficult to euthanize and perform a post mortem worm burden examination on a sample of horses than a sample of sheep. These factors may have contributed to the lack of standardisation between FECRT methods in the horse (Vidyashankar and Kaplan, 2005), despite the efforts of both the World Association for the Advancement of Veterinary

Parasitology (Coles et al., 1992) and the “International Co-operation on Harmonisation of technical Requirements for Registration of Veterinary Medicinal Products” (Vercruysse et al., 2002).

It is difficult to demonstrate resistance or susceptibility within a parasite population if the FECRT data are analysed inappropriately (Duncan et al., 2002). The use of point estimates for the reduction in faecal egg count takes no account of variability, and results in large confidence intervals which reduce the value of the results (Martin et al., 1989). Such methods also fail to take into account the aggregation of parasites within the population, which not only results in potentially valuable information being lost, but can also be a source of statistical error if not taken into account when analysing the results of FECRT (Morgan et al., 2005). The difficulty in analysing changes in low faecal egg counts has prompted some authors to advocate using only horses with an initial faecal egg count of greater than 150 EPG (Craven et al., 1998; Coles et al., 2006), which in some situations drastically reduces the number of horses available for study and could potentially bias the results. The need to distinguish treatment failure due to clinical resistance from treatment failure due to other causes, such as incorrect dosage and rapid reinfection, is also important for the detection of true anthelmintic resistance (Uhlinger, 1993). However, it is essential that these factors are overcome so that early and accurate determination of resistance can be achieved.

5.2 Development of an MCMC model for analysis of FECRT data

There are several different ways of formulating a MCMC model to analyse FECRT data. The majority of datasets include a pre-treatment count for each animal in a similar way to the data analysed in Chapters 2 to 4, and so the same distribution could be fitted to these data. A separate distribution could then be fitted to the post-treatment data, with a mean which has been scaled relative to the pre-treatment mean and a variance parameter that has been scaled relative to the pre-treatment variance parameter. This scale in mean would be used to infer the efficacy of the reduction, and most closely resembles the empirical mean reduction that is currently used to analyse such data. Either the lognormal or gamma distributions could be used to describe the distribution of mean counts, although use of the lognormal would make calculation of the change in geometric mean, equivalent to change in arithmetic mean on the log scale, simpler to calculate than the change in arithmetic mean, whereas the change in arithmetic mean is easier to calculate for the gamma distribution. Both the geometric and arithmetic means have been used by parasitologists, although often with no justification, so that either formulation could be chosen. However, it can be argued that the arithmetic mean is more useful on an epidemiological basis as it represents the pasture contamination rate more closely, and most published efficacy studies seem to be concerned with the arithmetic mean,

5.2 DEVELOPMENT OF AN MCMC MODEL FOR ANALYSIS OF FECRT DATA

so the change in arithmetic mean using a gamma formulation will be used. As before, the gamma-Poisson can be used both with and without a zero-inflated component, depending on biological reasoning with individual datasets. Typically, two counts, representing the pre- and post-treatment counts, are available from each animal from a FECRT, therefore a common group classification will be used for the two counts for each animal. This is to say that if an animal is part of the zero-inflated group at pre-treatment, then it must be part of the same group at post-treatment. Consequently, if a non-zero count is observed at either pre- or post-treatment then the animal must be classified as part of the infected group for both.

This formulation allows variability to exist both within and between animals, although neither are modelled explicitly and only the combined variability is modelled. The variability is allowed to change after treatment, therefore the variation between animals may change, possibly as a result of differing efficacy between animals, and the variability within an animal may also change, possibly as a result of variation in fecundity following treatment. Again, these individual sources of variability are not explicitly modelled, and are combined into the single change in variability parameter. It is not possible to separate these sources of variability with only a single observation per animal before and after treatment, but as they are accounted for in the model the estimate of efficacy should be robust to these effects. One drawback of this formulation is that the ‘paired’ effect of animal between pre- and post-treatment data is ignored, potentially resulting in a loss of power. This could be addressed by using a different formulation, whereby the same distribution is applied to the pre-treatment counts and a distribution of efficacy is applied to the count means, resulting in the means for the post-treatment observed counts. The variability sources are addressed slightly differently with this model, with the pre-treatment within and between animal variability accounted for in the same way before, and the shape of the distribution of efficacy reflecting the variability within animals before and after treatment as well as the change in variability between animals after treatment, i.e. the ‘true’ efficacy distribution. A beta distribution could be used for this efficacy distribution, which would constrain the individual mean changes to between zero and one. However, this may not be appropriate because the true mean of the post-treatment sample may truly be higher than that of the pre-treatment sample given the affect of within animal variability on this parameter. More appropriately, a gamma or lognormal distribution could be used which would allow the individual change in mean to be above one, representing an individual efficacy of below zero, but with a prior on the mean value restricting it to values of less than one, representing a mean efficacy of greater than zero. However, in pilot tests of this model formulation using both beta and gamma distributions on simulated data with values of dataset and mean taken from real data, extremely poor convergence was observed. This may be due to model identifiability issues associated with the multiple sources of variability associated with the efficacy distribution. It is also possible that the comparatively good convergence of the first model formulation is as a result of inferring the post-treatment variability, where there is comparatively little information as a result of the lower mean counts, from the pre-treatment variability where there is more

information. Pilot tests of model formulations using totally separate values of variability for pre- and post-treatment data resulted in worse convergence for the post-treatment variance parameter compared to that observed for the change in variance parameter when this is modelled explicitly. The shape of the distribution of efficacy is not directly comparable to the pre-treatment variance parameter, so it is not possible to aid identifiability by reformulating the model using a distribution of efficacy in this way. In addition, the apparent distribution of efficacy could arise as a combination of several other factors, including the aggregated distribution of eggs in faeces, variations in worm fecundity over time, variations in faecal consistency, and variations in the numbers of worms present, which are impossible to separate using only a single faecal sample per individual. This model formulation was therefore not pursued further here, but may be of use in determining the true variability in efficacy between animals in larger datasets containing repeated measures to improve identifiability.

The formulation of the final FECRT model based on two zero-inflated gamma-Poisson distributions is given in [Model 5.1](#). The priors used for pre-treatment mean and shape are consistent with those used in [Chapter 2](#). The change in mean is given a $\text{beta}(1, 1)$ prior to restrict post-treatment mean egg count to less than pre-treatment mean egg count, because it would not be expected for true mean egg count to increase after treatment given the short period of time being considered. The change in shape parameter is given a diffuse lognormal prior with a mean of zero and precision of 0.01, so that values of less than one and more than one are equally likely. This parameter is truncated to between 0.001 and 1000 to prevent sampler failures associated with extreme values, and the same lower truncation of 10^{-200} is used for the gamma distribution as discussed in [Section 3.6](#). Where zero-inflation is used, the pre- and post-treatment group assignment is consistent for each animal.

5.3 Comparison of three methods for analysis of FECRT data

5.3.1 Introduction

Analysis of FECRT data can be performed in one of several different ways. The method currently advocated by the World Association for the Advancement of Veterinary Parasitology (WAAVP) involves calculating the empirical mean and variance before and after treatment, and calculating the empirical mean proportional reduction and estimates of the 95% confidence interval for the true reduction according to [Equation 5.2](#) ([Coles et al., 1992](#)). This method takes no account of the difference between uncertainty regarding the true mean of a sample, introduced by the Poisson variability of the counting process, and variability in the true mean of different samples. Calculation of 95% confidence intervals in this manner also makes the assumption that the distribution of error for the true mean reduction is symmetrical about the empirical mean reduction.

A non-parametric bootstrapping approach has recently been suggested as an appropriate method to generate confidence limits from equine FECRT data (Vidyashankar et al., 2007). The technique involves re-sampling and summarising the observed data to estimate parameters of interest (Efron, 1979), and does not require knowledge of the distribution or processes generating the data (Mooney and Duval, 1993). Non-parametric bootstrapping approaches are therefore useful when the distribution of data is unknown. The assumption underlying this approach is that the data obtained are completely representative of the full distribution describing the population, which is potentially violated when dealing with small sample sizes leading to misleading results.

Compared to the case in ruminants, the relatively small sample size, low pre-treatment mean FEC, and frequent zero FEC observations typically encountered with equine FECRT data (Kaplan, 2002; Nielsen et al., 2006a) may provide a challenge to the use of a non-parametric bootstrapping procedure, because there are relatively few data points from which to sample. Computationally intensive parametric methods have an advantage in this regard in that the data are fit to a distribution, so that not all possible values have to be observed in order to be sampled, although there is also often insufficient data to be able to analyse the underlying distribution, which prevents validation of the choice of distribution used by a parametric analysis such as MCMC. The aim of this study was to assess the usefulness of 95% confidence intervals generated by applying these three methods to simulated data, and then to assess the impact of the assumptions being made for each method.

5.3.2 Materials and Methods

5.3.2.1 Simulated data generation

A total of 1000 parameters for a simulated FECRT were generated in the R statistical programming language. The true FEC reduction was simulated from a Uniform(0.75, 1) distribution, so that true egg count reductions varied in a range that is likely to be observed clinically. The pre-treatment mean number of eggs counted and sample size were chosen to reflect the values seen in real equine FECRT data obtained from the 64 Danish equine datasets presented in Section 4.3. Pre-treatment mean number of eggs counted was taken from a Uniform(1.45, 53.1) distribution, and sample size per group was sampled randomly from integers between 4 and 16. The cv between samples before treatment was sampled from a Uniform(1, 1.41) distribution, corresponding to a pre-treatment shape parameter of the gamma distribution, k , of between 1 and 0.5. The proportional increase in cv after treatment was sampled from the same distribution. These values were also chosen to reflect the values most likely to be encountered in real FECRT data; published values of k are usually less than one (Shaw et al., 1998), and differing efficacy of anthelmintic between animals would

be expected to result in an increase in variability post-treatment unless such differences were non-random due to a systemic effect of animal age or parasite burden, for example.

In order to test the implications of the distributional assumptions made by the MCMC and WAAVP methods, simulated datasets were generated using the following three different distributions of underlying sample means; gamma, multi-modal lognormal, and uniform. The observed count was modelled as a Poisson process on top of these means. For each dataset, the meta-population mean and variance was the same for all distributions. The gamma-Poisson (negative binomial) distribution was parametrised by the pre-treatment means and shape parameters simulated above for the pre-treatment data, and the product of the pre-treatment means and one minus the true reduction in faecal egg count, and the product of the pre-treatment shape parameter and the change in shape parameter, for the post-treatment data. For the multi-modal data, the number of groups was sampled as between two and ten for each dataset, and the mean of each group was sampled from a lognormal distribution with mean equal to the population mean. Further lognormal distributions were then used to describe the distribution of samples within each group, with the number of animals in each group taken from a multinomial distribution with equal probability of assignment to each group for each animal. The cv for each lognormal distribution within the compound distribution was calculated according to the two conditions that each value is equal (representing an equal contribution of variability from both between and within sub-group), and that the total population variance would be equal to the total population variance of a single gamma distribution with the given mean and shape parameter. This approximation does not hold when the number of modes is not equal to the number of animals because the variability between animals from the same mode is less than that between animals from different modes. This should not affect the comparisons because the ability of the models to correctly identify the true variance is not being tested. Equation 5.5 was used to calculate the effective cv of the population given the cv of two compound distributions. The derivation of this equation from the combination of two gamma distributions is given in Appendix B. The lognormal distribution is similar in shape to the gamma distribution, so the error introduced by using the lognormal distribution in its place will be relatively small and should also not affect the inference made.

$$cv_{total} = \sqrt{cv_1^2 + cv_2^2 + (cv_1^2 * cv_2^2)} \quad (5.5)$$

Where cv_{total} is the effective coefficient of variation (cv) of the population, cv_1 is the cv in the first distribution, and cv_2 is the cv in the second distribution.

For the final distribution, the lower (L) and upper (U) limits of the uniform distribution were calculated using Equation 5.6 where μ and cv are the population mean and coefficient of variation, respectively. If the calculated value for L was less than zero, then a log-uniform

distribution (a distribution which is uniform on the log scale) was used in place of the uniform distribution. In this case, an optimisation algorithm was used to find a solution to Equation 5.7 that fit the pre/post-treatment values of μ and cv generated.

$$\text{Lower and Upper limits} = \mu \pm \frac{cv \times \mu \times \sqrt{12}}{2} \quad (5.6)$$

$$U = 2 * \frac{\mu^2 + (cv * \mu)^2}{\mu - L} \quad (5.7)$$

$$\mu = \frac{U - L}{(\log U - \log L)}$$

The log-uniform distribution was only used when the uniform distribution would have required $L < 0$, so that it was possible for pre-treatment data to be of a different distribution to post-treatment data.

Pre- and post-treatment egg count data were generated using each of these three distributions with the 1000 parameter values, to simulate a FECRT for a total of 3000 datasets. These datasets were then analysed using each of the three techniques.

5.3.2.2 Statistical Analysis

The analysis currently recommended by the World Association for the Advancement of Veterinary Parasitology was performed as described by Coles et al. (1992). The median estimate along with upper/lower 95% confidence intervals of the mean reduction are calculated using Equations 5.8, taken from the appendix of Coles et al. (1992).

$$Y^2 = \frac{\bar{\sigma}_2^2}{N * \bar{x}_2^2} + \frac{\bar{\sigma}_1^2}{N * \bar{x}_1^2}$$

$$\text{mean estimate} = 100 * \left(1 - \frac{\bar{x}_2}{\bar{x}_1} \right) \quad (5.8)$$

$$\text{confidence interval} = 100 * \left(1 - \left(\frac{\bar{x}_2}{\bar{x}_1} * \exp \left(\pm 2.048 * \sqrt{Y^2} \right) \right) \right)$$

Where \bar{x}_1 is the empirical mean pre-treatment FEC, \bar{x}_2 is the empirical mean post-treatment FEC, $\bar{\sigma}_1^2$ is the empirical variance in FEC pre-treatment, $\bar{\sigma}_2^2$ is the empirical variance in FEC post-treatment, Y^2 is the variance of the reduction, and N is the number of animals.

Bootstrapping was conducted using a function written in the R statistical programming language. For each dataset, new pre-and post-treatment pseudo-datasets were sampled with replacement to the size of the original data 10,000 times. At each iteration the mean reduction was calculated as the mean of the post-treatment pseudo-data divided by the mean of the pre-treatment pseudo-data. The median and 95% confidence intervals of the mean reductions for each dataset were then calculated and recorded from these 10,000 iterations.

Bayesian MCMC analysis was performed using [Model 5.1](#), implemented using JAGS. The model fits a gamma-Poisson distribution to the pre and post-treatment raw count data, with parameters for pre- and post-treatment means and shape parameters. None of the data were simulated with zero-inflation components, so zero-inflation was not used in the model. The priors used for pre-treatment mean and shape are consistent with those used in [Chapter 2](#). Post-treatment mean and shape parameters are calculated by multiplying the pre-treatment mean and shape parameters by a “change in mean” and “change in shape” parameter, respectively. From this model, estimates of the mean anthelmintic efficacy and the variability in anthelmintic efficacy between animals can be obtained. The “change in mean” is given an uninformative Beta(1, 1) (equivalent to Uniform(0, 1)) prior, and the “change in shape” a diffuse lognormal prior with a mean of one. This allows inference on the true change in mean egg shedding, with an additional parameter reflecting the true change in variability between egg counts. The true % FEC reduction is derived from $(1 - \text{change in mean}) * 100$. Calling JAGS to run each simulation and summarising of MCMC chains was automated using the `runjags` package for R, with two chains. Convergence was assessed using the Gelman-Rubin statistic ([Gelman and Rubin, 1992a](#)), and necessary sample size using Raftery and Lewis’s diagnostic ([Raftery and Lewis, 1995](#)). The median estimate and 95% credible intervals for the true egg count reduction were calculated in R using the MCMC output.

For all techniques, credible intervals for the true proportion of datasets contained within 95% confidence intervals were calculated using a Bayesian approach with an uninformative Beta(1, 1) prior. The mean relative size of these confidence intervals was calculated using [Equation 5.9](#).

$$\text{confidence interval size} = \frac{\sum \frac{U-L}{T}}{N} \quad (5.9)$$

Where L denotes the lower confidence interval, U the upper confidence interval, T the true parameter value, and N the number of datasets

To assess the accuracy of the median estimates, the relative RMSE was calculated using the simulation value for each parameter as before.

5.3.3 Results

The relationship between the simulated and empirical % reductions is shown in [Figure 5.1](#). A degree of correlation is apparent, but with substantial variability. The uniform-Poisson distribution appears to have produced the most extreme deviations. Of the 3000 datasets, 100 (3.3%) gave a mean reduction of 100% using the WAAVP method. The median (95% simulation interval) simulated reduction for these empirical 100% reduction datasets was 99.13% (82.23% - 99.97%). As the post-treatment variance for these datasets was zero, the WAAVP method of calculating 95% confidence intervals could not be applied. In practice, these datasets would be assumed to represent a 100% reduction, so 95% confidence limits of 100% - 100% were assigned to these datasets. The non-parametric bootstrapping approach also generated confidence limits of 100% to 100% for these datasets, as all possible combinations of datapoints give a 100% reduction. Results are shown both with all 3000 datasets, and with these 100 datasets excluded.

In [Figure 5.2](#), the proportion of true reductions that were contained within the notional 95% confidence intervals for each method are shown, both for all datasets and with the 100 datasets that demonstrated a mean reduction of 100% removed. There is no evidence that the MCMC method did not estimate true 95% confidence intervals for both the gamma-Poisson and (log) Uniform data, but the confidence was lower for the multi-modal data. Non-parametric bootstrapping and the WAAVP method both returned notional 95% confidence intervals that

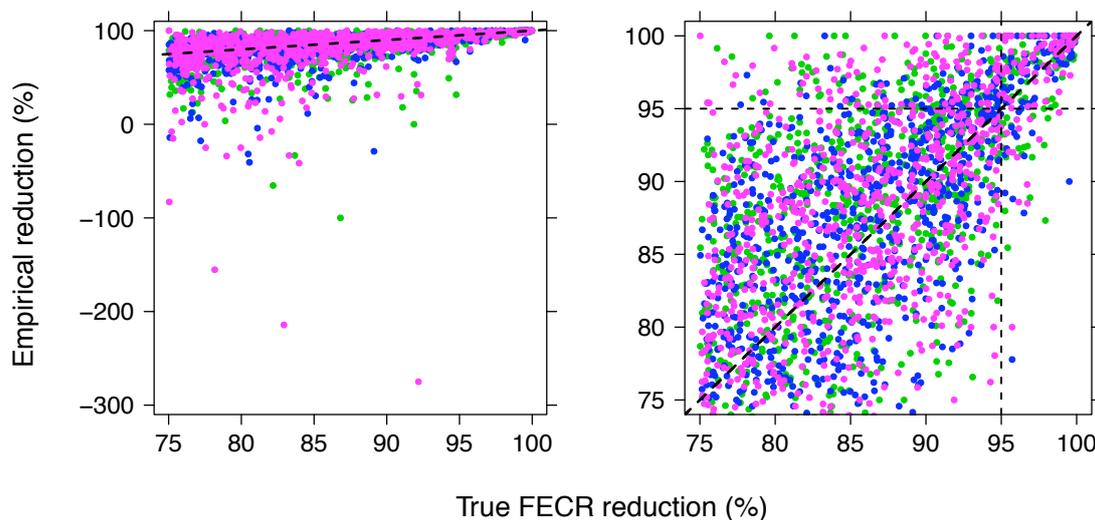


Figure 5.1: The relationship between the simulated % reduction and the empirical mean % reduction from 3000 simulated FECRT datasets - all values shown on left, some omitted on right. Gamma-Poisson data shown in green, multi-modal in blue and uniform-Poisson in pink. Diagonal dotted line denotes 1:1 correlation, vertical and horizontal dotted lines on right hand side denote limit of True/False positive and True/False negative results (assuming an efficacy threshold of 95%)

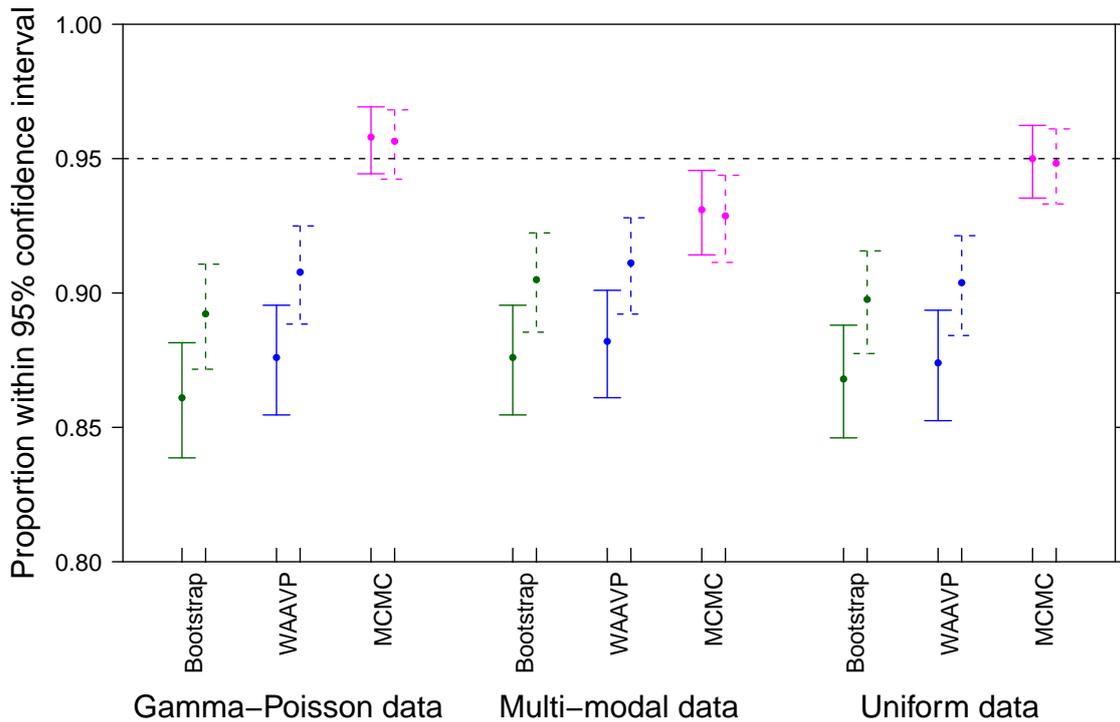


Figure 5.2: The proportion of 95% confidence intervals containing the simulated true FEC reduction parameter for each method from the analysis of all simulated FECRT datasets (solid lines) and the 2900 simulated datasets that did not give an empirical mean reduction of 100% (dashed lines)

would be expected to contain the true value between 85% and 90% of the time for all data types, although this improved to between 88% and 93% by removal of datasets with a mean reduction of 100%.

The mean relative size of the notional 95% confidence intervals for each method, both with and without the datasets that demonstrated a 100% empirical reduction, are shown in Tables 5.3 and 5.4, and the relative RMSE for each combination in Tables 5.5 and 5.6. The MCMC median estimates produced a comparable relative RMSE to the WAAVP and bootstrap estimates with the gamma-Poisson and multi-modal data, but lower relative RMSE for the uniform-Poisson data. Confidence interval sizes were comparable for all techniques with each data type.

5.3.4 Discussion

For all datasets, simulated from each of the distributions tested, the MCMC method provided confidence intervals with the best defined properties, as well as the most precise median estimates for the true FEC reduction when datasets with 100% reductions were removed.

Table 5.3: Mean size of 95% confidence intervals relative to the median (MCMC and bootstrap) or mean (WAAVP) estimate of the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of 3000 simulated FECRT datasets

	Gamma-Poisson	Multi-modal	Uniform-Poisson
Bootstrapping	0.702	0.459	0.803
WAAVP	0.673	0.532	0.786
MCMC	0.750	0.554	0.808

Table 5.4: Mean size of 95% confidence intervals relative to the median (MCMC and bootstrap) or mean (WAAVP) estimate of the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of the 2900 simulated FECRT datasets that did not give an empirical mean reduction of 100%

	Gamma-Poisson	Multi-modal	Uniform-Poisson
Bootstrapping	0.728	0.475	0.830
WAAVP	0.697	0.550	0.813
MCMC	0.741	0.541	0.802

Table 5.5: Relative root-mean-square-error (RMSE) for median (MCMC and bootstrap) or mean (WAAVP) estimate for the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of 3000 simulated FECRT datasets

	Gamma-Poisson	Multi-modal	Uniform-Poisson
Bootstrapping	1.69	1.49	2.60
WAAVP	1.70	1.53	2.66
MCMC	1.83	1.64	1.99

Table 5.6: Relative root-mean-square-error (RMSE) for median (MCMC and bootstrap) or mean (WAAVP) estimate for the true FEC reduction produced by the MCMC, bootstrap and WAAVP techniques from analysis of the 2900 simulated FECRT datasets that did not give an empirical mean reduction of 100%

	Gamma-Poisson	Multi-modal	Uniform-Poisson
Bootstrapping	1.72	1.51	2.64
WAAVP	1.73	1.55	2.70
MCMC	1.69	1.50	1.87

The size of the 95% confidence intervals produced was no greater for the MCMC method when datasets with 100% reductions were removed, indicating that the MCMC methods were producing more appropriate 95% confidence intervals rather than merely larger 95% confidence intervals. The relative RMSE and size of 95% confidence intervals was greater for each method with the Uniform-Poisson data than the other datasets; this is possibly a result of the shape of the Uniform or log-Uniform distribution creating more frequent moderately large values than would be expected with a gamma or lognormal distribution with the same variance. Interestingly, the relative RMSE of the MCMC method seemed to be less affected by the Uniform-Poisson data than the other methods. The reasons for this were not clear, and it may be a chance observation, or it may be due to the MCMC method accounting for these moderately large values with an increased estimate for the variance. Conversely, the relative RMSE and size of 95% confidence intervals was smallest for each method with the multi-modal data. This is most likely as a result of the effective population variance being reduced by the clustering effect of the multi-modal distribution, as follows. The variance of the two lognormal distributions contributing to the multi-modal distribution was calculated so that the total variance of the compound distribution would be equal to the variance used for the uni-modal distributions. The effective total variance of the multi-modal distribution would only equal this if one animal was used per group, so that each animal was independently sampled from the compound distribution. However, more than one animal was used per group, which results in a reduction of the total effective variance because animals are no longer independently sampled from the compound distribution. This would on average reduce the overall variance for the multi-modal distribution compared to the uni-modal distributions, but does not affect the conclusions drawn.

The MCMC method out-performed the WAAVP and bootstrapping methods not only for data simulated from a gamma-Poisson distribution, where the MCMC method using the same distribution would be expected to perform well, but also using data simulated from different distributions. The performance of the MCMC method was less optimal using the multi-modal data, but even here it out-performed the other two methods. This may be due to the distributional assumptions made by the MCMC method having less practical impact on the results of the analysis of these types of FECRT data than the assumption that bootstrapping a limited number of data points can capture all the variability of an inherently very variable system. [Vidyashankar et al. \(2007\)](#) proposed dealing with this effect by taking into account inter-farm variability, in effect increasing the sample size of the study. It is likely that this would improve the confidence of the bootstrapping method as long as the total effective sample size is sufficient. The MCMC method is also capable of analysing data from multiple sites, for example by defining a distribution of efficacy that describes the mean FEC reduction at each site and using this extra information to reduce uncertainty in the estimate for the true mean efficacy. Where data from multiple sites are available, and it is reasonable to assume a distribution of efficacy describing the sites, the additional data can be taken into account in order to reduce uncertainty in parameter estimates. However,

the intention here was to assess the performance of each method when analysing individual datasets in the absence of any other comparable datasets, so that taking into account inter-farm variability would not have been possible. Based on the results presented, parametric techniques appear to out-perform non-parametric bootstrapping in this regard when sample sizes are small. It is also possible to perform the bootstrapping calculation differently to that used here; for example bootstrapping the ratio between paired pre- and post-treatment counts rather than the pre- and post-treatment pseudo-datasets themselves. This may have altered the performance of the bootstrapping technique, but less closely resembles the formulation of the MCMC model used and therefore makes comparison of the techniques themselves more difficult.

The field equine data used to inform the parameter generation for the simulations were obtained from Equilab Laboratory, and represented voluntary submission of FECRT data from a wide variety of equine establishments across Denmark. As such, the data are representative of equine establishments in Denmark, although it could be argued that the distribution of mean FEC may be different in Denmark to other European countries and the USA as a result of the Danish regulatory requirements for prescriptions for anthelmintics. The relative performance of the bootstrapping and WAAVP procedures would be expected to be worse in other countries if a tendency towards smaller herd sizes and/or lower mean FEC exists in Denmark. For example, the mean pre-treatment FEC of the data presented in [Schumacher et al. \(2009\)](#) is 11 EPG which is much lower than the median (95% CI) value of 217 EPG (115 - 863 EPG) in the Danish FECRT data presented here.

Several of the datasets generated with parameters similar to observed equine FECRT data gave an empirical reduction of 100%, even where the true mean reductions were close to 75%. These datasets present difficulties when using both the WAAVP and bootstrap methods, which are unable to generate appropriate 95% confidence limits. Nineteen (19%) of these datasets were simulated using empirical reductions of less than 95%, and so represent a consistent source of false negatives for these methods. The MCMC method was the only method examined here which is capable of analysing datasets with 100% empirical reductions in an appropriate fashion, but analysis of these datasets still poses some difficulties. The lower 95% confidence interval for efficacy is substantially decreased when analysing data with a post-treatment dataset of all zero relative to post-treatment data with a single egg counted, with the same pre-treatment data. This may superficially appear as though the model is not appropriately analysing the data, but this phenomenon is related to the issue of analysing all zeros datasets discussed in [Section 2.7](#). When no eggs are observed post-treatment, the model has no information from the data on which to base the estimate for post-treatment cv , or change in cv . This means that it is possible that the post-treatment data have a relatively high mean combined with a large cv which is resulting in zero count observations. This effect is likely to worsen with smaller sample sizes. The prior for efficacy is Beta(1,1), so the effective prior for the post-treatment mean would tend towards values

uniformly between zero and the pre-treatment mean. In order to account for this prior, the estimate for post-treatment cv is inflated to increase the likelihood of observing zero counts with a high mean. When more than one or two low counts are observed post-treatment, the likelihood of a high cv decreases, so that the model tends towards lower change in cv and consequently lower post-treatment mean. This behaviour can be addressed by placing a more informative prior on the post-treatment cv to make such high values less likely, or a prior on efficacy that tends towards higher values. However, this may lead towards erroneous inference unless these priors can be properly justified. Fundamentally, it is quite possible that such datasets really do represent instances of extremely poor efficacy along with high variability. The reason that this initially seems strange to the casual observer is that the prior belief when carrying out a FECRT is that efficacy is closer to 100% than 0%. The inference made by the MCMC model is therefore appropriate given the priors, which are conservative in the respect that the data is required to prove acceptable efficacy. Alternatively, it may be possible to derive information from the magnitude of the change in variability to infer that the efficacy cannot be close to zero; it may be illogical to believe that the efficacy is truly zero if a change in variability is observed due to the treatment. The model specification could be altered in some way to reflect this, although as the confidence intervals for the change in efficacy were generally large it may provide little practical benefit.

The model formulation presented here uses a separate pre- and post-treatment distributions to describe the distribution of counts, with the efficacy inferred from the ratio in means. As discussed previously, an alternative formulation could calculate a distribution of efficacy values derived from the paired pre- and post-treatment count data from each animal, with the mean efficacy calculated from this. Such a formulation would require either a compound distribution separating within and between animal variability, or a fixed effects model with an appropriate error structure, such as gamma-Poisson, relating the observed counts to true animal means. In pilot studies of such formulations the lack of identifiability between the change in within animal variability and shape of the efficacy distribution precluded their use for the available data, although their use may well be preferred given sufficient information in the data. This model formulation will be presented later in this chapter as a method of analysing these data at a population level.

5.4 Further Analysis of Non-Parametric Bootstrapping

5.4.1 Introduction

It is apparent from the analysis presented previously that the bootstrapping procedure discussed produces unreliable 95% confidence estimates of true mean FEC reduction with the sample sizes typically encountered in equine FECRT data. The usefulness of the method is

likely to be increased with increasing sample size and mean FEC; an analysis of this relationship would therefore be useful to determine the minimum values at which the bootstrapping procedure can be used without making erroneous inference.

5.4.2 Materials and Methods

Pre- and post-treatment egg count data were generated using two gamma-Poisson distributions, to simulate a FECRT. For this analysis, sample size and mean FEC were fixed for each dataset so that the effect of increasing sample size and mean FEC on the performance of the bootstrapping method could be assessed. Sample size was drawn from the set $\{5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, and pre-treatment mean number of eggs counted from the set $\{1, 5, 10, 20, 30, 40, 50, 75, 100\}$, so that the range covered the most likely values observed in field FEC data. The pre-treatment cv was chosen from the set $\{1, 1.41\}$ to reflect values commonly seen with equine FECRT, and the change in cv post-treatment chosen from the set $\{1, 1.41\}$ simulating either no change in the cv after treatment, or an increase in cv between animals after treatment. Each of these 396 combinations was used to generate 250 datasets using two gamma-Poisson distributions and a true FEC reduction randomly generated from a Uniform(0.75, 1) distribution. Each dataset was analysed using the bootstrap method to provide a median and 95% confidence interval estimate as before. The 95% confidence interval produced by the bootstrap analysis was compared to the simulation parameter for each dataset, and the number of analyses where the true value was not contained within the 95% confidence intervals was recorded for each combination of parameters.

5.4.3 Results

The effect of changing cv and Δcv on the confidence of the bootstrapping method is shown in [Figure 5.3](#). Each plotted point represents the true confidence of the bootstrapping technique with a cv or Δcv of 1 compared to that with a cv or Δcv of 1.41 for the same combination of other parameters. The median (95% CI) proportion of 95% confidence intervals containing the true parameter with $cv = 1$ was 0.924 (0.731 - 0.960), compared to 0.920 (0.713 - 0.960) with $cv = 1.41$. For $\Delta cv = 1$ the median (95% CI) was 0.924 (0.727 - 0.964), compared to 0.916 (0.713 - 0.956) for $\Delta cv = 1.41$. Both cv and Δcv have a significant effect on the proportion of 95% confidence intervals containing the true parameter ($p < 0.001$), but this effect is relatively small. Each of the four datasets with different cv and Δcv at each combination of mean and cv were therefore combined to give 99 datasets of 1000 before analysis of the effect of mean and sample size.

The effect of increasing pre-treatment mean FEC and sample size on the ability of the bootstrapping method to accurately predict the true FEC reduction is shown in [Figure 5.4](#). With

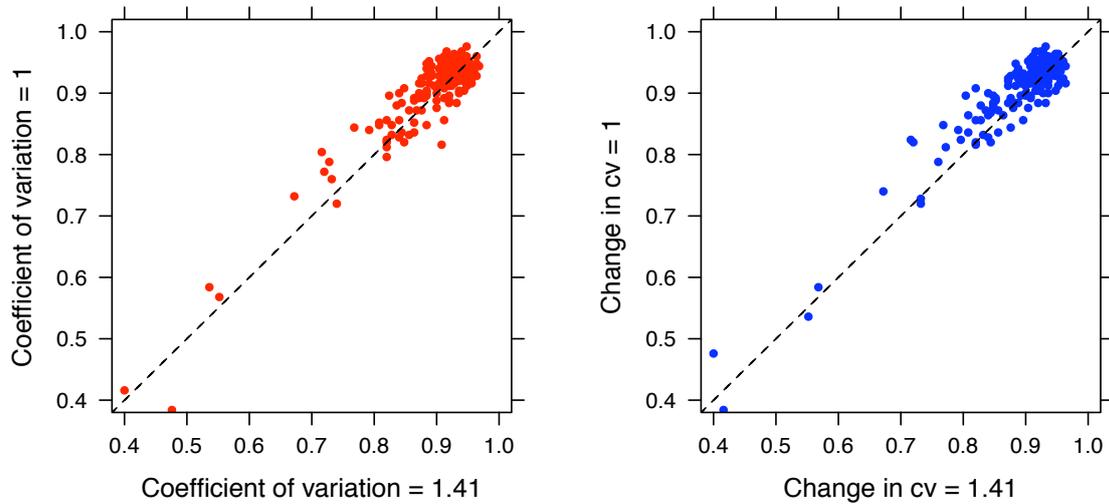


Figure 5.3: Effect of cv (left) and Δcv (right) on the proportion of bootstrapped 95% confidence intervals that contained the true efficacy parameter from 250 simulated FECRT datasets - each point represents the true confidence of the bootstrapping technique with a cv or Δcv of 1 compared to that with a cv or Δcv of 1.41 for the same combination of other parameters

a sample size of five, the notional 95% confidence intervals contained the true parameter as little as 40% of the time with a very low mean FEC. This improved to between approximately 93% and 95% for sample sizes of 50 and above with pre-treatment mean FEC of at least ten counted eggs.

Conversely, the confidence of the estimates produced by the MCMC method using a subset of the same data did not appear to be decreased by a reduced mean and sample size; notional 95% confidence intervals contained the true value 99% and 97% of the time with a sample size of five and means of 1 and 100 respectively, and 97% and 96% of the time with a sample size of 100 and means of 1 and 100.

5.4.4 Discussion

The mean, sample size, variability and change in variability all effect the utility of the notional 95% CI for mean FEC reduction produced by the bootstrapping method. The confidence was lower with increased pre- and post- treatment cv , reflecting the increased probability of extreme values with increasing variability, although the magnitude of this effect was relatively low compared to that of low mean and sample size. Confidence was very low when an average of one pre-treatment egg was counted, but rapidly increased with a mean of five and above. At first glance this provides partial validation to the practise of removing animals with low pre-treatment FEC, although the negative effect of the associated reduction in sample size is

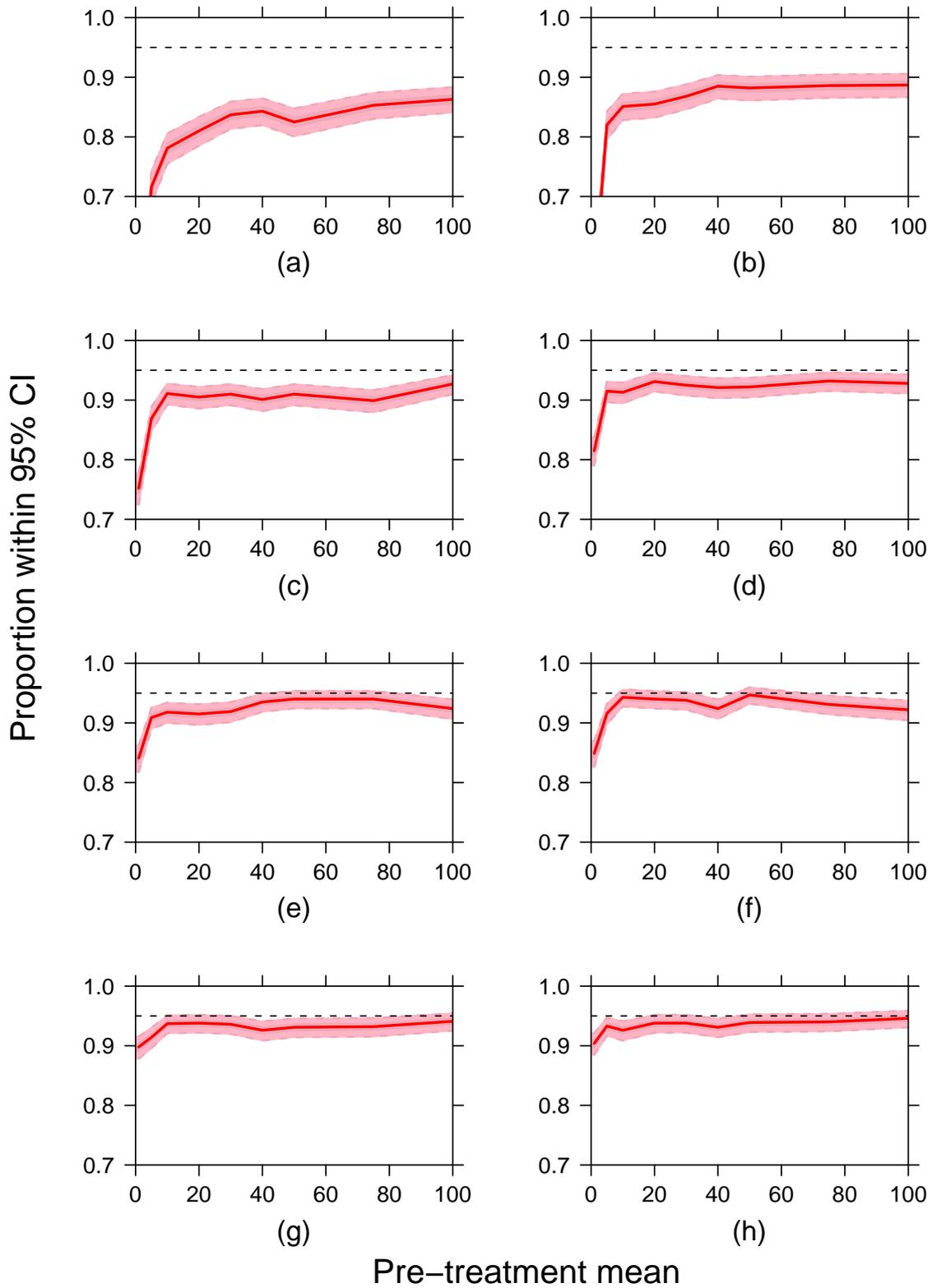


Figure 5.4: Proportion of 95% confidence intervals produced using the bootstrap method that contained the true parameter from 1000 simulated FECRT datasets at each pre-treatment mean number of eggs counted (95% credible intervals shown shaded) - sample sizes 5 (a), 10 (b), 20 (c), 30 (d), 40 (e), 50 (f), 80 (g) and 100 (h) shown. The dotted line indicates the optimum proportion.

likely to more than outweigh the potential benefits of increasing the mean. Confidence was generally very low with smaller sample sizes, even with a much higher pre-treatment mean.

The true distribution of data is unknown, therefore the most conservative estimate would be to use the data from [Section 5.3](#) at which the MCMC method performed worst. This suggests that a computationally intensive parametric method such as MCMC can and should be used in preference to the bootstrap method with a sample size of less than 40 with a pre-treatment mean FEC of 40 counted eggs, or with a sample size of less than 50 with smaller pre-treatment mean FEC. With larger datasets, the data distribution independence and reduced computational effort associated with the non-parametric bootstrap procedure make this method more attractive. These findings are specific to the type of data examined, and may not be directly applicable to other problems. [Cummings et al. \(2003\)](#) found bootstrapping to outperform MCMC based estimates when applied to a specific problem in the field of phylogenetics, although [Alfaro et al. \(2003\)](#) found that MCMC could outperform non-parametric bootstrapping with smaller sample sizes for similar applications. MCMC has also been shown to outperform bootstrapping for quantifying variance parameters ([Ros et al., 2002](#)). Given the need to make assumptions with any type of statistical technique, it is sensible to make use of simulation studies to validate these assumptions for complex problems before relying on inference made from real data.

5.5 Field data analysis

5.5.1 Introduction

There are several variations on the method of carrying out a FECRT. For ruminants, where larger sample sizes predominate and seasonal variation in FEC with animals is typically observed, the majority of the methods involve a control group along with the treatment group so that the change in FEC attributable to the drug can be separated from any change in FEC due to other factors also affecting the two control groups. With the smaller sample sizes encountered with equine FECRT data this would not be advisable due to the unavoidable loss of power in the treatment group, so the tendency is to have all animals in the treatment group. A further variation often encountered, and used with the 64 groups of horses discussed in [Sections 4.3](#) and [5.3](#), is the use of targeted dosing of higher FEC animals with all animals demonstrating FEC under a threshold being removed from the test ([Craven et al., 1998](#); [Coles et al., 2006](#)). This is done partly to reduce anthelmintic use in the hope of delaying the onset of resistance, and also partly because of a perceived difficulty in analysing data from low pre-treatment FEC animals. Unfortunately, due to the effects of within animal variability discussed previously, it does not necessarily follow that animals removed as a result of a single low FEC were really the low shedding animals, and some of the higher FEC observed

may have been due to an unusually high result in a low shedding animal. The result is to present a further difficulty in analysing the data as a result of the bias towards high FEC for pre-treatment data, possibly tending the pre-treatment mean towards a higher value than the ‘true’ group mean, and therefore the reduction towards higher efficacy than would otherwise be expected due to regression towards the mean.

While it has not previously been done, it is possible to use all the pre-treatment data collected along with the model presented in [Section 5.3](#) to provide a more accurate, or less biased, estimate of the pre-treatment group mean. Unfortunately, this presents a further difficulty in analysis of the data in that the post-treatment data, which is only available from high pre-treatment shedding animals, will be biased towards animals with a higher true mean FEC. Modelling the missing post-treatment data as randomly missing is therefore not necessarily accurate. An alternative is to use a more complicated model where the within animal and between animal variability are modelled separately, and the true FEC reduction is derived from the reduction in true animal mean for each animal. The pre-treatment animal means are modelled as part of a group distribution; the post-treatment animal means are derived from this and therefore follow the same shape but with a change in mean due to varying scale parameter. There is an additional variability source compared with the previous model so that pre-treatment variability can arise as a result of between animal variation or pre-treatment within animal variation, which is fixed between animals. Post-treatment variability can arise as a result of between animal variability, which is not modelled implicitly but equal to that pre-treatment, or post-treatment within animal variability, which is scaled relative to the pre-treatment within animal variability in the same way as the total variability used in the previous model. This model carries the further advantages that the paired effect of animal present in the data is modelled explicitly, possibly resulting in a greater power, and a method of increasing power by repeat sampling from individuals is introduced. The disadvantage is that two separate layers of variability are modelled, which may impact on convergence and model stability. This will be especially apparent when only one observation is available per individual, although the model should be able to partition the variability to an extent by using the paired pre and post treatment data.

In this section, analysis of the 64 datasets discussed in [Sections 4.3](#) and [5.3](#) are compared between the two model types using both the selected data and the full data for each group of horses. Differences in inference and the impact of model assumptions will be discussed.

5.5.2 Materials and methods

The FECRT data obtained from 64 herds of Danish horses and previously presented in [Section 4.3](#) was used to compare the inference made using several different FECRT models. All datasets had more animals sampled pre-treatment than post-treatment, on the basis of selec-

tion on pre-treatment FEC being greater than a threshold. The difficulty of analysing data with all-zero counts has previously been discussed in [Section 2.7](#). Three of the datasets demonstrated 100% empirical reductions, and so presented difficulties in estimating the change in variability for both models. These results were included in the analysis, but removed when making inference on the change in variability.

All datasets were first analysed using the model outlined in [Section 5.3](#), removing the data from animals that demonstrated a low FEC (denoted ‘partial data’). The data were then analysed using a paired pre- and post- treatment data model, with both the partial data and the data including the previously removed pre-treatment data (denoted ‘full data’). In this model, pre-treatment data are assumed to follow a compound gamma-gamma-Poisson distribution, with the first gamma distribution describing between individual variability, and the second gamma distribution describing pre-treatment within individual variability. The post-treatment data are assumed to follow a gamma-Poisson distribution with a mean which has been scaled relative to the second pre-treatment gamma distribution for that animal by a reduction which is common to all animals, and a variability equal to the post-treatment within animal variability. The post-treatment within-animal variability is scaled relative to the pre-treatment within-animal variability, and both are common for all animals, a simplifying assumption, and given the same non-informative prior as used previously and in [Chapters 2 to 4](#). The post-treatment between animal variability is not directly modelled, but is equal to the pre-treatment between animal variability because efficacy is fixed between animals. All gamma values were truncated at a lower bound of 10^{-200} as described in [Section 3.6](#). The parameter representing the change in dispersion was bounded between 0.001 and 1000 to prevent failure of the sampling algorithm associated with selection of extreme values. The model descriptions are given in [Models 5.1 and 5.2](#). There was significant evidence for zero-inflation in some of the full datasets used based on the work presented in [Chapter 4](#), so zero-inflated distributions were used for all models so that results can be more easily compared.

5.5.3 Results

The median-relative size of the 95% CI and median estimates of efficacy for the two models using the partial data is shown in [Figure 5.5](#). The individual distribution model produced lower median estimates for efficacy ($p < 0.001$), and a borderline significant increase in 95% CI size ($p = 0.07$) compared to the paired model. The difference in median estimate between the two models was correlated to the amount of information in the data as defined by the relative size of the 95% CI produced by the paired model (Spearman’s rank correlation $p < 0.001$; [Figure 5.6](#)). The 95% CI for efficacy overlapped between the models in every case. The median of the median estimates for change in shape parameter for the single distribution model was 0.22, and the median of the median-relative 95% sizes was 16.4. For the paired

model, these values were 0.07 for the median and 66.7 for the median-relative 95% CI size. These values are not directly comparable between the models, although a value of below one indicates an increase in variability post-treatment for both. The lower 95% CI for change in shape parameter was below one in every case for both models, and the median was below one for 50 (82% of datasets with post-treatment counts) of the paired model results and for 45 (74% of datasets with post-treatment counts) of the single distribution model results, of which 44 were the same datasets. The upper 95% confidence interval was below one in 26 (43%) and 27 (44%) of the datasets for the individual and paired models respectively, of which 25 were the same datasets. All 64 datasets eventually converged with both models, although the paired model was more heavily autocorrelated. Using the required number of samples calculated by the Raftery and Lewis diagnostic as a proxy for the most autocorrelated parameter, the paired model required significantly more ($p < 0.001$); with a median (95% CI) of 79,760 (35,636-187,181) samples compared to 41,172 (22,197-163,798) samples for the individual distribution model.

The median estimates of efficacy and median-relative size of the 95% CI estimates of efficacy are compared using the paired model with the full and truncated datasets [Figure 5.7](#). The full data produced lower median estimates for efficacy ($p < 0.001$), but there was no significant effect on the 95% CI size ($p = 0.92$). Again, the 95% CI for efficacy overlapped between the models in every case. The difference in median estimate relative to the median estimate for the full data is shown plotted against the median estimate for the full data in [Figure 5.8](#). There was no significant correlation (Spearman's rank correlation $p = 0.55$). The median

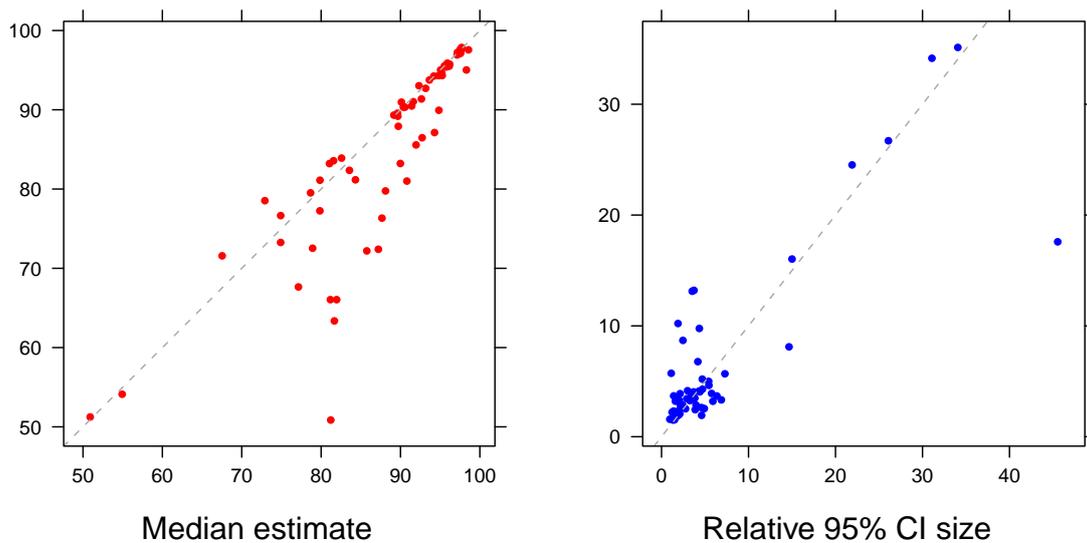


Figure 5.5: Median estimates of mean FEC reduction and median-relative 95% CI size produced for 64 Danish FECRT datasets using the single distribution (y axis) and paired (x axis) models

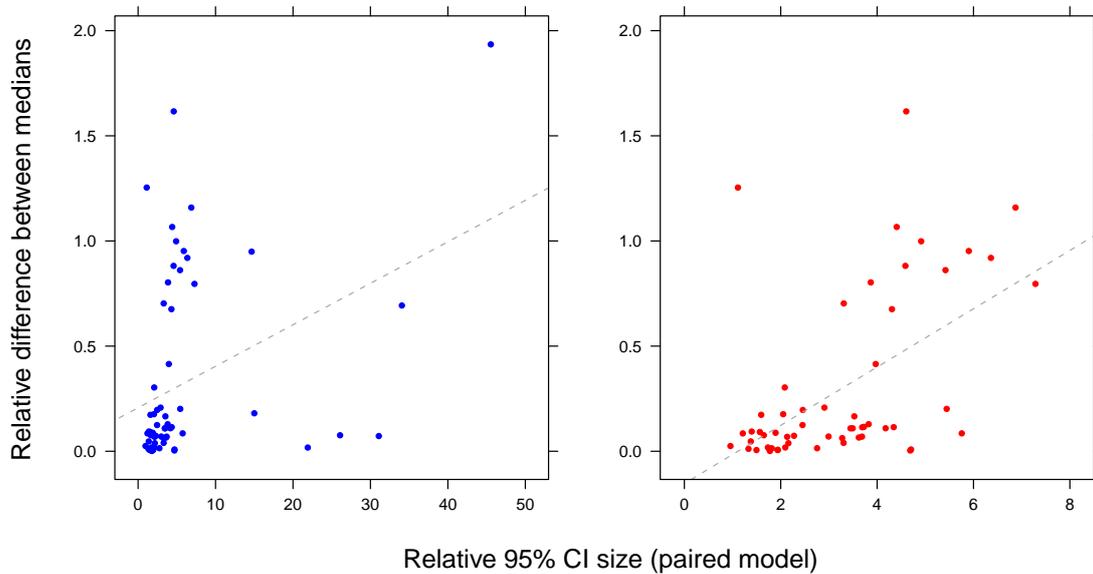


Figure 5.6: Effect of the amount of information in the data (relative size of 95% CI produced by the paired model) on the relative difference between the paired and single distribution median estimates of mean FEC reduction produced from 64 Danish FECRT datasets - all datasets shown on left, 7 points with very large 95% CI removed on right. Regression line shown dashed.

(95% CI) of the median estimates for change in shape parameter for the full data was 0.31 (0.0007 - 11.22), and the median (95% CI) of the median-relative 95% sizes was 69.9 (4.2 - 1910.1). Compared to the same figures quoted for the truncated data above, the full data produces a larger median estimate of change in shape with similar sized 95% CI (Figure 5.9). The lower 95% CI was again below one in every case, with reduced numbers of 39 (64% of datasets with post-treatment counts) and 15 (26% of datasets with post-treatment counts) of the median and lower 95% CI estimates below one respectively.

For the full data, the estimates of zero-inflation returned were almost identical to those produced using the simple ZIGP model Figure 5.10. There was evidence of more than 1% zero-inflation in ten (16%) of the datasets, with four datasets returning lower 95% CI for zero-inflation of between 10% and 20%, and one dataset returning a lower 95% CI of 26%.

5.5.4 Discussion

Using the same truncated datasets, the individual distribution model inferred slightly lower estimates for efficacy, i.e. tended towards lower efficacy, compared to the paired model. This difference between the median estimates appeared to be greater when the paired model returned larger relative 95% CI sizes for efficacy, suggesting that this effect is exacerbated by

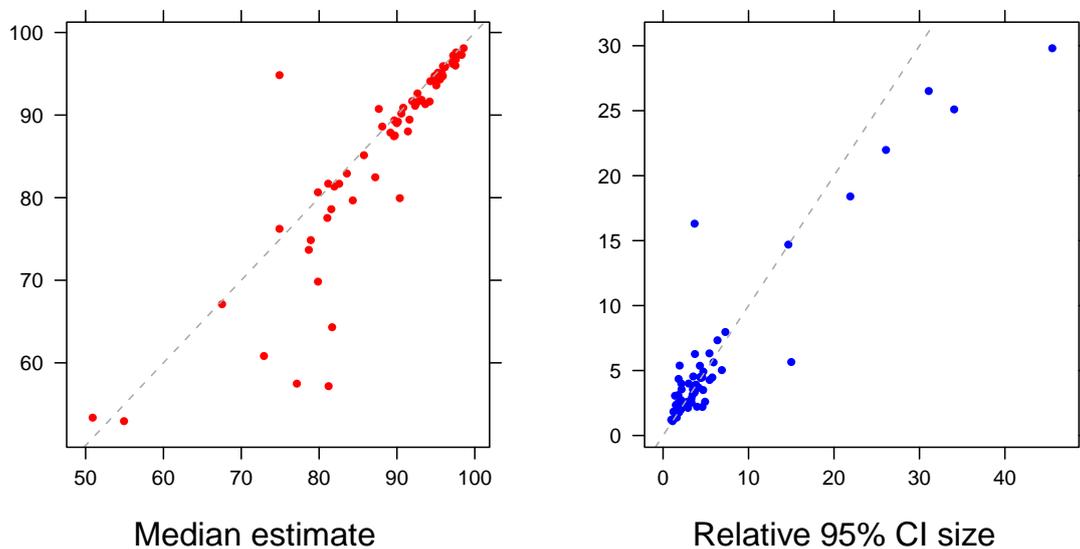


Figure 5.7: Median estimates of mean FEC reduction and median-relative 95% CI size produced for 64 Danish FECRT datasets using the full dataset (y axis) and truncated dataset (x axis) with the paired model - the dotted line represents equality

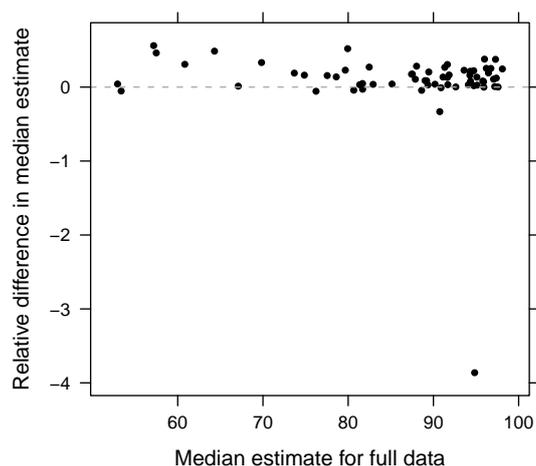


Figure 5.8: The difference in median estimate of mean FEC reduction between the full and part data relative to the median estimate for the full data (y axis), plotted against the median estimate for the full data (x axis), produced for 64 Danish FECRT datasets - the dotted line represents equality between the estimates produced from the full and part data

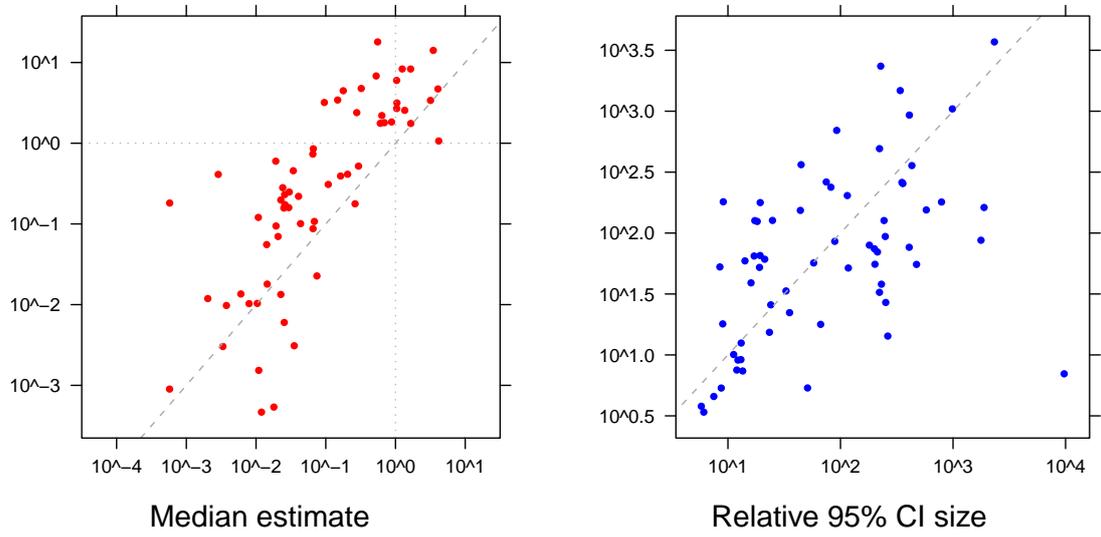


Figure 5.9: Median estimates of change in shape and median-relative 95% CI size produced for 64 Danish FECRT datasets using the full dataset (y axis) and truncated dataset (x axis) with the paired model

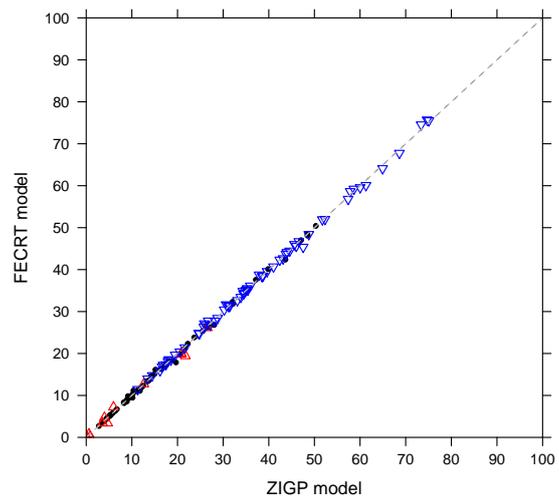


Figure 5.10: Median (black) and 95% CI (red-blue) estimates for zero-inflation produced by the FECRT model and the simple ZIGP model with the full pre-treatment data from 64 Danish FECRT datasets

a lack of information in the data. This implies that the individual distribution model returns results that are more heavily influenced by the prior, which given the bias towards higher efficacy in the datasets relative to the $\text{Beta}(1, 1)$ prior would result in a tendency towards lower values of efficacy. This is consistent with the smaller 95% CI returned by the paired model, suggesting that the paired model is able to extract more information from the data. However, the difference in medians may also be due to a potentially greater sensitivity to the effect of pre-treatment sampling bias as a result of the paired pre- and post-treatment data. The median-relative 95% CI sizes for efficacy appear to be similar between the models, indicating that little extra information is gained from the paired effect when only one sample is available per animal. Both models indicated that variability is likely to increase after treatment, with the paired model inferring a greater increase. The change in variability is not directly comparable between the models, as for the single distribution model the variability is applied to the full distribution of combined within and between animal variability, and therefore encompasses all the changes in variability that may be expected as a result of the change in within and between animal variability as a result of treatment. Within animal variability may change as a result of, for example, increased variability in the fecundity of worms in response to treatment with anthelmintic. Between animal variability may increase due to differing efficacy of drug between animals. For the paired model, this change in variability is applied only to the within animal variability, so in theory a change in variability between animals after treatment must be directly modelled as an effect of animal on efficacy. However, where only one post-treatment sample is available it is impossible to separate an increase in within animal variability from an effect of animal on efficacy, so that the model may have been compensating for a non-uniform efficacy by increasing the change in variability. This may explain why the paired model produced smaller estimates for the change in shape; in order to effect the same post-treatment variability a larger adjustment would have to be made to the within animal variability alone than the total variability. Ideally these effects would be separated using a distribution of efficacy, or a random effect of animal on efficacy, which would allow separate inference on the change in within animal variability and the change in between animal variability. This would require sufficient replicated samples from each animal to make the model identifiable, which is not currently available but would be interesting for a future study. The change in between animal variability would be modelled indirectly as a result of the distribution of efficacy, but could be calculated or at least approximated using the returned chains. This would allow direct comparison of the change in within and between animal variability as a result of anthelmintic treatment.

Three of the datasets had all zero count observations at post-treatment, and produced large 95% CI for efficacy and high estimates of post-treatment variability using the individual model for the reasons explained in [Section 5.3.4](#). Although the variability sources are formulated slightly differently in the paired model, a similar effect on estimates for efficacy and change in within animal variability is seen with only zero counts post-treatment. The weighting towards higher values of post-treatment mean driven by the prior on efficacy is compensated

for by inferring a large post-treatment variability and large change in variability from pre-treatment variability. As discussed previously, this inference is a result of the prior used for efficacy and the lack of information in the data; as there are no post-treatment counts from which to infer the variability it would be inappropriate to use the model to try to estimate this. Therefore, estimates of cv produced by these models are ignored, although inference on efficacy is still appropriate because this is data driven. More useful analysis of these data could be achieved using more informative priors for efficacy and/or the change in variability, if such information were available.

When applied to the full data, the paired model produced lower estimates for efficacy than when applied to the truncated data. This supports the supposition that selecting pre-treatment data on the basis of high FEC biases the resultant efficacy estimate towards higher values, forcing the drug efficacy to appear greater than it truly is. [Figure 5.7](#) shows a cluster of four points with low efficacy where this effect is much stronger, suggesting that a decrease in efficacy may exaggerate the effect of selection, although [Figure 5.8](#) shows that the relative difference is not correlated to efficacy. The same outlier point can be seen on both graphs, and is likely due to a lack of information in the data making a large difference to the median. The 95% CI estimates overlapped between the models in every case. This finding is a result of the data collection not the analysis technique, and so applies equally well to the use of the WAAVP or bootstrapping methodologies previously discussed. Use of the partial dataset compared to the full dataset also resulted in a difference in inference on the change in variability. This is also likely to be due to the effect of the selection of the pre-treatment data; variability is reduced in the pre-treatment data so a greater increase in variability is required to affect the same value of post-treatment variability. However, the models were relatively consistent with the direction of change in variability - even with the full data the possibility of an increase in variability was retained with all models, and there was sufficient information to confirm this in 28% of the datasets. It is not clear whether this reflects an increase in within animal variability, between animal variability or both, although there are reasonable biological arguments for an increase in both sources of variability.

5.6 Meta-population modelling

5.6.1 Introduction

The 64 datasets of equine FECRT data were all obtained from the Equilab Laboratory, and represented voluntary submission of FECRT data from a wide variety of equine establishments across Denmark. All reductions were obtained using the same anthelmintic Pyrantel embonate, and farms are likely to be similar in many respects such as the same species, similar age demographic etc, although the data were not taken at the same time of year on

each farm. It is therefore possible to analyse all 64 datasets together using a single meta-population MCMC model. The paired type model is most appropriate when treatment is on the basis of pre-treatment FEC, so only that model will be used. The results obtained will be compared to the results from the analysis of the herds separately and these results discussed.

5.6.2 Materials and methods

The paired data model discussed previously is adapted so that the mean efficacy and change in within animal variability for each dataset are random effects derived from a gamma distribution describing all datasets. The pre-treatment mean, variability and zero-inflation are fixed effects and independent for each farm. The mean efficacy is given a Beta(1, 1) prior to restrict this value to less than one, and the mean change in shape parameter a diffuse lognormal parameter with a mean of one and log precision of 0.01. The inverse of the shape parameter of the gamma distributions describing efficacy and change in shape are given log uniform priors between 0.001 and 1000. Due to the slower updating rate because of the large amount of data, the maximum convergence time was extended to 12 hours. The model specification is given in [Model 5.3](#).

5.6.3 Results

[Figure 5.11](#) shows the agreement between median and lower/upper 95% CI estimates produced by the metapopulation and individual paired models for efficacy of each of the 64 datasets. The regression line shown for each is less than one, indicating that the metapopulation model produced less variation between datasets for all estimates. The lower 95% CI and median estimates for datasets with no post-treatment observations (shown in triangles) were particularly altered by use of the metapopulation model. The median-relative size of the 95% CI for efficacy produced by the metapopulation model were smaller than those produced by the individual paired models for 60 (94%) of the datasets. The 95% CI for efficacy overlapped between models for each dataset.

The posterior probability densities for the mean and cv of the distributions describing the efficacy and change in shape parameters are shown in [Figure 5.12](#). The median (95% CI) of the mean efficacy was 0.065 (0.053 - 0.080), corresponding to a median (95% CI) estimate of 93.46% (92.02 - 94.74%) for the metapopulation mean FEC reduction. The median (95% CI) estimate for the mean change in shape parameter was 0.35 (0.20 - 0.57) indicating that the average total variability after treatment is greater than that before treatment. There was sufficient information in the data to estimate the cv of the efficacy distribution (the variability in efficacy between farms) and the change in shape distribution (the variability in change in the change in variability after treatment between farms) with only a relatively wide 95% CI,

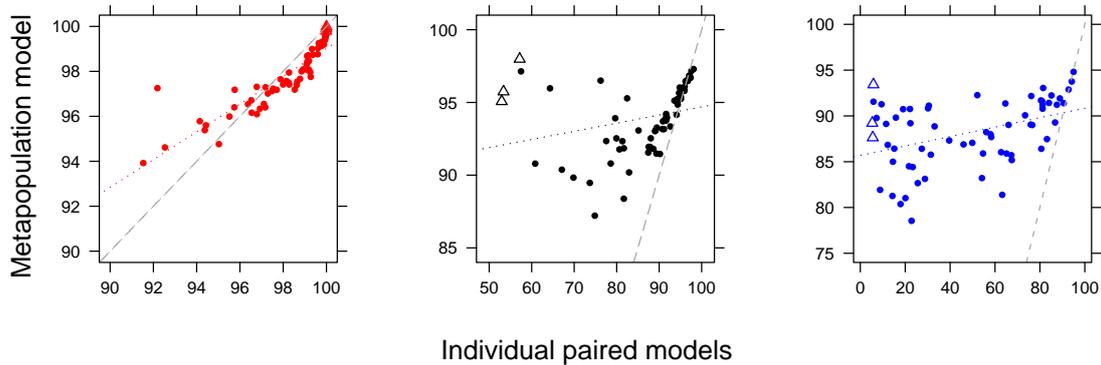


Figure 5.11: Median (black) and 95% CI (red-blue) estimates for efficacy produced by the metapopulation model (y axis) and the individual paired models (x axis) from analysis of 64 Danish FECRT datasets - regression lines shown dotted, agreement line (slope 1, intercept 0) shown dashed. Datasets with no post-treatment observations in triangles.

although it appears that the change in variability is more consistent between farms than the efficacy.

5.6.4 Discussion

As would be expected, fitting farm efficacy as a random effect derived from a single global distribution dramatically reduced 95% CI for estimates of efficacy on individual farms. The metapopulation model produced higher estimates for 62 (96.9%) of the lower 95% CI, and lower estimates for 51 (79.7%) of the upper 95% CI, which is a consequence of the random effects model. There also appeared to be a tendency for the metapopulation model to shift the posterior density for each farm efficacy towards the metapopulation mean value compared to the individual models, evidenced by the slope of the regression lines shown in [Figure 5.11](#). This is a result of the assumption that farm efficacies are derived from a common distribution, making outliers appear to be as a result of extreme random variation rather than an extreme true efficacy. The validity of this assumption depends on the validity of the distribution used to describe these efficacies which is difficult to assess; although this effect decreases as information about farm specific values increases. The variability in efficacy between farms could be the result of the variability in farm management practices, age of the animals, time of year when the sampling was performed as well as differences in the resident population of parasites. However, this is not necessarily a continuous distribution and may not be well approximated by the gamma distribution.

The distributional assumptions of efficacy and the change in shape parameter also allow data where all post-treatment counts are zero to be analysed. The change in shape pa-

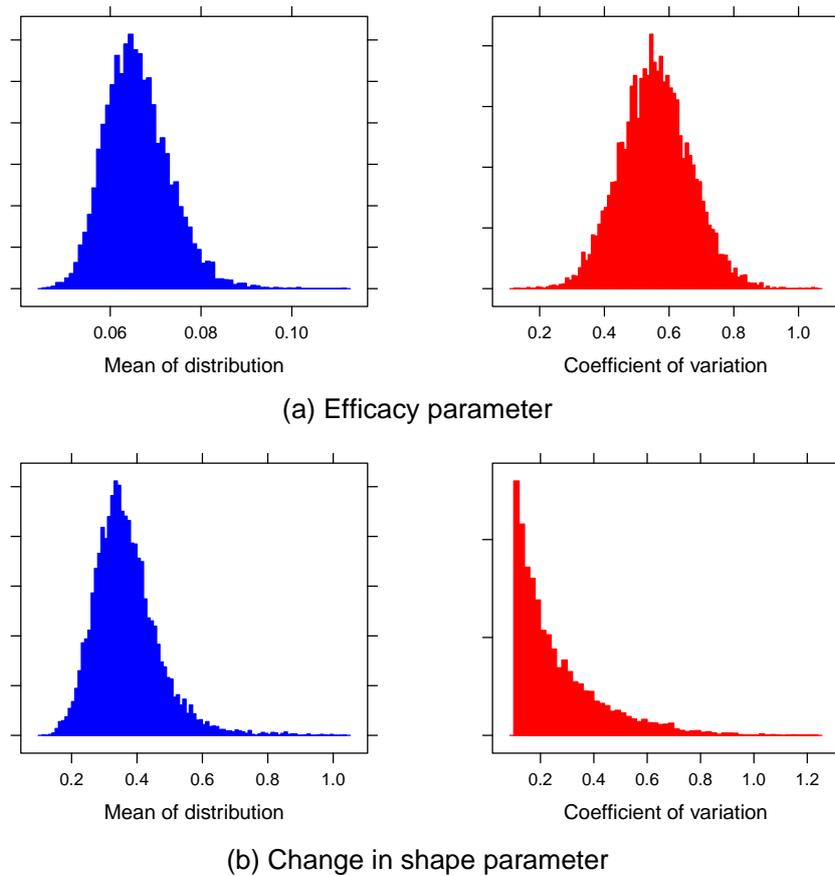


Figure 5.12: Posterior probability densities for the mean (blue) and cv (red) of the efficacy (top) and change in shape parameter (bottom) from metapopulation modelling of 64 Danish FECRT datasets

parameter for the farms with no post-treatment observations is sampled from the distribution informed by datasets from which information about this parameter is available. Effectively, the post-treatment shape parameter is informed from the other datasets rather than from the prior. The efficacy parameter for these datasets is also partially informed from the efficacy parameter in other datasets, rather than the Beta(1, 1) prior used in the individual model. This additional information allows the confidence interval size for efficacy to be reduced to a greater degree than for other datasets (Figure 5.12).

The 95% confidence interval for the mean efficacy was 92.0 - 94.7%, indicating that the mean efficacy of the metapopulation is likely to be below 95%; the published efficacy value for Pyrantel. However, the interpretation of this value is not clear because no similar efficacy studies exist with which to compare it. Conceptually, the variability in efficacy between farms could arise due to several factors, including systemic biases in management factors and the age and breed of animals, along with differences in susceptibility of worm populations between farms. The posterior density for the cv of the efficacy distribution indicates that

efficacy is quite variable between farms, implying that these factors have an appreciable effect on the observed variability in efficacy between reduction tests performed on different farms. Animal age was recorded in the data so could have been incorporated into the model, but there is no way to separate the effect of management practises and worm susceptibility on farm efficacy. It would be reasonable to assume that a large proportion of this effect is due to differences in worm susceptibility, but a detailed study to confirm this would be worthwhile. One difference between the meta-population and individual based models is the restriction on farm efficacy. In the individual model, farm efficacy is restricted to values of greater than zero, but in the meta-population model only the mean efficacy has this restriction and individual farm FEC reductions can be negative. This could have been addressed using a bounded gamma distribution, although this would not have had any effect because none of the estimated farm reductions were below 75% (Figure 5.11).

The 95% confidence interval for the mean change in shape parameter did not include one, indicating that there is a tendency for the coefficient of variation in FEC to increase after treatment with anthelmintic. This could be due to randomly varying efficacy of drug between animals, varying dosing between animals, or an increase in within animal variability after treatment due to increased variability in the fecundity of worms after treatment. This increase in variability is observed alongside a decrease in mean, so it is also possible that there is a biological explanation for an increase in cv with lower FEC. For example, if worms release eggs in boluses, a larger number of worms would decrease the variability in total fecundity because the bolus release of eggs is averaged out over more worms. Interestingly, the posterior density for the estimate of the cv of the distribution of change in shape indicated a relatively low cv , suggesting that the magnitude of the change in cv after treatment is more consistent between farms than the magnitude of the change in efficacy. This implies that the variability in efficacy between animals on a farm may be quite consistent between farms. Further studies to validate this would certainly be of benefit.

All the inference presented here should be interpreted with a degree of caution as the assumptions that farm efficacy and the change in variability can be described by a gamma distribution may not be valid. This would not be true if the true distributions were multi-modal, which might reasonably be expected if anthelmintic resistance was dichotomous rather than graded, i.e. present to a degree on all farms. Adaptation of the model to allow separate distributions of these parameters for farms where resistance is and is not present may provide different inference, although this would require a definition of ‘resistance’ that is currently lacking and auto-correlation of the mixed model would likely be very high. Further studies to determine the validity and impact of the meta-population assumptions is required before such inference can be relied upon.

5.7 Discussion

Equine FECRT data are frequently characterised by small sample sizes, low means and a high degree of variability (Uhlinger, 1993). Based on the results presented here, it would appear that such data are most appropriately analysed using MCMC, although the bootstrapping procedure is to be preferred with sample sizes of 50 or more due to the distribution independence and reduced computational complexity involved in the analysis. Analysis of data with treatment and untreated control groups is also possible using MCMC, but has not been addressed in this chapter. The main reason for this is that equine FECRT datasets tend to be small, so halving the sample size on which to base a reduction is quite detrimental (Coles et al., 1992; Pook et al., 2002) and not easily justified. Where sufficient data are available to do this, the additional inference on change in ‘baseline’ FEC may be valuable if there is reason to suspect that it may be changing, although work in sheep suggests that little is to be gained in terms of diagnostic test sensitivity and specificity from such a procedure (McKenna, 2006b).

The practise of selection of animals for a FECRT on the basis of pre-treatment FEC is currently routine (Craven et al., 1998; Coles et al., 2006). This may increase the utility of the WAAMP and bootstrapping methods by ensuring that the pre-treatment mean is high enough to mitigate some of the disadvantages associated with these techniques, but has the serious disadvantage that the effective sample size is reduced, and is not necessary for the MCMC method because the error structure of the data is inherently accounted for. Furthermore, based on the results presented here, it is evident that selection of animals for a FECRT on the basis of a single high FEC will bias the result towards a higher than true mean efficacy using any analysis method unless this effect is taken into account. Use of the paired data MCMC model is able to do this, and has the added benefit of allowing several replicate samples to be taken from each animal to add extra information about the pre-treatment mean FEC. Use of replicate post-treatment samples would require a modification to the model adding an effect of animal on efficacy so that the change in within animal variability can be modelled independently to the variability in efficacy between animals, and possibly incorporating effects of age or gender etc. This modification would likely reduce identifiability, but this should be improved as more and more samples per animal are modelled. Given the relatively poor power usually encountered with equine FECRT data especially, this alternative may provide a useful way of increasing the utility of the method albeit at the expense of laboratory time. The additional inference on change in variability at different levels, as well as the potential to robustly examine the effect of age and gender on efficacy, would provide an interesting avenue for future work with large datasets. The inclusion of a separate parameter describing the distribution of efficacy to the paired model would also allow the true efficacy to be analysed appropriately at an individual animal level. This is not possible using the individual model, where the pre and post treatment means represent single sample means not animal means,

and with the paired model without a distribution of efficacy, where the post-treatment mean does not reflect variability in efficacy between animals. This further inference could provide the considerable advantage of allowing quarantine or removal from pasture of individuals with the lowest true efficacy, preserving anthelmintic efficacy on the premises for as long as possible.

However, the paired model does carry several drawbacks. The most obvious is the increased auto-correlation caused by segregating the within and between animal variability, which is only a computational issue as long as identifiability is sufficient to allow convergence given a long enough run time. There are also several assumptions inherent to this model specification that are not made with the simpler model. In order to reduce the number of degrees of freedom, the within animal variability is common to all animals in the model. This is definitely a simplifying assumption, the impact of which is not clear and the validity of which can not be assessed due to the absence of information in the data or the literature. Additionally, the change in variability both within and between animals is modelled as being independent of the mean. It is possible that the change in variability may be higher when the mean of an animal is lower, for example as a result of time-clustered fecundity of individual worms having more of an effect with a smaller number of worms. The simpler model is robust to all of these potential effects because individual counts are assumed to be randomly distributed over a distribution incorporating all potential sources of variability combined. This model is therefore more conservative in terms of model assumptions, but cannot be used when animals are selected for treatment on the basis of FEC, which may be legitimately required in order to delay emergence of anthelmintic resistance, and is unable to benefit from the extra data when multiple FEC per individual are available.

It is also apparent from the wide confidence intervals for efficacy presented in [Section 5.5](#) that analysis of a single equine FECRT dataset will often prove inconclusive. This is consistent with the conclusions made by [Miller et al. \(2006\)](#) that the results of a FECRT based on an arithmetic mean reduction can be inconsistent. The analysis of data from multiple sites presented in [Section 5.6](#) could yield more conclusive results, but carries several un-testable assumptions of meta-parameter distributions and may also be less sensitive to a reduced efficacy at a single site. Use of a bootstrapping procedure to analyse data from multiple sites as advocated by [Vidyashankar et al. \(2007\)](#) also carries similar drawbacks. Alternatively, the utility of the method could be increased by performing a suitable sample size calculation prior to performing the FECRT, and increasing the number of samples taken and/or reducing the egg detection threshold accordingly. This may not be practical for routine clinical tests due to the added cost and time associated with taking more samples and counting more eggs. A more useful solution might be to use a longitudinal approach to identify any result that is considered ‘exceptional’, indicating a potential cause for concern to be investigated (for more details, see [Reinke \(1991\)](#)). A more detailed FECRT with multiple samples taken from each animal, and a model similar to the paired model presented in [Section 5.6](#), could then be

used when the data indicated a potential or arising problem. This is similar to the statistical process control approach that has been used extensively in the manufacturing sector, and more recently as an aid to improve medical care (Duncan et al., 2005), monitor adherence to clinical guidelines (Peek et al., 2008), and as a tool to improve epidemiological disease surveillance (Williamson and Weatherby Hudson, 1999; Morton et al., 2001; Arantes et al., 2003). The application to anthelmintic resistance surveillance may represent both a more efficient use of resources, and a greater overall diagnostic test sensitivity and specificity, than the current use of repeated reduction tests viewed in isolation and without the necessary sample size calculations.

In this chapter, inference has been made only on the true FEC reduction, somewhat incorrectly abbreviated to ‘efficacy’, with no attempt made to define ‘resistance’. This is mainly because defining a dataset as susceptible or resistant requires a figure below which the true FEC reduction is incompatible with susceptibility to the drug, which is often unknown (Bauer et al., 1986; Martin et al., 1989; Kaplan, 2002; Vidyashankar and Kaplan, 2005). There are currently a lack of general guidelines on an appropriate figure to be chosen for the FEC reduction. Studies of the FEC reduction after treatment with the same formulation of Pyrantel as used to generate these data have found reductions between 93% and 99% (Boersema et al., 1995; von Samson-Himmelstjerna et al., 2007). A figure of 95% seems to be tentatively accepted as the best estimate of the efficacy of Pyrantel in a naïve population, although Bauer et al. (1986) used a lower value of 90% to define resistance. This is at least partly attributable to a confusion in the non-statistical literature between the effects of magnitude and significance; the target FEC reduction often appears to be adjusted on the basis of an anticipated lack of confidence in inferences made. Compounding this, the current guidelines based on interpretation of the mean reduction and lower 95% confidence interval while ignoring the upper 95% confidence interval also precludes any distinction to be made between datasets that lack evidence for susceptibility from datasets where resistance is likely. Both the bootstrapping and MCMC methods return a probability distribution for efficacy in place of the mean and 95% CI returned by the WAAMP method, allowing more intuitive inference to be made from the output. The efficacy of reductions could be classified according to the probability that the true reduction was below a given threshold (\hat{p}), allowing the distinction to be drawn between cases where there is clear evidence for resistance and cases where there is insufficient evidence to demonstrate acceptable efficacy. Both the bootstrap and MCMC procedures also allow any figure to be used for the required FEC reduction, depending on the minimal acceptable efficacy of the drug in each clinical situation. This could be the published efficacy of the drug used, or a figure chosen to reflect the point at which other drugs or control methods would provide better parasite control. If standardised in the veterinary parasitology literature, such a classification scheme could improve the interpretation of FECRT results substantially.

Analysis of equine FECRT is compounded with difficulties introduced by high variability

and small numbers of observations (Uhlinger, 1993; Kaplan, 2002; Miller et al., 2006; Coles et al., 2006). Based on the results presented here, such data are best analysed using parametric techniques that describe the known sources of variability within the data in order to make the most informed inference. Meta-population modelling can aid inference on individual farms, although the distributional assumptions made by using such techniques require further investigation before any inference made can be relied upon. Analysis of data with no post-treatment observed eggs is especially difficult, and is likely to yield inconclusive results without prior information on the expected efficacy and/or post-treatment variability. This is not a limitation of the analysis technique *per se*, rather a function of the lack of information in the data. Alternative approaches such as bio-equivalence testing could be investigated for analysis of these data, or inference from previous studies in similar establishments used to generate prior distributions. Most usefully, repeated observations could be made for each animal in order to reduce the likelihood of failing to observe eggs. Finally, the consequences of selecting animals for a FECRT based on a single high pre-treatment FEC need to be addressed during the analysis of such data. If targeted dosing of animals is required on the grounds of anthelmintic resistance then the resultant data should be analysed using a model with appropriate variability structure. If targeted dosing is not necessary, then analysis of FECRT using an appropriate technique is best informed by use of all available animals regardless of pre-treatment FEC.

One of the main conclusions of this chapter has been that the analysis of equine FECRT is frequently associated with a large degree of uncertainty in the true efficacy, as a result of a lack of study power. Despite this, prospective power calculations are rarely, if ever, performed prior to undertaking a FECRT or FEC study. In the next chapter, methods of performing such calculations are developed and parameterised using inference on the variability structure of FEC gained from intensively sampled FEC data.

CHAPTER 6

Variability and repeatability of FEC

Variability and repeatability of FEC

6.1 Introduction

The FEC test is the most widely used method of quantifying the nematode parasite burden of domestic animals, and is essential for both routine surveillance of parasite burdens as well as surveillance for anthelmintic resistance (Coles et al., 2006). Despite this, the statistical methodology behind the FEC is poorly understood, and the usefulness of any information obtained from these data are therefore also poorly understood. Studies have been made to attempt to quantify repeatability in FEC (McKenna, 1981; Hoste et al., 2002; Miller et al., 2006), and even to quantify the variability sources (Mes, 2003) although with limited statistical vigour (Morrison, 2004). It is currently therefore very difficult to perform suitable precision analysis for FEC data and prospective power calculations for FECRT data, frequently leading to under-powered studies with equine FECRT datasets (as discussed in Section 5.5). A quantitative study of these potential variability sources is therefore essential to allow future studies to be undertaken with an adequate probability of obtaining a useful conclusion, and may suggest ways in which the procedure could be altered to improve repeatability.

The variability structure of FEC data is likely to be quite complex, with sources of variability aggregated at different levels of the biological system introducing a nested structure to the data that precludes the use of simple uni-modal continuous distributions to analyse the data. The use of intra-class coefficients (Fisher, 1954), which have been used to separate, for example, variability within a rater to variability between raters (Shrout and Fleiss, 1979) and the structure of variability within and between families (Stanish and Taylor, 1983), present one possible solution. However, it is also possible to model the variability structure directly using similar MCMC methods to those used to analyse FEC data previously, and this approach will be used here. The variability of FEC are first considered at an individual animal level before being expanded to multiple animals, and the parameter estimates used to inform development of a precision analysis method for FEC and FECRT data. Although throughout this chapter an arithmetic mean simulated egg count is used to calculate the precision, the results also apply to more statistically complex methods of calculating the mean egg count such as maximum likelihood analysis or MCMC, in which case the precision

will reflect the relative size of the posterior 95% CI produced.

6.2 Quantifying variability within an individual

6.2.1 Introduction

To improve the usefulness of the standard FEC procedure, it is necessary to first identify the source of the observed variability in FEC. It is widely understood that the FEC measurement varies even when applied to individual animals. The source of this variability has not been well studied but would theoretically be as a result of some combination of the following:

1. The variability in egg shedding rate within an animal over a long period of time, due to changes in worm burden
2. The variability in egg concentration between faecal piles over a short period of time, due to variable fecundity of a relatively fixed population of worms and variability in faecal consistency
3. The variability in egg concentration between areas or boluses of the same faecal pile, due to incomplete mixing of eggs within faeces on a wide scale
4. The variability in egg concentration between samples taken from the same area of a faecal pile, which can be further subdivided into:
 - (a) The degree of aggregation of eggs on a local scale within faeces
 - (b) The amount of faeces taken per sample; for example 3 grams in the standard method
5. The variability introduced by the McMasters technique itself, due to factors such as errors in weighing faeces and the effect of sieving material
6. The variability introduced by the process of counting eggs

Faecal egg counts are most frequently used to obtain a ‘snapshot’ of the parasitism of a group of animals at a single point in time, so that the variability in egg shedding rate within an animal over a long period of time, for example over a grazing season, will not have an effect. This source of variability will therefore be discounted. The variability due to the counting process would be expected to follow a Poisson distribution if eggs are randomly mixed by the process of blending and mixing, but may in reality follow an over-dispersed Poisson distribution as a result of clumping of eggs within a sample even after mixing. Of

the remaining four potential sources of variability, there is little available information in the literature concerning their relative importance. The McMasters technique itself has been used for several decades and has proved to be useful and reliable in sheep, although it has been suggested that the process does not recover all eggs available in the sample [Mike Stear pers. comm.]. If the proportion of eggs recovered is fixed, i.e. the McMasters technique recovers exactly the same proportion of eggs in each sample, then this will not represent a source of variability because the mean of each sample is scaled equally. If the proportion of eggs is not fixed, i.e. the proportion of eggs recovered varies randomly, then this will introduce a source of variability into the resultant FEC.

It is possible to quantify these sources of variability 2-6 by repeated sampling of an individual over a short space of time, with several samples taken from each area of each faecal pile, and the McMasters process replicated several times on each sample so that the variability sources can be separated and quantified. The variability due to the process of counting the eggs can be quantified separately by counting the number of eggs in several McMasters chambers from the same sample and recording the results individually. It is also possible to separate the variability introduced by the McMasters technique from the variability in egg concentration between samples taken from the same area of a faecal pile by varying the amount of faeces taken per sample and examining the relationship between the mean and the cv . This allows the sources of variability that are independent of the mean total egg content of the sample, with a fixed cv , to be separated from the sources of variability that are dependent on the total egg content of the sample, with a variable cv . The latter represents variability in the mean egg count of the homogenised sample of faeces, which would be expected to vary more with a smaller homogenised sample than with a larger homogenised sample. The former represents any source which is independent of the size of the sample, such as variability introduced by inaccuracies in weighing faeces or measuring liquid levels as long as the standard deviation of these is proportional to the mean. These sources of variability will be attributed to the McMasters technique, although it is possible that further mean-dependent sources of variability introduced by the McMasters technique exist. Assuming that both process result in a gamma distribution, the total effective cv of the variability due to the McMasters technique (that which is independent of the size of the sample) and the variability between samples taken from the same area of a faecal pile (that which is dependent on the size of the sample) is given by [Equation 6.2](#) (for the derivation see [Appendix B](#)). It is therefore possible to separate the two sources of variability by examining the effect of varying the amount of faeces used on the cv_{total} . The combined variation of two independently sampled distributions is given by:

$$cv_{total} = \sqrt{cv_1^2 + cv_2^2 + cv_1^2 \times cv_2^2} \quad (6.1)$$

Where cv_{total} is the effective combined coefficient of variation (cv), cv_1 is the cv in the first distribution, and cv_2 is the cv in the second independently sampled distribution. This can be extended to consider the average of multiple samples from one of these distributions by dividing the cv of this distribution by the square root of the number of samples taken, or the relative size of the single homogenised sample taken. When applied to the combination of variability due to the McMasters technique and the variability between replicates, this gives:

$$cv_{total} = \sqrt{cv_1^2 + \left(\frac{cv_2}{\sqrt{g}}\right)^2 + cv_1^2 \times \left(\frac{cv_2}{\sqrt{g}}\right)^2} \quad (6.2)$$

Where cv_{total} is the effective combined coefficient of variation (cv), cv_1 is the cv due to the variability of the McMasters technique, cv_2 is the cv due to the degree of aggregation of eggs on a local scale within faeces, and g is the amount of faeces (in grams) used per sample.

The remaining sources of variability 2-3 can be quantified by taking samples from several faecal piles produced by the same animal, and from different areas or boluses of a single faecal pile. The resultant data could then be analysed using MCMC to provide parameter estimates for the cv introduced at each level of variability within an individual, and consequently the total variability that would be expected from samples taken from an individual over a short period of time. This could then be compared to the total variability reported between similar groups of animals to quantify the true variability between animals.

6.2.2 Model specification and validation

6.2.2.1 Introduction

There are several different ways of specifying an MCMC model to quantify the sources of variability 2-5 given above. The first is a multi-level hierarchical model, using a series of compound distributions to describe the variability in the true mean egg count at each of the sources of variability. At each level of the model, the mean of the distribution is used to populate the distribution at the higher tier. This formulation has the advantage of estimating the mean egg count, but multi-level hierarchical models are frequently associated with poor convergence. A second formulation could involve modelling the variability between egg counts at each level separately, calculating the observed variability at each level using [Equation 6.3](#) (again, see [Appendix B](#) for the derivation of this formula). The mean of the distribution at each level could be modelled as part of the distribution at the tier above, using a different value for effective variability to reflect the removal of variability at that level, preserving the hierarchical part of the model where this information is available.

$$cv.obs_i = \sqrt{cv.obs_{i-1}^2 + cv.actual_i^2 + (cv.obs_{i-1}^2 * cv.actual_i^2)} \quad (6.3)$$

Where $cv.obs_i$ is the effective coefficient of variation (cv) at level i , $cv.obs_{i-1}$ is the effective cv at level $i - 1$, and $cv.actual_i$ is the actual cv due to the variability described at tier i . The variability at tier one is calculated using [Equation 6.2](#).

In order to evaluate which of these two choices of model is most useful, a comparison of the accuracy of the estimates obtained using each model was performed.

6.2.2.2 Materials and Methods

Simulated datasets were generated using a series of compound gamma distributions to describe the variability around the mean value at each level. For each dataset, a mean faecal egg count of 250EPG was used. Twenty faecal piles were simulated from a single simulated animal, with 15 of these piles represented by a single sample and the remaining five represented by twenty samples from different locations. Five of these samples were represented by 30 replicates, ten of each using 1g, 3g and 9g of faeces. The remaining 15 samples were represented by a single replicate. For each replicate, the number of eggs in ten McMasters chambers was simulated using a Poisson distribution. This resulted in a dataset consisting of 8400 egg count observations from 840 replicate samples.

The variability between faecal piles, between different areas of a faecal pile, between different samples from the same area, and the variability introduced by the McMasters process were varied between datasets. Ten datasets were generated using every combination of high ($cv = 0.5$) and low ($cv = 0.1$) values for the four possible sources of variation, giving a total of 160 datasets. [Models 6.1](#) and [6.2](#) were used to analyse the 160 datasets. Gamma distributions were used to model the variability at each level, and priors consistent with those used for [Model 2.2a](#) were used for the variability at each tier. Each model was allowed to run for a maximum of five hours to achieve convergence.

6.2.2.3 Results

[Figure 6.1](#) shows the number of samples required according to the Raftery and Lewis diagnostic by each model. The non-hierarchical model required fewer samples (paired Wilcoxon $p < 0.001$), indicating a lower degree of auto-correlation than the hierarchical model. The non-hierarchical model failed to converge after the initial 10,000 samples for seven (4.4%) of the datasets compared to 32 (20.0%) of the datasets for the hierarchical model, also as

a result of a lower autocorrelation. Both models converged within the time allowed for all datasets.

A comparison of the proportion of 95% CI containing the simulation parameter, mean relative bias, mean relative RMSE and mean relative 95% CI size of the estimate for each of the four modelled sources of variability for both models is shown in Figure 6.2. The mean relative 95% CI size and relative RMSE were very similar between the two models, although the hierarchical produced smaller relative RMSE for three of the four sources of variability and smaller 95% CI for all. The hierarchical model also produced smaller bias for all sources of variability, and a proportion of 95% CI containing the simulation parameter that was closer to the optimal 95% for three of the measures.

6.2.2.4 Discussion

The accuracy of the inference on each of the four sources of variability was quite similar between the two models, although the hierarchical model appeared to perform better in this regard. The variability between faecal boluses and due to the McMasters technique produced the smallest bias, relative RMSE and 95% CI size for both models, possibly indicating that the data contained the most information about these parameters. The proportion of simulation

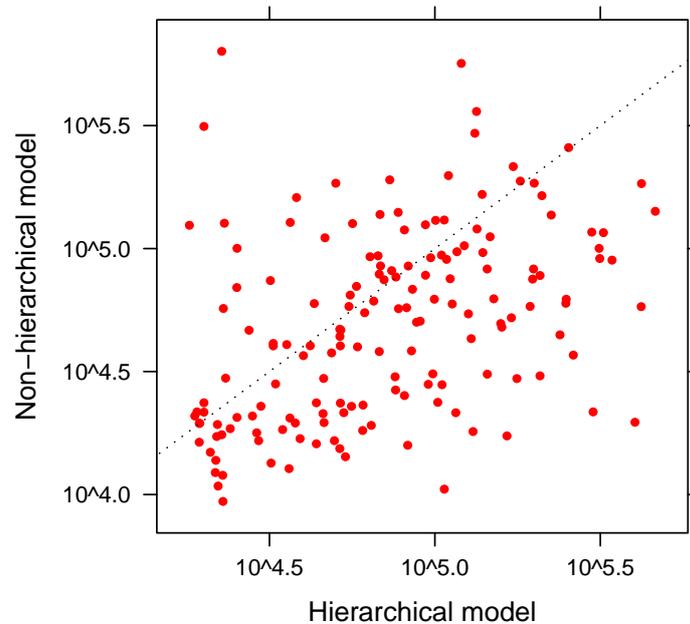


Figure 6.1: The required number of samples calculated by the Raftery and Lewis diagnostic for the hierarchical compared to non-hierarchical MCMC model for each of the 160 simulated hierarchical FEC datasets - equality line shown dashed

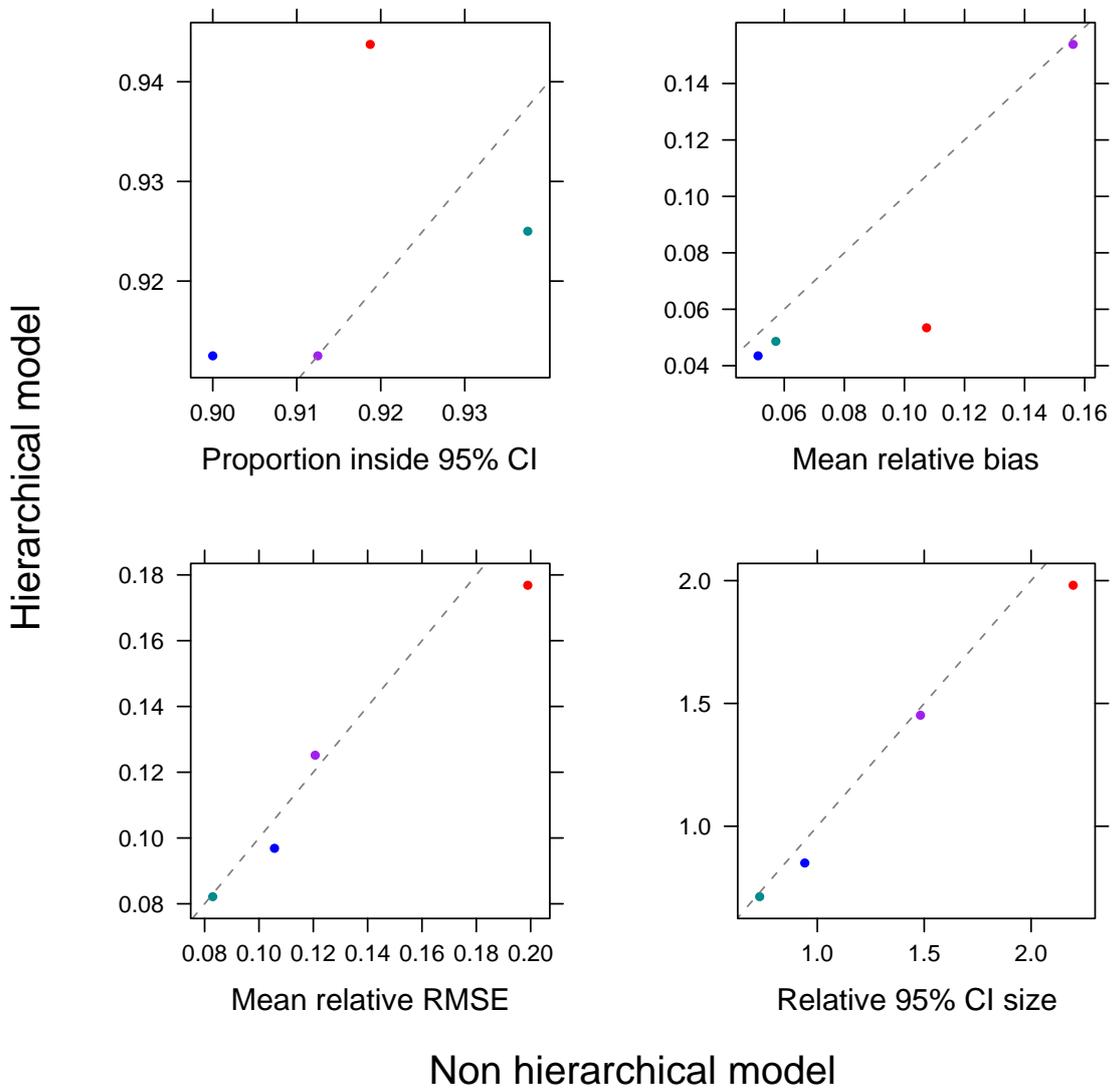


Figure 6.2: The proportion of 95% CI containing the simulation parameter, mean relative bias, mean relative RMSE and mean relative 95% CI size of the estimate for each source of variability produced by the hierarchical (y axis) and non-hierarchical model (x axis), from analysis of the 160 simulated hierarchical FEC datasets - variability between faecal piles shown in red, variability between faecal boluses in blue, variability between samples from the same bolus in purple, variability due to the McMasters technique in cyan.

parameters inside the 95% CI estimates produced by each model were in all cases less than 95%. This is likely an effect of the prior on the inverse shape parameter tending towards lower values of cv , and could be reduced by using uniform priors on cv rather than log-Uniform priors on the inverse shape parameter for consistency with the simulation distributions. The hierarchical model produced 95% CI that were more appropriate for three of the four sources of variability, possibly indicating that the hierarchical model was somehow able to extract more information from the data. Alternatively, the superior performance may be as a result of reduced reliance on Equation 6.3 in the hierarchical model resulting in fewer sampling issues associated with the calculation. However, the apparent improved performance may also be a chance observation.

In the model formulations presented, the variability at each level was assumed to follow a gamma distribution. The cv is a strictly positive parameter, the genesis of which could be considered as a series of multiplicative events, suggesting that a lognormal distribution may be appropriate. This would also be consistent with the usual implementation of a random effect. However, model specifications using the lognormal distribution produced notional 95% CI containing the true parameter only a small proportion of the time, possibly because Equation 6.3, which was derived from the gamma distribution, does not well approximate to the combination of lognormal distributions.

An alternative to using Equation 6.3 in the model specification to separate the variability between samples and due to the McMasters technique could be to model the process more literally, by modelling one distribution representing the variability between 1g sub-samples and sampling from this distribution repeatedly according to the number of grams of faeces used. The mean of these samples could then be used as the mean of the distribution representing the McMasters variability. This has the advantage of reflecting the process more intuitively, but introduces a considerable amount of extra computational effort and produces results that equivalent to using Equation 6.3 to calculate the combined distribution.

One disadvantage of using the hierarchical model appears to be the increased auto-correlation, resulting in a longer burn-in period before convergence is achieved and a larger number of required samples to compensate for the increased Monte Carlo error. This auto-correlation is caused by the inter-dependence of each of the four sources of variability using this model formulation, which is partly overcome by using the non-hierarchical model. This increased auto-correlation means that the hierarchical model has a higher computational requirement, but as long as the sample size and burn-in period are increased accordingly should not detrimentally affect the inference made. The hierarchical model appears to produce more accurate results than the non-hierarchical model, so this formulation will be used in the analysis of real data.

6.2.3 Analysis of equine FEC variability data

6.2.3.1 Introduction

FEC are widely used in horses, and in contrast to the usual implementation with sheep and cattle, are often used as the basis to quantify the parasite burden of individual animals rather than to obtain an indication of the parasite burden of the group as a whole. Factors affecting the variability of FEC from a single individual are therefore more important in horses than in other domestic species. In order to quantify potential sources of this variability, data were obtained from four horses by intensive sampling of faecal piles over a short period of time. The variability introduced by the process of counting eggs ([item 6](#) on the list of variability sources given in [Section 6.2.1](#)) was then examined using a gamma-Poisson model to quantify the over-dispersion, before quantifying the remaining sources of variability within an animal using [Model 6.1](#).

6.2.3.2 Materials and methods

Samples were taken from two individually stabled animals (denoted Animals 1 and 2) for a five day period in December 2007, taking faeces from the floor of the stable during the early morning and mid afternoon. The animals were selected from a group of six on the basis of having the highest worm egg count a fortnight previously, and had been stabled without grazing for four or five months prior to the sampling. For one of the ten sampling events, five samples of faeces were taken from different areas of the faeces and analysed separately, with the McMasters procedure performed on five separate 3g sub-samples for one of these five samples, and the remaining four samples having only one 3g sub-sample taken. The remaining nine sampling events involved a single sample of 3g of faeces taken from a randomly selected faecal bolus. The number of eggs in ten McMasters chambers were counted for each sub-sample.

Further samples were taken from two animals (denoted Animals 3 and 4) over a five day period in December 2008. These were also stabled individually, in a different location to the animals from the previous year. A more intensive sampling procedure was followed with these animals, with ten large samples being taken from the same faecal pile on two occasions. Two of these samples were divided into five 1g sub-samples, five 3g sub-samples and five 9g sub-samples. The McMasters procedure was performed on a single 9g sample for the remaining eight samples, and from each of the samples taken at the remaining eight sampling events. The number of eggs in six McMasters chambers was counted for each of these samples.

The resultant 168 sets of six replicate counts from individual samples, and 36 sets of ten

replicate counts, were analysed to quantify the amount of extra-Poisson variability present. A gamma-Poisson model was fitted to the 204 sets of replicate counts, assuming a common cv but allowing the 204 means to vary independently as fixed effects. The prior for cv used was identical to that used for [Model 2.2a](#).

A slightly modified form of the model developed in [Section 6.2.2](#) was then used to analyse all the data, allowing data from the four animals to be analysed together. The daily variability (variability between faecal piles) and sample variability (variability between faecal boluses) were modelled as fixed effects nested within animals, allowing each animal to have a separate value for each. The replicate variability (variability between samples from the same bolus) was modelled first as a fixed effect nested within animals ([Model 6.3](#)), and then as a single value common to all animals ([Model 6.4](#)). The McMasters variability (variability introduced by the McMasters technique) was fixed between animals for both models. The results obtained with these models were compared to results obtained with a Uniform(0.001, 100) prior directly on cv for each source of variability ([Model 6.3a](#)) to ensure that the results presented are robust to the choice of prior.

6.2.3.3 Results

Only strongyle eggs were detected in the faecal samples, although it was not possible to speciate the eggs further. The minimum, maximum and arithmetic mean of the observed counts is given in [Table 6.1](#). Animal 1 had the highest mean FEC, with Animals 2 and 4 having the lowest. Negative, or zero count, observations were made for all animals on at least one occasion. There was a large degree of variability in faecal egg concentration, from a minimum count of two (100EPG) to a maximum of 17 (850EPG) in the same faecal pat of Animal 1 (mean 454 EPG). If a single FEC had been used to determine if the true FEC of this animal was below a threshold of 350EPG, a false negative diagnosis would have been made in 69 of 180 (38.3%) observations.

The median (95% CI) estimate of the cv of the extra-Poisson variability due to the counting process was 0.028 (0.01- 0.11). [Figures 6.3](#) and [6.4](#) show the posterior probability densities

Table 6.1: Empirical summary of observed FEC from four animals over a five day period, with 18 samples each taken from Animals 1 and 2 and 84 samples each taken from Animals 3 and 4

Min	Mean	Max
0	8.62	24
0	0.37	3
0	2.96	11
0	0.74	5

for each of the remaining four sources of variability in the four animals, using the log scale prior on the inverse shape parameter and uniform prior on cv respectively. Figure 6.5 shows the posterior probability densities obtained when variability between sub-samples taken from the same faecal sample, the replicate variability, is modelled as being independent of the animal. The variance of the posterior estimates for all four sources of variability were reduced, especially the sample and replicate variability.

Figure 6.6 shows the value of combined daily and sample variability and the total within animal variability using Equation 6.1 to combine the estimates for all four variability sources presented in Figure 6.5. The median and 95% CI estimates for the total within-animal variability for each of the four animals is shown in Table 6.2. Animals 2, 3 and 4 showed similar values for total within-animal variability, with a median combined value of 0.32 and combined 95% CI of 0.19-0.68. There was an increased total within-animal variability for Animal 1, as a result of a higher daily variability compared to the other animals (Figure 6.5).

All results presented here have been based on 3g of faeces being used per sample, for consistency with the modified McMasters technique. However, the effective cv due to the replicate variability also depends on the amount of faeces used in each sample, as shown in Figure 6.7. The total within animal variability, calculated using the combined estimates from all animals

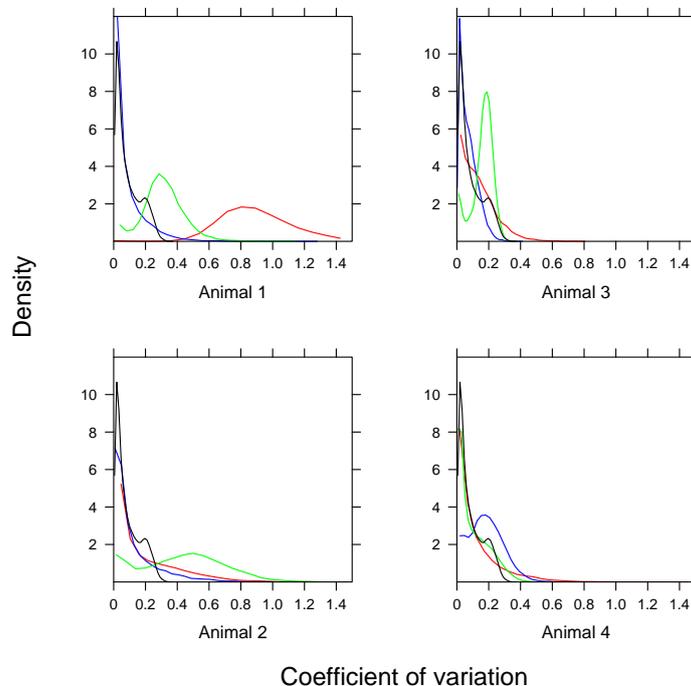


Figure 6.3: The cv associated with each of four sources of variability, all using a log-scale prior on the inverse shape parameter, observed in the FEC of four animals over a five day period - daily (red), sample (blue), replicate (green), McMasters (black) variability

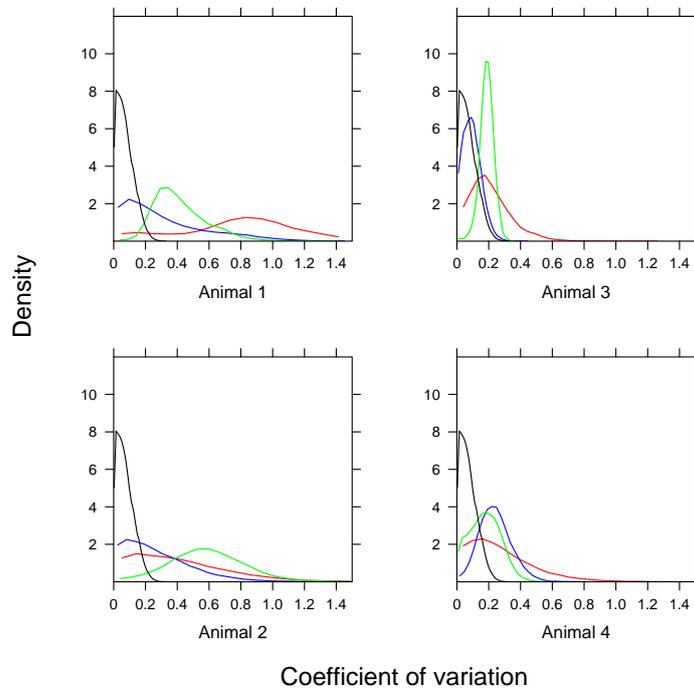


Figure 6.4: The *cv* associated with each of four sources of variability, all using a Uniform prior directly on *cv*, observed in the FEC of four animals taken over a five day period - daily (red), sample (blue), replicate (green), McMasters (black) variability

Table 6.2: Median and 95% CI estimates for the total within animal variability in each animal, assuming replicate variability is fixed between animals and 3g faeces are used per sample, in the FEC of four animals taken over a five day period

	Lower 95% CI	Median	Upper 95% CI
Animal 1	0.62	0.97	1.50
Animal 2	0.19	0.37	0.86
Animal 3	0.21	0.30	0.44
Animal 4	0.19	0.33	0.56

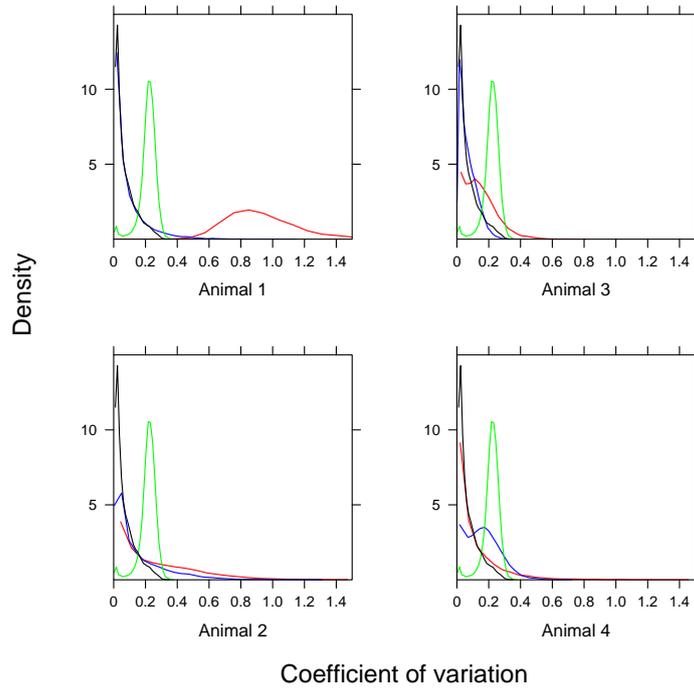


Figure 6.5: The cv associated with each of four sources of variability, assuming replicate variability is fixed between animals, observed in the FEC of four animals taken over a five day period - daily (red), sample (blue), replicate (green), McMasters (black) variability

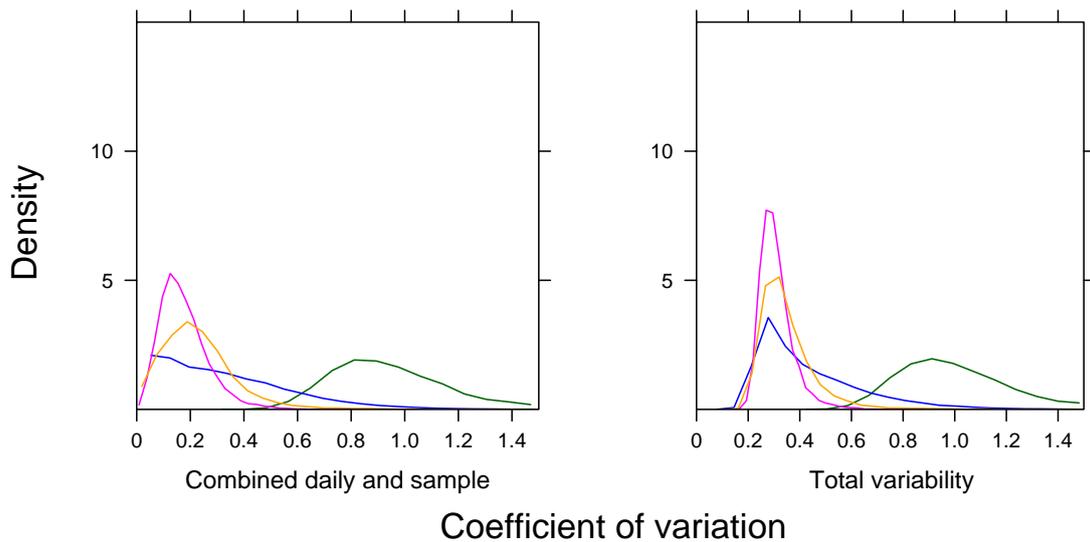


Figure 6.6: Probability density of the combined daily and sample variability (left) and the total within-animal variability (right), assuming replicate variability is fixed between animals and 3g faeces are used per sample, in the FEC of four animals taken over a five day period - animals 1 to 4 shown in dark green, blue, pink and orange respectively.

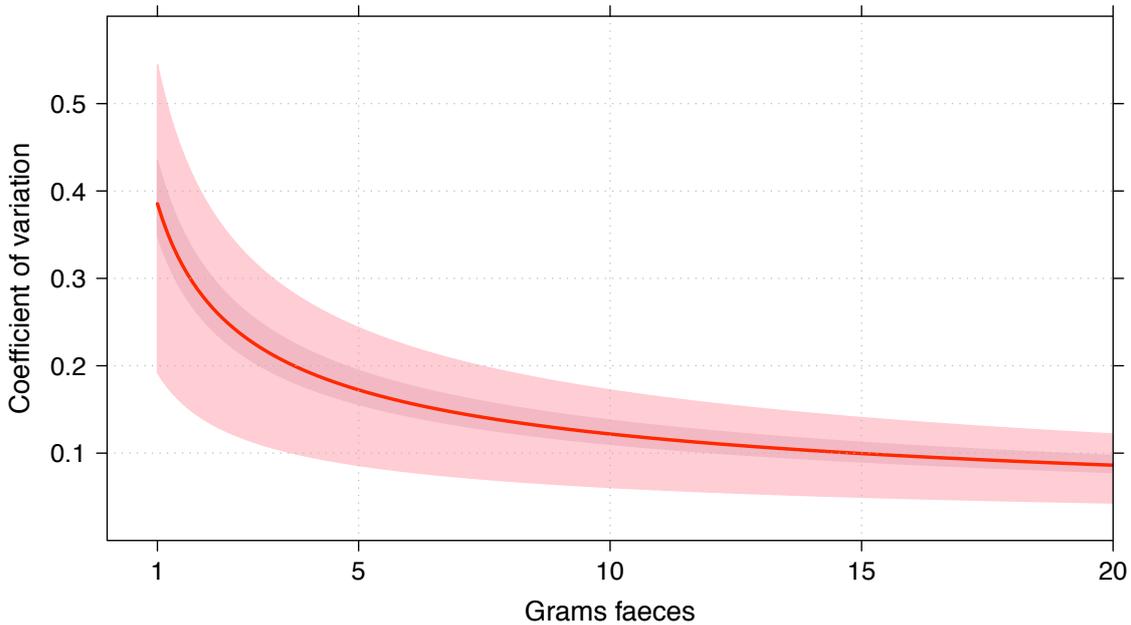


Figure 6.7: The relationship between the amount of faeces used and the effective cv associated with the variability in observed FEC between sub-samples from the same faecal sample - 95% CI shown shaded in pink, 50% CI in dark pink

shown in [Figure 6.5](#), therefore also depends on this measure ([Figure 6.8](#)). Based on these combined estimates, the median (95% CI) estimate for total within animal cv is 0.37 (0.19 - 1.34) for 3g faeces, 0.31 (0.12 - 1.30) for 10g faeces and 0.30 (0.10 - 1.30) for 20g faeces in the four animals observed, assuming a fixed source of replicate variability between animals. This therefore represents a source of variability that is likely to have a clinically relevant impact.

6.2.3.4 Discussion

The extra-Poisson variability between counts taken from the same sample was minimal, indicating that the process of counting eggs is approximated very well by the Poisson distribution. The 95% CI for extra-Poisson cv included the lower limit of the prior used, so it is possible that the data were truly Poisson distributed. The median and upper 95% CI estimates were also small, representing a gamma-Poisson distribution that is almost indistinguishable from a Poisson distribution given the sample size of six and ten counts. This is useful as it means that the sum of the number of eggs in a given number of chambers contains as much information as the individual counts, removing the need to record the counts individually. It also allows the gamma distribution describing the extra-Poisson variability to be discounted from further analyses, reducing the potential for auto-correlation in a multi-level hierarchical model.

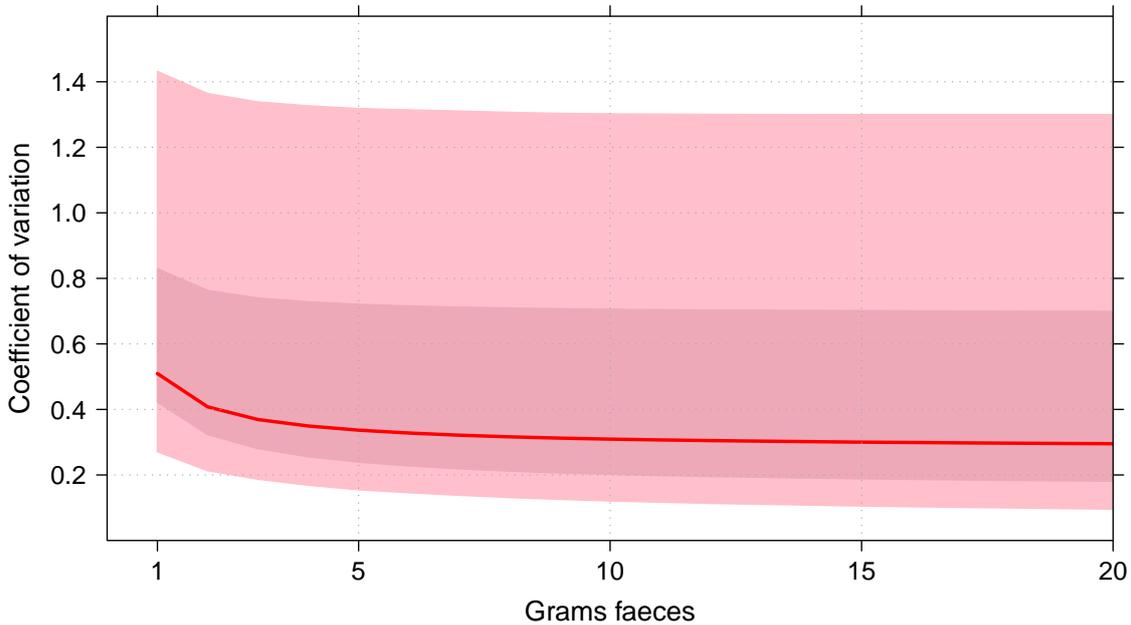


Figure 6.8: The relationship between the amount of faeces used and the total effective cv associated with the variability in observed FEC within an individual over a period of a few days - 95% CI shown shaded in pink, 50% CI in dark pink

Of the four remaining sources of within animal variability, the estimates for sample and replicate variability were quite similar in all animals, although the posterior for sample variability was increased for Animal 4, and the posterior for replicate variability slightly increased for Animal 2 and decreased for Animal 4. However, the 95% confidence intervals of the estimates in each animal overlapped in every case, and were relatively unaffected by the choice of prior. This verifies that the results were predominantly likelihood driven, with the change in prior affecting only the left tail of the posterior distributions. With the replicate variability fixed between animals, the estimate for this source of variability was much larger than that for the McMasters variability and sample variability in animals one to four, although the estimate of sample variability in animal four was only slightly lower than the estimate for replicate variability. The daily variability was also quite consistent between Animals 2, 3 and 4, but markedly increased in Animal 1. [Stear et al. \(1995b\)](#) found that FEC were more consistent when taken from an individual sheep over a short period of time than a longer period of time, suggesting that this source of variability is important for *Ostertagia spp.* infections in sheep. Animal 1 also had the highest mean FEC, which may be coincidental finding but could suggest that the variability between egg concentration of faecal piles is somehow related to the mean FEC shedding rate. It is possible that the stochastic nature of worm fecundity could result in FEC variability that is somehow related to the mean FEC, but this would be expected to decrease variability as mean increases due to the averaging of the egg release timings of more worms. Alternatively, this animal may have been experiencing one of several

factors that could have resulted in a more variable consistency of faecal piles, such as gastrointestinal upset or a change in diet, which would increase the variability in FEC over a short period of time. As the sampling was performed in December it is also possible that the older age of the worms within the animal caused an increase in variability, although this would be expected to be the same for all animals. More work clearly needs to be done to study this further. Another potential source of variability that was not directly examined in this study is a potential diurnal variability in FEC. It has been suggested anecdotally that FEC in horses have a tendency to be higher in the morning than the afternoon or evening, although [Bennett \(1990\)](#) found no such tendency in the study of repeated FEC from ten individuals. There was also no significant evidence of diurnal variability between the samples collected in the morning and afternoon in this study. If such a diurnal variation truly exists, it would be incorporated into the faecal pile variability using the statistical analysis demonstrated here.

For the purposes of this study, the variability introduced by the McMasters technique was assumed to be the same for all animals. This is justifiable as the laboratory work was carried out by one of two personnel in the same laboratory using the same procedure. The importance of this source of variability may differ with laboratory, although given the relative simplicity of the McMasters technique this seems unlikely as long as mixing of the samples remains sufficient. The variability between replicates from the same sample of faeces was modelled both as a fixed effect of animal, and assuming an equal source for all animals. The former is obviously more conservative, but conceptually replicate variability could be expected to be dependent on factors such as the time between releasing of eggs from worms and the peristaltic activity and length of time eggs remain in the intestine, which should be more similar between animals than other factors such as faecal consistency. This could be altered by host characteristics such as age and breed and the predominant site of parasitism and species distribution of parasites, and could vary substantially between host species. Further work to examine these effects would certainly be desirable before reliance is placed on this work in a clinical setting, but the estimates obtained assuming a fixed source of replicate variability will tentatively be used here applied to adult horses similar in age and breed to those sampled from.

6.3 Quantifying variability between individuals

6.3.1 Introduction

The variability in FEC within an individual animal has been quantified in [Section 6.2.3](#), although it was not possible to make inference on the variability between animals using this data as the four animals sampled were not grazed as a group. The total between animal variability was estimated for data obtained from several groups of horses in [Chapter 4](#), and

can be combined with the estimates for within animal variability obtained in this chapter to obtain estimates for the variability in true mean FEC between animals. This assumes that the variability between FEC observed in the four animals studied is representative of the wider equine population. Given the discrepancy between the variability between faecal pats observed in Animal 1 compared to Animals 2, 3 and 4, this may not be the case. However, these are the only estimates currently available, and a cautious comparison of the relative importance of within and between animal variability was considered appropriate.

6.3.2 Materials and methods

The 24 datasets with a sample size of 50 or more from the datasets presented in [Section 4.3](#) were used to estimate the total variability between animals. Each dataset was analysed using the ZIGP [Model 2.2a](#) to obtain a posterior distribution for the total variability (cv_t), from which five thousand samples were taken for each dataset. Five thousand random samples of total within animal variability (cv_w), based on the use of 3g of faeces, were taken from the combined posterior estimates obtained from the four horses studied in [Section 6.2.3](#). These five thousand estimates of cv_t and cv_w were then combined to yield five thousand estimates of the true between animal variability, cv_b , using [Equation 6.4](#) (derived from [Equation 6.1](#)).

$$\begin{aligned}
 cv_b^2 + cv_w^2 + cv_b^2 cv_w^2 &= cv_t^2 \\
 cv_b^2 + cv_b^2 cv_w^2 &= cv_t^2 - cv_w^2 \\
 cv_b^2 (1 + cv_w^2) &= cv_t^2 - cv_w^2 \\
 cv_b^2 &= \frac{cv_t^2 - cv_w^2}{1 + cv_w^2}
 \end{aligned}
 \tag{6.4}$$

Median, 95% CI and 50% CI estimates were then calculated for cv_b for each of the 24 datasets, and compared to the same quantile estimates for cv_w .

6.3.3 Results

[Figure 6.9](#) shows the 95% CI (shown in light pink), 50% CI (darker pink) and median estimate (red) for cv_w , relative to the 95% CI (blue lines), 50% CI (light blue rectangles) and median (dark blue points) for cv_b in the 24 datasets. The 95% and 50% CI estimates for cv_w and cv_b overlapped for every dataset, although the upper 50% CI was higher for cv_b in every case and the upper 95% CI for cv_b higher for all but two datasets. The median estimate for cv_b was also higher than that for cv_w in every case.

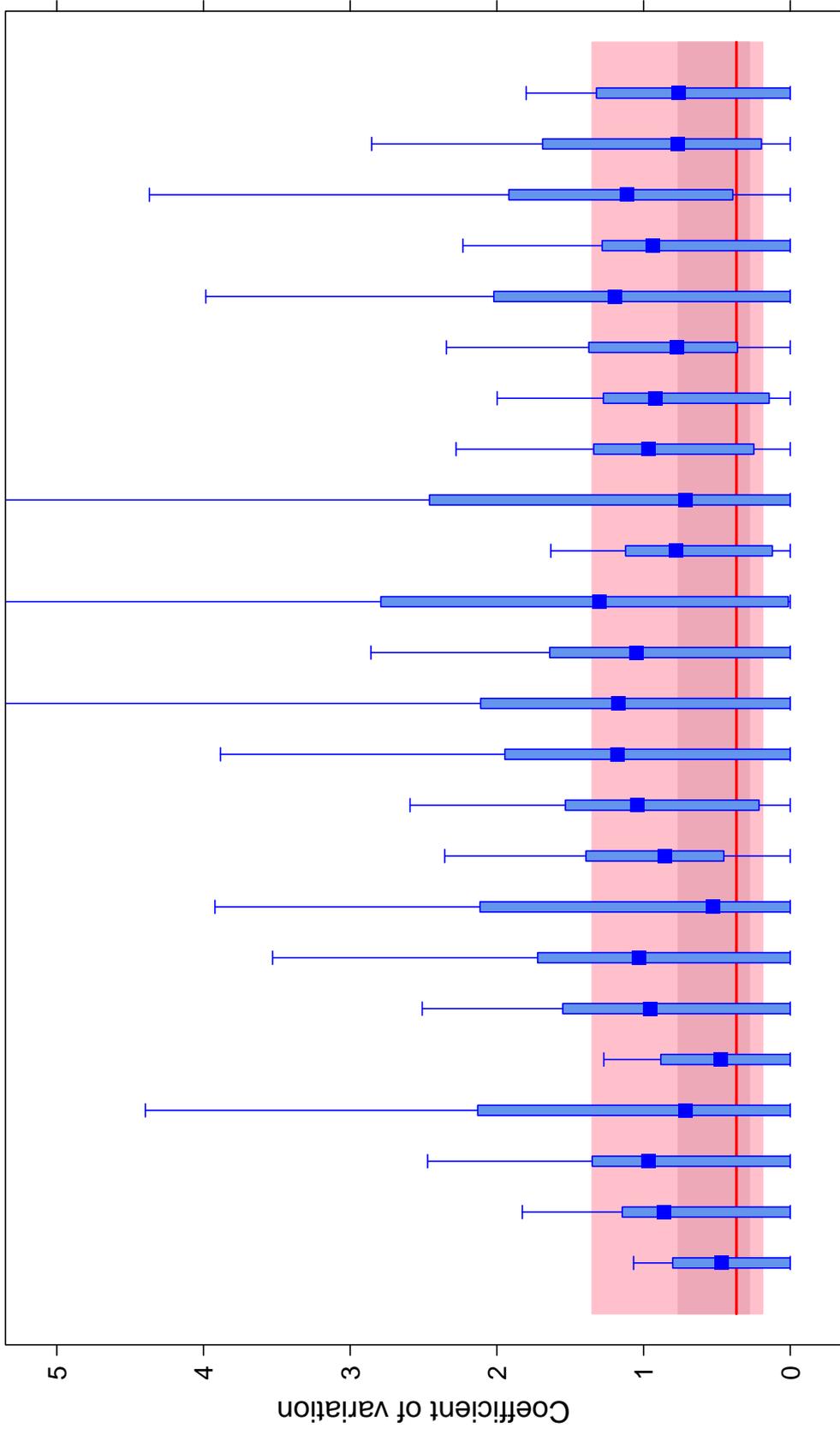


Figure 6.9: Estimates for the between animal variability (cv_b) obtained from 24 equine datasets with sample sizes of 50 or more, and within animal variability (cv_w) obtained from a total of 204 samples taken from four animals over a five day period - 95% CI, 50% CI and median estimates for cv_w , shown in light pink, darker pink and red respectively. 95% CI, 50% CI and median estimates for cv_b , shown using blue lines, light blue rectangles and dark blue points respectively.

6.3.4 Discussion

Based on the results presented here, it appears that between animal variability is a more important source of observed variability in FEC than within animal variability in horses. The median estimate for within animal variability was 0.37, compared to the median estimates of variability between animals of between 0.47 and 1.30, although confidence intervals were relatively wide as a result of large confidence intervals for the total variability and large confidence intervals for the within animal variability. However, these results indicate that the variability within an animal is sufficient to represent an important contributor to the observed variability in FEC in the majority of these datasets, and may even be equally or more important than the variability between animals in some datasets.

The results presented here are dependent on the four animals studied in [Section 6.2.3](#) having similar components of within animal variability to the animals studied in [Chapter 4](#). The validity of this assumption is not verifiable with the available data, and the inference on between animal variability may be different if information regarding the variability within animals was available from the datasets presented in [Chapter 4](#). Corroboration of the results obtained here in further groups of animals would therefore be essential before relying on the inference made here for clinical applications. However, the data presented here are sufficient to conclude that variability within an animal is likely to be an important contributor to the observed variability in FEC, and should be accounted for in statistical analysis of FEC data. In light of this, many of the assumptions underlying the study designs commonly used to perform a FECRT are likely to be violated. This supports the conclusion made in [Section 5.5](#) that exclusion of animals from post-treatment sampling on the basis of their pre-treatment FEC is not justified and likely to introduce bias. It also suggests that repeat sampling within an individual would be beneficial to improve the accuracy of individual estimates of true mean FEC.

6.4 Precision analysis

6.4.1 Introduction

The ability to predict accurately the true value of an indirectly observed parameter, such as the true mean FEC of a group or individual, is here referred to as the ‘precision’ of the procedure. When considering a hypothesis driven test such as a FECRT, it is also useful to consider the ‘power’ of the procedure, or the probability that a significant difference between the true FEC reduction and desired FEC reduction will be observed. The most common use of power calculations is as an aid to study design, when attempting to calculate the number of samples that need to be collected in order to be able to detect a given value of difference

in population means, at an appropriate significance level. Consideration of the study power may also be useful when reporting the significance level of the results of a hypothesis test, and when considering summary statistics of a sample of data, because these methods do not provide any perspective of how much discrepancy is expected between the population and sample means. However, there are also several disadvantages to the use of post-hoc power calculations in this way, and methods of quantifying uncertainty using confidence intervals are certainly preferable (Hoenig and Heisey, 2001).

Calculating the precision or power of a study before data are collected also provides an opportunity to change the sampling procedure to optimise the amount of data collected, or to abandon the study altogether if the calculated precision is so small that useful results are unlikely to be obtained. This saves time, conserves laboratory resources and is vital when performing studies with an animal welfare concern. Although rarely considered for this purpose, precision calculations can also be applied to the collection of FEC data, by calculating the number of faecal samples that need to be collected in order to be able to calculate the population or true individual mean FEC to within the desired accuracy at an appropriate confidence level. In this case, the desired confidence level can be interpreted as the proportion of the time that the observed mean will lie within the desired accuracy range of the true mean. A similar mechanism can be considered for the FECRT, and has been published by Gill *et al.* (1986) although does not seem to have been widely used clinically. By considering the precision of FEC data collection when designing studies or deciding on the minimal egg detection threshold of the technique to be used, the optimal combination of laboratory time and usefulness of resultant data can be achieved.

There are conceptually two subtly different parameters that could both be referred to as the ‘mean faecal egg count’ of a population. The first is the population mean egg count of the distribution from which the sampled animals are derived, conceptually representing a population of an infinite number of animals. The second is the true mean egg count of the animals sampled, which represents the mean of those animals only, and is an estimate of the true mean of a population from which the samples were taken, or one realisation of potential means for that group of animals. This parameter is likely to be more relevant when all individuals in a group are sampled, and when we are interested in the pasture contamination or egg count reduction of those animals. The first measure would be more appropriate when only a random representative sample of animals is taken from a larger group of animals, as the true mean egg count of the sampled animals is an estimate for the true mean of the larger population. This parameter will always be more difficult to estimate because the possible variation in mean egg count from unobserved animals has to be taken into account. Both parameters are relevant in different situations, so both will be examined here. In order to avoid confusion the population mean egg count of the distribution from which the sampled animals are derived will be referred to as the ‘population mean’, whereas the true mean egg count of the sampled animals will be referred to as the ‘true sample mean’.

The standard error of a mean FEC will depend on the variability between counts (further subdivided in [Section 6.2.1](#)), as well as the number of samples taken. The variability (cv) associated with the Poisson process of counting eggs decreases with increasing mean. This is because the variance is equal to the mean, so the cv is therefore equal to $\frac{\sqrt{\text{mean}}}{\text{mean}}$. The mean number of eggs counted (mean of the Poisson distribution) will therefore also have an effect on the precision. In addition, the proportion of the time that a mean estimate will lie within a given accuracy of the population or true sample mean will be affected by the given accuracy (tolerance). The possible factors that could affect the precision of a mean FEC are therefore:

1. The number of animals sampled (not including those, if any, with means that are truly zero and are therefore not part of the distribution describing other animals)
2. The variability between animals
3. The number of samples taken from each individual
4. The variability (cv) between true mean counts, further subdivided into:
 - The variability between samples (the total variability between faecal piles and between samples from within a pile)
 - The amount of faeces taken per sample
 - The variability between 1g replicates
 - The variability introduced by the McMasters technique
5. The uncertainty introduced by the counting process, further subdivided into:
 - The cv of the counting process (estimated at ≈ 0 in [Section 6.2.1](#), so discounted here)
 - The mean number of eggs counted per sample, which is dependent on:
 - The number of McMasters chambers counted
 - The true mean FEC of the animals
6. The desired accuracy

The relationship between these factors and the precision of a FEC was explored. This work is presented in the remainder of this chapter.

6.4.2 Precision analysis of a FEC from an individual animal

6.4.2.1 Development of the parametrisation

The probability that a single observed faecal egg count will be within a given tolerance of the true mean of an individual depends on several factors as discussed previously, but calculating this for the true mean a sample is relatively simple. The probability that an arithmetic mean observation ($\hat{\mu}$) will be $\pm T * \mu$ (where T is the tolerance and μ is the true mean number of eggs counted), and is given by the following probability expression:

$$\begin{aligned} \text{Probability} &= \text{P}(\hat{\mu} \leq \mu(1 + T), \quad \hat{\mu} \geq \mu(1 - T) \mid \lambda) \\ \text{where :} \\ \lambda &= \mu \end{aligned}$$

This probability can be calculated using the cumulative probability function of the Poisson distribution. The variability (as *cv*) of the Poisson distribution decreases with increasing mean, so the power as given above will therefore increase with increasing μ . This expression requires that an equal proportion of faeces is observed to that on which the mean is based, but can also be expanded to represent either a proportion or multiple of this, obtained from multiple observations from the same sample, using a basic property of the Poisson distribution. The sum of a series of independent Poisson observations are Poisson distributed with a mean equal to the sum of the independent Poisson means (Evans et al., 2000), so that:

$$\begin{aligned} \text{Power} &= \text{P}(O \leq \mu(1 + T) \times S, \quad O \geq \mu(1 - T) \times S \mid \lambda) \\ \text{where :} \\ \lambda &= \mu \times S \end{aligned}$$

The additional parameter S reflects the proportion, or multiple, of the sample that is observed (denoted the counting sensitivity), and O the total number of observed eggs. The overall precision therefore depends on the mean number of eggs in the total observed sample, rather than either the counting sensitivity of the technique or the mean FEC directly.

This can be extended to calculate the probability that a FEC will be within a tolerance of the true mean of an individual by accounting for the variability within an individual. This single source of variability is actually comprised of variability between samples (cv_s), variability

between 1g replicates (cv_r), variability introduced by the McMasters technique (cv_m), and the amount of faeces taken per sample (g). However, if only one replicate is performed per sample then these individual sources of variability are inseparable, so only the combination needs to be considered for a precision calculation. This combination is given by:

$$\begin{aligned}
 cv_w^2 = & cv_s^2 + cv_m^2 + cv_a^2 + \\
 & (cv_s^2 * cv_m^2) + (cv_s^2 * cv_a^2) + \\
 & (cv_a^2 * cv_m^2) + (cv_s^2 * cv_m^2 * cv_a^2)
 \end{aligned}
 \tag{6.5}$$

where :

$$cv_a = cv_r * \frac{1}{\sqrt{g}}$$

If this combined source of variability is considered to approximate well to a gamma distribution, it can be incorporated into a closed form calculation by using the negative binomial in place of the Poisson distribution. The probability is then given by the following:

$$\begin{aligned}
 Probability = & P(O \leq \mu(1+T) \times R \times S, \quad O \geq \mu(1-T) \times R \times S \quad | \quad \alpha, \beta) \\
 & where :
 \end{aligned}
 \tag{6.6}$$

$$\alpha = \frac{R}{cv_w^2}$$

$$\beta = \mu \times S \times cv_w^2$$

Where the additional parameter R represents the number of repeat samples taken, S the proportion of a gram observed per sample, or the minimum egg detection threshold or counting sensitivity, and O the total number of observed eggs. This probability can then be calculated using the cumulative probability function of the negative binomial.

6.4.2.2 The effect of different parameters on precision

As previously stated, the precision of a FEC depends on the total number of eggs counted rather than the number of samples, counting sensitivity or mean FEC directly. The probability that an observed mean FEC will be within a tolerance of the true mean also depends on the desired tolerance, T , because the probability that a randomly drawn sample will lie within a range of values is dependent on the relative size of the range. This probability with

negligible cv_w can therefore be expressed as a non-linear function of R , S , μ and T . Calculating the probability using a negligible value of cv_w and every combination of parameter values with R drawn from the set $\{1, 2, 3, 4, 5\}$, S drawn from the set $\{0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20\}$, μ drawn from the set $\{200, 400, 600, 800, 1000\}$ and T drawn from the set $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.07, 0.1, 0.15, 0.2, 0.5\}$, and plotting the resultant values against $R \times S \times \mu \times T^2$ yields **Figure 6.10**. A small amount of variation around the agreement line is observed due to the integer nature of FEC having different effects at different values of counting $S \times \mu$ (shown in darker colours on **Figure 6.10**), but a constant value of $R \times S \times \mu \times T^2$ can be considered to yield approximately the same probability value given a negligible value of cv_w .

Examining the affect of cv_w introduces a degree of complexity to the relationship between R , S , μ , and the precision of a FEC. This can be simplified by considering only the value of T that yields an ‘acceptable’ probability, for example 95%, that the observed mean FEC will be within tolerance of the true mean FEC. The value of T for which the observed mean FEC will lie $\pm T \times$ the true mean FEC 95% of the time given the combination of R , S , μ and cv_w

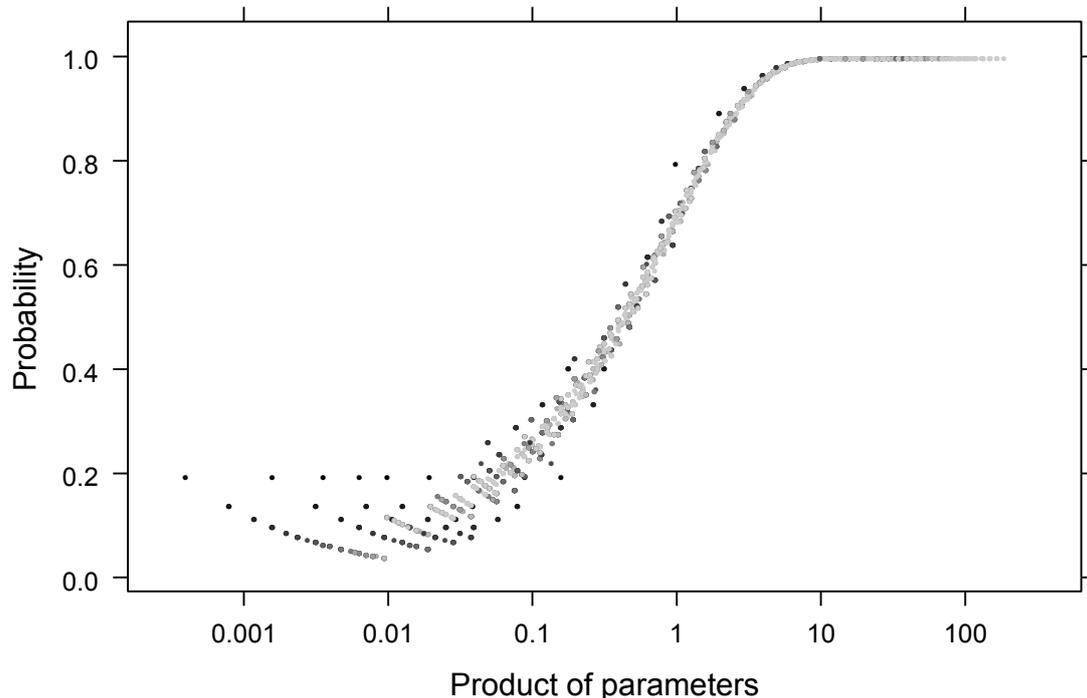


Figure 6.10: The effect of $R \times S \times \mu \times T^2$ on the probability that an observed mean FEC will lie within a tolerance (T) of the true mean FEC of an individual animal with a negligible cv_w of 0.001, calculated using the cumulative probability density function of the negative binomial distribution - smaller values of $\mu \times S$ shown as darker grey

in the dataset can be calculated using an optimisation algorithm. A three dimensional plot of minimum required $T^2 \times R$ to yield 95% power given cv_w and $\mu \times S$ is given in Figure 6.11. All optimisations were performed using a large value of $R = 100$ to allow more precise calculation of T with low μ and S . This reduces the impact of the integer nature of FEC with low mean and/or counting sensitivity, and avoids the situation where the probability can not be close to 95% due to the optimal tolerance requiring theoretical fractions of observed eggs. As would be expected, an increase in cv_w results in a larger required T in order to maintain the same 95% probability, regardless of the mean, counting sensitivity and number of repeat samples. With negligible cv , the required T is decreased with increasing mean and S in line with Figure 6.10. With larger values of cv , the mean and S has little effect provided that the product of these is greater than between 5 and 200 depending on the cv_w . The only method

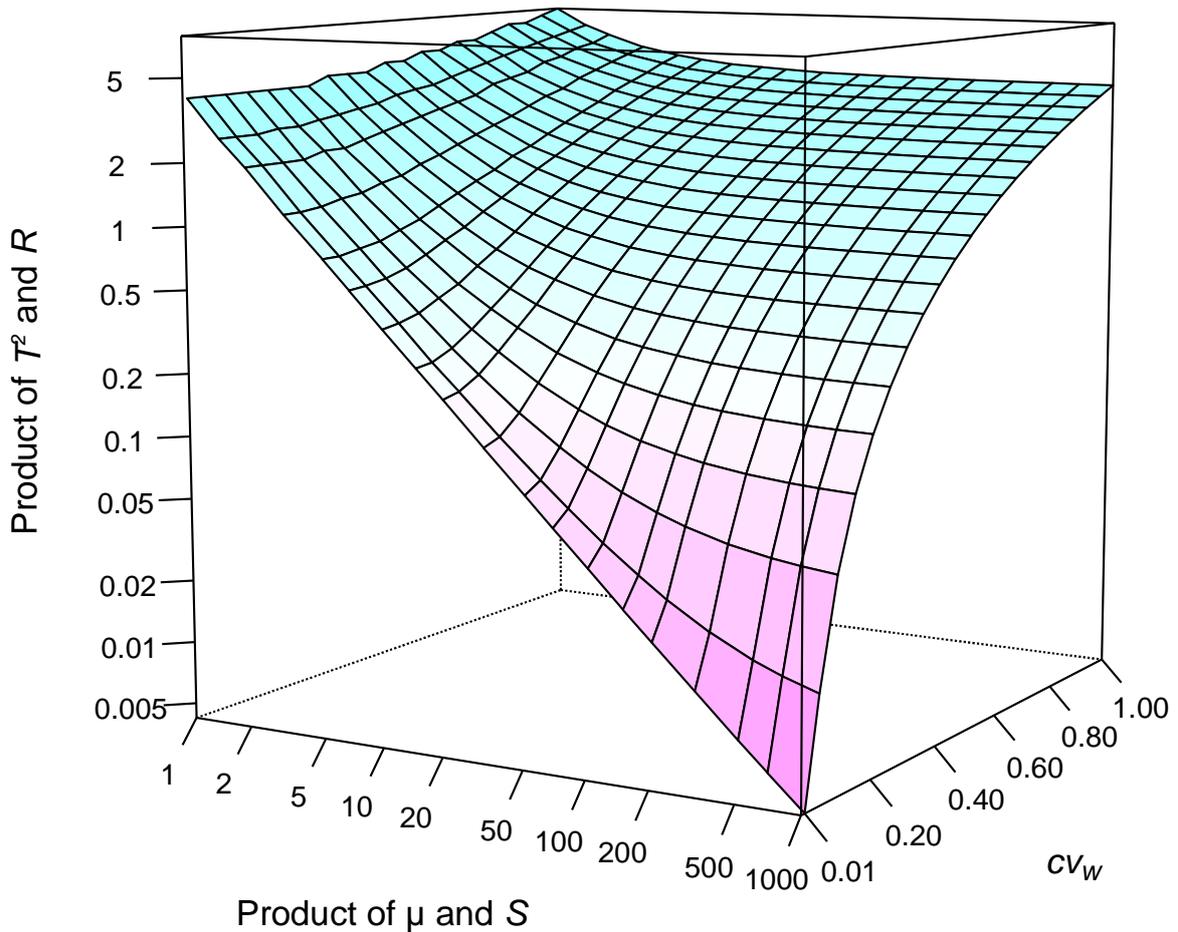


Figure 6.11: The effect of $\mu \times S$ and cv_w on the $T^2 \times R$ required to maintain a 95% probability of observing a mean FEC within $\mu \pm T \times \mu$ (the true mean FEC) from an individual animal, calculated using an optimisation algorithm and the cumulative probability density function of the negative binomial distribution

of obtaining a more accurate estimate of the mean FEC in such circumstances is therefore to take more replicate samples, or to reduce the cv_w by increasing the amount of faeces used per sample.

The separate effects of the minimum egg detection threshold (S), mean FEC and number of replicate samples taken (R) with a fixed $cv_w = 0.37$ (the median estimate based on 3g faeces obtained from Section 6.2.3) on the tolerance with which a 95% probability is achieved is given in Figure 6.12. When cv_w is fixed at this value, the most important factor on the precision of a FEC is the number of replicate samples taken. The minimum egg detection threshold has less effect with a larger mean FEC, and increasing this provides diminishing returns beyond approximately $\frac{1}{12.5}$ EPG (equivalent to counting the number of eggs in four McMasters chambers) with means of 500EPG and greater. Using only a single sample and with a counting sensitivity of $\frac{1}{50}$ EPG, the empirical mean observed FEC was $\pm 1 \times$ the true mean FEC 95% of the time, which is unlikely to be sufficient on which to base clinical decisions. Even with a mean FEC of 1000EPG and low egg detection threshold, the empirical

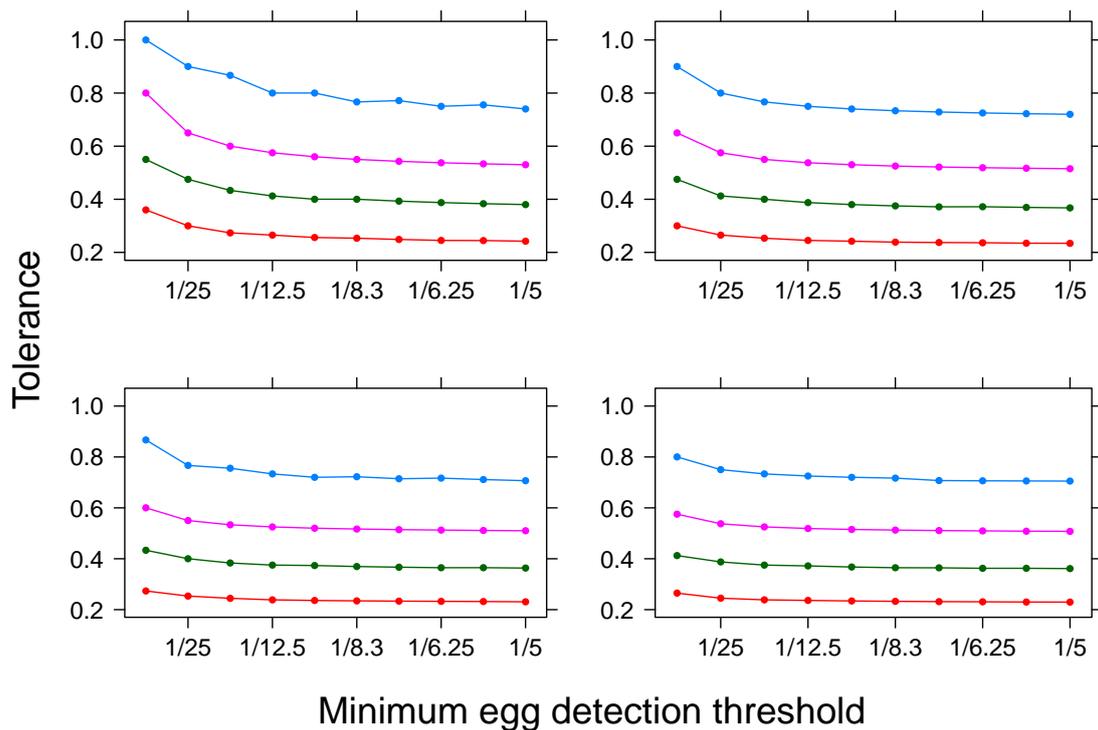


Figure 6.12: The effect of varying the minimum egg detection threshold (S) on the required tolerance (T) to maintain a 95% probability of observing $\mu \pm T \times \mu$ (the true mean FEC) from an individual animal, calculated using an optimisation algorithm and the cumulative probability density function of the negative binomial distribution - mean FEC of 250 (top left), 500 (top right), 750 (bottom left) and 1000 (bottom right). One, two, four and ten replicate samples shown in blue, pink, dark green and red respectively. cv equal to 0.37 for all plots.

mean observed FEC was $\pm 0.6 \times$ the true mean FEC 95% of the time. By taking four samples this can be reduced to a little over $\pm 0.3 \times$ the true mean FEC, or $\pm 0.2 \times$ the true mean FEC with ten replicate samples.

6.4.3 Precision analysis of a FEC from a group of animals

Extending this to analyse the precision of a true mean FEC of a group of animals requires that the variability between animals and the number of animals are taken into consideration. The latter is relatively simple to incorporate into the parametrisation of the negative binomial by multiplying the scale parameter by the number of animals sampled. It is more difficult to incorporate the variability between animals. Only one sample is taken per animal, therefore it is possible to combine the variability observed within an animal (cv_w) with the variability between animals (cv_b) to obtain an effective variability that describes the single distribution observed. This assumes that animals are sampled at random from a herd with an effectively infinite population, which is rarely the case. In reality, each of the available animals is likely to be sampled once, a form of structured sampling, which results in an effective variability of less than would be expected from simply combining the variability within and between animals into a single distribution. This effect increases when the variability between animals increases relative to the variability within an animal, and is negligible when $cv_b \ll cv_w$ as all animals have a similar individual mean FEC reducing the importance of which animals are sampled. However, this is unlikely to be the case in most populations. It is also impossible to estimate the precision of a FEC procedure for the true sample mean rather than the population mean (as explained in [Section 6.4.1](#)) in this way. The true sample mean represents the true mean FEC of the observed animals, which is more likely to be of interest clinically than the population mean which describes some theoretical distribution from which the observed animals, or one realisation of the observed animals, were drawn. There are some situations where the population mean would be more useful to consider, such as when not all animals in a population were sampled, or when considering the true difference in FEC between different groups of animals.

The effects of structured sampling and requirement to be able to calculate the precision for the true sample mean preclude the use of a closed form solution based on the negative binomial, although it is possible to use a simulation model to do this. A conceptual outline of this model is explained below.

- The mean egg count of each individual is derived from a distribution which describes the mean egg count of each individual in that population. The variability of this distribution accounts for the variability in mean egg count between individuals, and the mean of this distribution describes the population mean egg count. The true sample mean is

taken as the average mean egg count of all sampled individuals, which will converge on the population mean as the number of animals sampled tends to infinity.

- The number of eggs in each gram of faeces within an animal is taken from a distribution which describes the number of eggs in every gram of faeces within that animal. The mean of this distribution is equal to the mean egg count of the individual, derived from the distribution above, and the variability of the distribution describes the variability in the number of eggs between processed samples. This is a combination of the variability between samples, the variability between 1g replicates, the amount of faeces taken per sample, and the variability due to the McMasters technique.
- Eggs are randomly distributed throughout each homogenised sample of faeces. The number of observed eggs is Poisson distributed with mean equal to the product of counting sensitivity and the number of eggs per gram of the sample.

The vector of observed means output by this model can be used to either calculate the probability that an observed mean will be within a tolerance of the true mean, subject to uncertainty given by a beta distribution, or to calculate quantiles corresponding to a desired probability of the same, with uncertainty, from which the required tolerance can be derived. This is equivalent to optimising the tolerance to achieve the desired probability discussed for precision analysis using the negative binomial. This approach can also be extended to calculate the precision associated with a FECRT by including two sets of FEC from the same animals, allowing for each set of FEC to have different values of mean, cv_b and cv_w . For this model, additional parameters reflecting the change in true mean FEC and change in variability within and between animals are required. As with the precision for a group mean FEC, the precision of a FEC reduction can also be considered in terms of the change in true sample mean and change in population mean, although the change in population mean provides a more conservative estimate of the precision given that some parasites may be *in refugia* at the time the test was carried out. This is also more consistent with the population efficacy as calculated in [Chapter 5](#). It is also possible to extend these calculations to calculate the power of a FECRT by also determining if each observable reduction would indicate a significant difference to a given required reduction given a true reduction of less than this value. This simply requires calculation of the distribution of observed values, using the same mathematical model, given the null hypothesis of a true reduction equal to the required reduction, and determination of the lower 0.025 quantile (assuming that the desired two-tailed significance level is 5%) of this distribution. As any value equal to or below this significance threshold is extreme given the null hypothesis, it can be considered significant with no need to calculate the actual p-value. The mathematical model would then be repeated using an alternative value for the true reduction, and the proportion of observed reductions that are below the calculated significance threshold calculated. This value, with

appropriate confidence intervals calculated using a beta distribution, is equal to the power of the procedure.

The precision and power calculations discussed above assume that all required parameters, such as the variability, mean FEC etc, are known, but it is unlikely that this would be the case in practice. An extension of this work to include uncertainty in some of the parameters of the population can be achieved by using Monte Carlo integration. For example, the desired tolerance, number of animals, counting sensitivity and number of grams of faeces used are likely to be known, but the true mean FEC, variability between animals and variability within an animal is subject to uncertainty. In this case, the probability that an observed mean will be within a tolerance of the true mean can be calculated using several independent random samples from distributions describing the uncertainty in true mean FEC, variability between animals and variability within an animal, and a probability density distribution describing uncertainty in the true probability obtained. Another alternative to using a value of tolerance (where *tolerance* = T and *probability* = $P(\hat{\mu} \pm T \times \mu)$) is to specify only a lower limit or upper limit on the tolerance, i.e. *probability* = $P(\hat{\mu} < U)$ or *power* = $P(\hat{\mu} > L)$, where L and U are the specified lower and upper limits respectively. This formulation may be more useful in considering the power of a FECRT, where it is important to consider the probability that the observed mean ($\hat{\mu}$) is significantly below an efficacy threshold given that the true mean (μ) is above this threshold (the probability of a false positive), or the probability that the observed mean ($\hat{\mu}$) is above an efficacy threshold given that the true mean (μ) is below this threshold (the probability of a false negative). This could be combined with the use of Monte Carlo integration to quantify the uncertainty in the true false positive or false negative rates given uncertainty in one or more population parameters. For example, if a true reduction of less than 95% were considered to represent a positive case of drug resistance, then the probability that an empirical mean reduction that is significantly less than 95% will be observed after treatment of 20 animals with a drug of true population efficacy 85% could be calculated given uncertainty in the pre-treatment mean and variability sources, with one 3g replicate per animal and a minimum egg detection threshold of 5 EPG. This probability represents the power of the FECRT, or one minus the false negative rate. For the unknown population parameters, the pre-treatment mean FEC may be predicted to be between a minimum of 50 EPG and a maximum of 400 EPG with a most likely value of 200 EPG, the pre-treatment replicate and individual *cv* predicted to be between 0.1 and 0.5 with a most likely value of 0.3, the pre-treatment *cv* between animals predicted to be between 0.2 and 1.2 with a most likely value of 0.7, and each post-treatment *cv* predicted to be between 1 and 1.5 times the corresponding pre-treatment *cv* with a most likely value of 1.2 times the corresponding pre-treatment *cv*. In this case, the distribution describing the uncertainty in the power is shown in [Figure 6.13](#).

Mathematical models written in a mixture of R and C based on the calculations demonstrated in this chapter are included in the contributed R package ‘bayescount’ (see the list

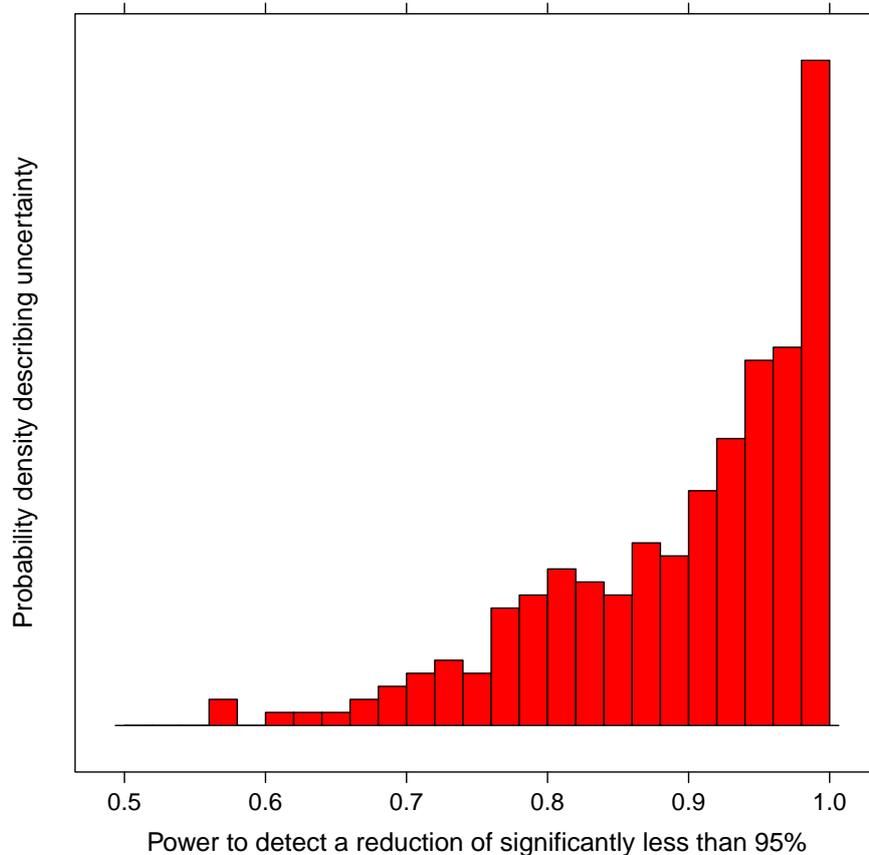


Figure 6.13: Distribution describing the uncertainty in the power to detect an empirical mean reduction of less than 95% after treatment of 20 animals with a drug of true population efficacy 85% - obtained using 250 Monte-Carlo samples from distributions describing uncertainty in the mean pre-treatment FEC and degree of variability in the data.

of publications for more details). This allows the precision for the true sample mean FEC and population mean FEC to be calculated for each theoretical combination of parameters given above, and for the power associated with a FECRT to be calculated. For the precision analysis for the true sample mean, the simulated individual animal means are adjusted so that the sample mean is identical to the population mean. Because of the iterative nature of the simulation, the estimate for power is subject to uncertainty given by a beta distribution. The R package also contains a function to calculate the precision using a negative binomial distribution, which can be used to calculate the precision for an individual animal FEC, or the population mean FEC where animals are randomly drawn from a very large population or the variability between animals is considered to be negligible compared to the variability within an animal.

6.5 Discussion

FEC data in clinical parasitology are widely accepted to be quite variable ([McKenna, 1981](#); [Uhlinger, 1993](#); [Hoste et al., 2002](#); [Miller et al., 2006](#)), but the statistical reasons for this poor repeatability have not been well studied. Where FEC data are taken from a group of animals, a common assumption is that the predominant variability in the observed FEC is as a result of variability in parasite shedding between animals, be it due to genetic, exposure level, or other factors, and the variability in FEC within an animal is therefore unimportant. The estimates of the relative importance of each potential source of variability demonstrated here indicate that this is not the case, although the relatively small number of animals from which the data were taken should be considered. It would be beneficial to apply the methods developed here to data obtained from other animals, so that consistency of each source of variability in the wider population can be assessed. Analysis of data obtained from animals of different breeds and age groups should also be made. Further studies in sheep and cattle would also be of interest, to examine the effect of different parasite locations within an animal on the variability of fecundity, although an analysis of the use of composite FEC in sheep suggests that less variability may be observed in this species ([Morgan et al., 2005](#)). [Hunter and Quenouille \(1952\)](#) consider the between samples taken from an individual sheep to follow a Poisson distribution very closely, although it is not clear if they are referring to a single sample or multiple samples from an individual.

In the horses examined in this study, the replicate variability, or variability between subsamples taken from the same sample of faeces, appears to be a large source of the total within animal variability using the standard McMasters technique. This is consistent with the findings of [Sinniah \(1982\)](#) and [Gasbarre et al. \(1996\)](#), and is an important observation because this source of variability can be reduced substantially by simply increasing the amount of faeces used per homogenised sample. Diminishing returns appear to be obtained beyond the use of approximately 10g of faeces, but assuming that the sample continues to be completely homogenised even larger samples than this can be used to further decrease the effective variability. Samples can also be amalgamated from several areas of a single faecal pile, reducing the effect of the sample variability which was found to be important in Animal 4. Based on these simple, low to zero cost recommendations alone, the clinical usefulness of a single FEC could be increased substantially. The variability in FEC from an individual over time should also be considered in clinical decision making, and where necessary repeated sampling over an animal on more than one day in order to reduce the impact of factors such as faecal consistency could be considered ([Gasbarre et al., 1996](#)). However, it appears that the largest source of variability between FEC of horses is the true variability between individuals, which is consistent with the observations in sheep ([Hunter and Quenouille, 1952](#)) and cattle ([Gasbarre et al., 1996](#)).

The variability introduced by the counting process was found to be very close to a Poisson distribution, with only a small extra amount of variability added by the mechanism of the McMasters technique. This is consistent with previously published work (Hunter and Quenouille, 1952; Presland et al., 2005), and indicates that a Poisson distribution describing the number of eggs observed has a mean very closely related to the number of eggs in the sample, although based on the findings of Rossanigo and Gruner (1991) it is likely that this represents a fixed proportion of the true number of eggs in the sample rather than all of the eggs. As a result, the majority of the observed variability in FEC between samples taken from the same animal is caused by Poisson variation and true variation in the egg content between the samples. Novel methods of detecting eggs in faeces such as the procedure advocated by Presland et al. (2005) may therefore provide little advantage over the McMasters procedure beyond the use of a larger sample of faeces to reduce replicate variability and counting more eggs to reduce the Poisson variability, both of which are easily achievable using the McMasters technique (Hunter and Quenouille, 1952). Homogenisation of larger samples, in combination with counting more eggs, is also a cheaper and quicker alternative to the examination of multiple samples per individual advocated by Gasbarre et al. (1996), and would give the same precision of estimates for the true mean FEC.

The experimental setup of the precision calculations was intended to closely mimic a FEC procedure using a McMasters technique, and could theoretically have been accomplished using the real technique given enough FEC samples of known differing mean FEC. Despite this, some assumptions regarding the real procedure were inevitable and are outlined below:

- Laboratory technique including the weighing of faeces and decanting of liquid introduces a source of variability with a standard deviation proportionate to the mean. This assumes that any loss of eggs occurring during the process of centrifuging or filtering, is a variable *proportion* of the number of eggs in the sample, and not a fixed number of eggs regardless of the number of eggs present.
- The technician analysing the samples does not make a mistake in counting or fail to report apparently unusually low or high individual FEC, i.e. there are neither type I or type II errors when counting eggs.
- The variability between samples within an individual is common for all individuals. This is a simplifying assumption, and based on the results presented in [Section 6.2.3](#) does not hold for real data. The effect may be expected to be random between animals, in which case it would not represent a source of bias, although it is also possible that the variability may be in some way related to the mean egg count of the individual.
- The distribution of between and within animal variability in mean egg count is adequately described by a gamma distribution. This is a distributional approximation, but given the small sample sizes and flexibility of a gamma distribution it is likely to be

reasonable. A lognormal distribution may be better justified and could have been used in place of the gamma, but would be expected to give similar results with values of cv less than one. Multi-modality in the distribution, if present, may not be adequately described by either of these distributions.

A single FEC is often used as a basis for making clinical decisions on treatment within equine clinical parasitology (Uhlinger, 1993), but consideration to the associated uncertainty of the observation is rarely, if ever, made. Performing prospective precision analysis calculations is not straightforward, but can be achieved using statistical software such as R, or even using the in-built distribution functions of Excel. Given the potential impact of clinical decisions regarding parasite treatment on pasture egg contamination, and the difficulty associated with generating appropriate 95% CI for the true mean FEC given a single estimate, it is important to consider the statistical issues of FEC in an individual animal. Although better estimates and easier methods of calculating 95% CI are provided by the distribution of data collected from groups of animals, prospective sample size calculations would be beneficial to evaluate the loss of information compared to reduction in cost associated with using pooled FEC samples, and could also be used when sample sizes are small so that the benefit resultant from taking repeat samples from within animals can be assessed. An arguably greater benefit would be achieved by application of prospective power calculations to the FECRT. One of the main conclusions made in Chapter 5 is that estimates of the true efficacy from typical FECRT data are subject to very poor confidence, which severely limits the usefulness of the technique. Use of power calculations to ensure that the proposed study design is likely to give useful inference on efficacy would represent a marked improvement of the utility of the method, albeit at the expense of extra laboratory costs associated with analysing replicate samples from within individuals and/or decreasing the minimum egg detection threshold.

CHAPTER 7

General discussion and conclusions

General discussion and conclusions

Over the last several decades, the leading method of analysis for parasitological data has changed from calculation of empirical summary statistics, through the use of basic generalised linear models (Wilson and Grenfell, 1997), to more complex likelihood based techniques such as maximum likelihood analysis (Wilson et al., 1996; Torgerson et al., 2005). In this thesis, analysis of such data using the next generation of computationally intensive statistical methods has been explored. Bayesian MCMC models representing ZIGP and ZILP distributions were developed and validated with simulated data, and were shown to produce superior results to analysis using maximum likelihood techniques. The inference on the variability parameter was particularly improved when using MCMC compared to the maximum likelihood method. For future studies concerned with the quantification of variability and/or zero-inflation in FEC and worm burden data, analysis of the data using computationally intensive statistical methods such as MCMC should therefore be considered essential.

The effects and validity of various distributional modelling assumptions have also been examined. Parasite data are almost ubiquitously assumed to follow a negative binomial distribution (May and Anderson, 1978; Wilson and Grenfell, 1997; Shaw et al., 1998). While several authors have justified the use of a Poisson distribution compounded to some underlying distribution describing the true variability between animals (Fisher, 1941; Crofton, 1962; Pacala and Dobson, 1988; Grenfell et al., 1995), the use of the gamma distribution to describe this underlying variability seems to be entirely based on the convenience of having a closed form solution for the negative binomial. Use of a lognormal distribution in place of this gamma distribution is more easily justifiable, and has previously been suggested (Elston et al., 2001). The gamma-Poisson and lognormal-Poisson distributions have a very similar shape with equal mean and cv of less than one, but can be distinguished with cv of greater than one and sufficient mean and sample size. Analysis of real equine and sheep datasets with high cv and sufficient mean and sample size would be expected to consistently produce a better fit to either the ZILP or ZIGP distributions, but this was not the case despite the comparative inference made using the two distributions being quite different for some parameters. There was also an indication of a zero-inflated component in the FEC and worm burden datasets examined, including in the equine data where the presence of true zero-inflation is difficult to justify biologically. This most likely indicates that the underlying variability in mean FEC and worm burden in infected animals does not follow either a gamma or lognor-

mal distribution. This has serious implications for the validity of analyses which make this assumption, particularly when using zero-inflation to model the shedding prevalence, and further examination of the issues outlined below should be considered a priority for future work.

The previously stated biological reasoning for worm burdens following a lognormal distribution depends on the multiplication of a series of distributions describing the variability in various host factors such as grazing habits, adversity of the intestinal environment to developing larvae, host immune responses and genetic fitness of the ingested larvae. If all of these distributions are independent, approximately continuous and uni-modal, the resultant distribution of worm burdens would be approximately lognormal. The possible flaws in these assumptions are as follows:

- It is possible that there are insufficient separate distributions for the combination to approximate a lognormal distribution. Given the complexity of the host-parasite interaction this seems unlikely.
- Any feedback effects of competition on worm development are not accounted for. This is likely to be incorrect because density dependent effects are thought to be important in the epidemiology of helminth parasites (Dietz, 1988; Churcher et al., 2005). It is also possible that heavily parasitised animals are immunocompromised by the presence of large numbers of worms, leading to a positive feedback effect on larval development.
- The combination of different parasite species may have an effect on the overall distribution of worm burdens for all species. This may have affected the sheep burdens where species of *Nematodirus* other than *battus* were present, albeit in lower numbers, and is very likely to have affected the equine data where several species of cyathostomins are likely to have been present.
- The assumption that all effects that contribute to the overall distribution of worm burdens are continuous may be invalid. While it is justifiable for effects with an environmental influence and/or multiple genetic influences to be approximately continuously distributed, heavily influential single genetic effects could introduce a bi-modality into the distribution. Each sub-group of the multi-modal distribution would still be expected to follow a lognormal distribution, as long as the remaining distributions were approximately continuous.
- The overall distribution of parasite numbers may be predominated by a few, or even a single, rate limiting step in the parasite development

The result of possible feedback and multiple species effects is likely to be complex (Dietz, 1988; Churcher et al., 2005), and the effect on the resultant distribution of worm burdens is

therefore difficult to estimate. If the distributions of different worm species were independent, then given sufficient numbers of different species the overall distribution would be expected to approximate a normal distribution. This is obviously not the case, suggesting that the distribution of one species is dependent either directly on the distribution of another species, or that each species has common dependencies such as the host immune response. If there are sufficient different species of worm that this could be considered to contribute an approximately continuous distribution, then worm species could be considered as another contributor to the resultant multiplicative distribution. With fewer different species the resultant distribution may not approximate to a lognormal, but as each animal would reasonably be expected to harbour more than one species of parasite, there is no reason to suspect that the resultant distribution would be multi-modal. It is conceivable that a negative feedback effect of the presence of developing larvae on the survival of adult worms could reduce the numbers of animals with high worm burdens, which would produce fewer high count observations than expected given the number of zero observations. This is a possible explanation for the excess zeros observed. A positive feedback effect would be expected to increase the number of high count observations relative to zero observations, which does not fit the observed data.

The more likely explanation for observing a multi-modal distribution is that one or more of the distributions contributing to the overall distribution of worm burdens does not approximate a continuous distribution. One possible cause of this could be a single genetic effect such as homozygosity or heterozygosity at a heavily influential genetic element such as the MHC. If heterozygous animals are considered to be more resistant to parasitism than homozygotes, the resultant overall distribution would be bimodal with the homozygotes contributing a distribution with larger mean and the heterozygotes contributing a distribution with smaller mean. If the mean of the heterozygotes was close enough to zero, then the resultant distribution would appear to be a lognormal-Poisson with greater weight at zero, which would approximate very well to a ZILP or ZIGP distribution. This effect potentially explains the inconsistency in the fit to the ZILP and ZIGP distributions; as the gamma is more flexible than the lognormal, the fit of the ZIGP to a multi-modal distribution with two reasonably high means might be expected to be superior. The ZILP distribution would produce a greater maximum likelihood when the vast majority of the data is described by either a single lognormal distribution, or a lognormal distribution with other distributions contributing nearly all zero observations. The implications of potential multi-modality are important for control of parasitism; both in terms of identifying animals that are likely to be shedding large numbers of eggs, and in identifying animals that are likely to be more resistant to clinical disease. These may not necessarily be the same animals if the genetic effects of susceptibility to clinical parasitism and predisposition to high egg shedding rate are not the same. If any source of multi-modality in the data is important, then analysis of data assuming any single continuous distribution may give misleading results.

This explanation is consistent with the observations in FEC and worm burden data, but is

quite speculative and further work involving intensive sampling to identify any multi-modality in the data would be a useful extension of the work presented here. As part of this further work, mathematical modelling could be usefully employed to study the effect on the observed distribution of worm burdens of parasite feedback effects, the combination of different species, and the number of potentially important single genetic effects. These models could also be extended by modelling the suspected influence of co-parasitism and host immune response on fecundity, variability in nematode fecundity over time and mixing patterns of faeces to explain the observed variability in egg concentration on a local scale within faeces. The observed frequency distribution could then be compared to the lognormal-Poisson, gamma-Poisson, ZILP and ZIGP to identify the distribution that provides the best approximation to the data. Another study incorporating replicate sampling from individual animals would be useful in order to validate the inference on within animal variability obtained here, or to extend the data to include other species. Funding has recently been obtained to perform such a study in donkeys.

While it is possible that some species of parasite in host groups with certain characteristics such as *N. battus* in lambs may truly follow a zero-inflated distribution, it is likely that the inference on zero-inflation obtained from parasites that are known to be almost 100% prevalent does not represent a group of animals that are truly uninfected. The use of a zero-inflated distribution in this situation is likely to be controversial. [Lord et al. \(2007\)](#) examined the use of the zero-inflated distribution in the field of motor vehicle crashes, and conclude that the statistical fit of the zero-inflated distribution was not sufficient to justify its use without a theoretical justification to model the data as two distinct groups. The same applies to analysis of FEC data in situations where parasite prevalence is known to be close to 100%. In these situations, where the lower 95% CI for zero-inflation does not include zero, it is possible to infer that the data more closely follow the zero-inflated mixture distribution than the uni-modal distribution. However, this does not mean that a zero-group truly exists. If the underlying cause of the apparent zero-inflation is a multi-modal distribution of animals, one of which is a positive distribution contributing the vast majority of the observed non-zero counts and the others contributing mostly zero observations, then the zero-inflated distribution is a simplification of the true multi-modal distribution and is therefore justified. Analysis of such data without the use of a zero-inflated component would result in inferring a reduced mean and increased variability for the high shedding group of animals. Fitting a mixture distribution with several positive modes may be preferable to using a zero-inflated distribution, but would likely prove unidentifiable with insufficient data. Alternatively, if the extra zeros are arising from some other mechanism, then the use of a zero-inflated model is not justified and may give misleading estimates for the true mean and variability in the data. Selection of the most appropriate model will ultimately depend on the simplifications and assumptions being made about the biological system, and the most useful inference will always be dependent on the desired purpose of the study.

The novel MCMC analysis method was shown to produce more reliable 95% confidence intervals when applied to FECRT data than non-parametric bootstrapping, which did not perform well even compared to the calculation of 95% CI based on the empirical mean and variance. Implementing MCMC presents some technical difficulties relative to calculation of empirical means (Gilks et al., 1998), but the analysis of FEC and FECRT data presents sufficient inherent difficulties that the extra effort associated with the more sophisticated method is justifiable. The current (Duncan et al., 2002), and forthcoming (pers. comm.), WAAVP guidelines for analysis of equine FECRT data do not adequately account for these difficulties, and based on the results presented in this thesis are likely to be statistically inadequate. In addition, the current practise of selecting animals to be included in the FECRT on the basis of a pre-treatment single FEC (Craven et al., 1998; Coles et al., 2006) is counter-productive as the effect is likely to improve the apparent efficacy of the drug unless explicitly modelled, and the presence of low or zero observations in the pre-treatment data poses no problem to analysis of the data using MCMC. Studying the change in variability associated with a FECRT could potentially provide useful information on emerging resistance that would be missed when examining the change in mean alone, but requires more intensive sampling than is currently performed. Use of the paired model described in Chapter 5 also provides a statistically valid way of incorporating several replicate samples from each available animal, which would increase the power of a FECRT substantially. While the analysis of ruminant datasets, with smaller cv , larger sample size and larger mean relative to typical equine datasets may be less affected by these difficulties, the improvement of inference on cv and zero-inflation afforded by MCMC would still be beneficial, especially in terms of the usefulness of the 95% CI produced. The comparatively poor performance of some of the statistical techniques examined is also an indication that validation of statistical assumptions using simulated data is an under-utilised resource. Such simulation studies are relatively simple to implement using statistical programming languages such as R. There is also a strong argument for using similar techniques to validate MCMC models where more than one possible syntactic variation is possible, and to analyse the sensitivity of a model to superficially similar prior distributions.

The priors used for cv in the model specifications presented were intended to be minimally informative within a biologically sensible range, but as would be expected had a demonstrable effect on inference of cv and zero-inflation when there was comparatively little information in the data. The prior on mean count also had a strong effect when the data contained few count observations, superficially seeming to lead to inappropriate inference for these datasets. The fact that truly uninformative priors do not exist could be used as an argument against Bayesian techniques, although subjective Bayesians would argue that there is no conceivable situation in which there is no available prior information, and it can also be argued that frequentist methods are no more objective in that conclusions depend on the perceived experimental design (Berger and Berry, 1988). Where prior information is easily incorporated, the ability to include this information is a major advantage of Bayesian methods

(Goodman, 2005), especially when information in the data is limited as it often is with FEC data. Specifying a prior for cv that limits inference to values that are likely to be observed in real data improves the inference on mean count and zero-inflation, and where possible using data available in similar groups of animals to inform these priors is a very useful way of reducing uncertainty for all parameters. However, it is important to understand that even ‘minimally informative’ priors may still dominate the posterior in some situations, and to always be sure that the prior distribution used is appropriate.

To date, prospective consideration of the precision associated with FEC and FECRT procedures has been severely under-utilised. The poor repeatability inherent in the FEC has previously been demonstrated (for example Mes, 2003; Presland et al., 2005; Eysker et al., 2008), but with little attempt to quantify the source of this variability with any statistical vigour. By demonstrating that a large amount of this variability in horses is as a result of local scale variability within faeces, a practical method of improving the repeatability has been demonstrated, and the beneficial effect of improving counting sensitivity has been quantified. The repeatability of the McMasters technique itself was found to be very good, suggesting that little is to be gained from the use of more modern procedures such as FECPAK (Presland et al., 2005), beyond the use of larger samples and counting the number of eggs in a larger proportion of the sample. Further, and statistically valid, evaluation of these newer procedures should also be performed before their use can be considered as equivalent to the McMasters procedure. While the development of these precision analysis methods involved the use of some advanced statistical methods, and such calculations are dependent on figures for within and between animal variability which are currently uncertain, consideration of the likely precision of a sampling procedure should be considered essential to efficient experimental design. While it may not be always be possible to incorporate more animals into the study, taking replicate samples from each animal and increasing the counting sensitivity of the test and size of faecal sample analysed is always possible. Such calculations can also be used to examine the cost benefit ratio of undertaking more laboratory work, so that the optimal balance of resource expenditure for experimental power can be found. When carrying out hypothesis driven testing such as a FECRT, prospective power calculations are essential to ensure that the study has an appropriate power before undertaking sampling. Post-hoc power calculations can also be used to help interpret a significance result, although consideration of appropriate confidence intervals is certainly superior (Hoenig and Heisey, 2001). The power calculation methods developed in this thesis have been made freely available online as part of the ‘bayescount’ package for R, and can be freely downloaded from the CRAN website at <http://cran.r-project.org/web/packages/bayescount>

Within this thesis, it has been demonstrated that FEC data consist of a complex distribution of observations incorporating variability derived from multiple sources, as might be expected given the known complexity of the underlying biological system. It has also been demonstrated that the most frequently used methods of analysis of such data are unable to

deal with this complex variability structure, leading to inaccurate inference and potentially incorrect clinical decisions in many cases. Computationally intensive parametric techniques such as MCMC provide a superior method of analysis for FEC and FECRT data, including the potential to improve the accuracy of results by incorporating repeated samples, and should be considered mandatory for analysis of data from small groups of animals or when attempting to quantify the variability in the data. Combining prospective power calculations and a complex MCMC model to analyse repeated samples from each individual also provides the opportunity to drastically improve the usefulness of a FECRT for a small group of animals. The WAAVP guidelines currently make no recommendations regarding power calculations whatsoever (Coles et al., 1992; Duncan et al., 2002), and consider only the mean and lower 95% confidence interval in interpreting results. Given the considerable resource expenditure associated with monitoring anthelmintic resistance, the lack of appropriate statistical guidelines for carrying out a FECRT is unacceptable. It is therefore essential that appropriate prospective power calculations and statistically correct calculation and interpretation of the significance of observed FEC reductions be performed for all future FECRT procedures. The implementation of such methods is not trivial, but given the widespread availability of computing power and quantitative expertise in similar disciplines, persistence with flawed statistical techniques in veterinary parasitology on the basis of convenience is difficult to justify. Considering the welfare implications of clinical parasitism and anthelmintic resistance, there is a clear professional responsibility to ensure that faecal worm egg count data are collected from optimally designed studies, and analysed using appropriate statistical methods. The work presented in this thesis sets the tone for an overhaul of current thinking, in order to allow these ideals to be better incorporated within veterinary parasitology.

BIBLIOGRAPHY

Bibliography

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In B N Petrov & F Csaki (Eds), 2nd International Symposium on Information Theory (pp 267-281) Budapest: Akademiai Kiado.* **9**, **51**
- Alfaro, M. E., Zoller, S. and Lutzoni, F. (2003). Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evol*, **20**(2): 255–266. **139**
- Alston, C., Kuhnert, P., Choy, L. S., Mcvinish, R. and Mengersen, K. (2005). Bayesian Model Comparison: Review and Discussion. **11**
- Anderson, R. and May, R. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications. **4**, **7**
- Anderson, R. M. and May, R. M. (1982). Coevolution of hosts and parasites. *Parasitology*, **85** (Pt **2**): 411–26. **12**
- Andrieu, C., Djuric, P. M. and Doucet, A. (2001). Model selection by MCMC computation. *Signal Processing*, **81**(1): 19–37. **11**
- Anscombe, F. J. (1950). Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. *Biometrika*, **37**(3/4): 358–382. **5**
- Arantes, A., Carvalho, E. d. S., Medeiros, E. A. S., Farhat, C. K. and Mantese, O. C. (2003). [Use of statistical process control charts in the epidemiological surveillance of nosocomial infections]. *Rev Saude Publica*, **37**(6): 768–774. **154**
- Armour, J. and Coop, R. (1991). *Pathogenesis and control of gastrointestinal helminthiasis. In: W.B. Martin and I.D. Aitken, Editors, Diseases of Sheep*. Blackwell Scientific Publications, Oxford. **86**
- Armour, J., Jarrett, W. and Jennings, F. (1966). Experimental *Ostertagia circumcincta* infections in sheep: Development and pathogenesis of a single infection. *American Journal of Veterinary Research*, **27**: 1267–1278. **89**
- Barger, I. A. (1985). The statistical distribution of trichostrongylid nematodes in grazing lambs. *Int J Parasitol*, **15**(6): 645–649. **2**, **4**, **5**, **50**, **86**, **111**
- Barnes, E. H. and Dobson, R. J. (1990). Population dynamics of *Trichostrongylus colubriformis* in sheep: computer model to simulate grazing systems and the evolution of anthelmintic resistance. *Int J Parasitol*, **20**(7): 823–31. **17**
- Barrett, E. J., Farlam, J. and Proudman, C. J. (2004). Field trial of the efficacy of a combination of ivermectin and praziquantel in horses infected with roundworms and tapeworms. *Vet Rec*, **154**(11): 323–325. **88**

- Baudena, M. A., Chapman, M. R., French, D. D. and Klei, T. R. (2000). Seasonal development and survival of equine cyathostome larvae on pasture in south Louisiana. *Vet Parasitol*, **88**(1-2): 51–60. [19](#), [115](#)
- Bauer, C., Merkt, J. C., Janke-Grimm, G. and Burger, H. J. (1986). Prevalence and control of benzimidazole-resistant small strongyles on German thoroughbred studs. *Vet Parasitol*, **21**(3): 189–203. [115](#), [116](#), [120](#), [122](#), [154](#)
- Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, **53**: 370–418. [10](#)
- Bennett, M. F. (1990). Is there a morning-evening difference in egg production of bloodworms, strongyle parasites of equines? *Prog Clin Biol Res*, **341B**: 729–733. [19](#), [119](#), [172](#)
- Berger, J., Ghosh, J. and Mukhopadhyay, N. (2003). Approximations and consistency of Bayes factors as model dimension grows. *Journal of Statistical Planning and Inference*, **112**(1-2): 241–258. [11](#)
- Berger, J., Moreno, E., Pericchi, L., Bayarri, M., Bernardo, J., Cano, J., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., Dasgupta, A., Gustafson, P., Wasserman, L., Kadane, J., Srinivasan, C., Lavine, M., O'Hagan, A., Polasek, W., Robert, C., Goutis, C., Ruggeri, F., Salinetti, G. and Sivaganesan, S. (1994). An overview of robust Bayesian analysis. *TEST*, **3**(1): 5–124. [17](#)
- Berger, J. O. and Berry, D. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, **76**: 159–165. [17](#), [195](#)
- Besier, B. (2007). New anthelmintics for livestock: the time is right. *Trends Parasitol*, **23**(1): 21–4. [117](#)
- Bishop, S., Jackson, F., Coop, R. and Stear, M. (2004). Genetic parameters for resistance to nematode infections in Texel lambs and their utility in breeding programmes. *Animal Science*, **78**: 185–194. [87](#)
- Bisset, S. A., Vlassof, A., Morris, C. A., Sourthey, B. R., Baker, R. L. and Parker, A. G. H. (1992). Heritability of and genetic correlations among faecal egg counts and productivity traits in Romney sheep. *New Zeal J Agric Res*, **35**: 51–58. [19](#)
- Black, M. A. and Craig, B. A. (2002). Estimating disease prevalence in the absence of a gold standard. *Stat Med*, **21**(18): 2653–69. [11](#)
- Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., Vanfleteren, J. R., Mackey, L. Y., Dorris, M., Frisse, L. M., Vida, J. T. and Thomas, W. K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature*, **392**(6671): 71–75. [2](#)
- Bliss, C. and Fisher, R. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, **9**(2): 176–200. [4](#), [14](#)
- Boag, B. and Thomas, R. J. (1975). The population dynamics of nematode parasites of sheep in northern England. *Res Vet Sci*, **19**(3): 293–5. [4](#), [17](#)
- Boersema, J. H., Borgsteede, F. H., Eysker, M. and Saedt, I. (1995). The reappearance of strongyle eggs in faeces of horses treated with pyrantel embonate. *Vet Q*, **17**(1): 18–20. [154](#)
- Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L. and Kirchner, U. (2000). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal Of The Royal Statistical Society Series A*, **160**(1): 121–122. [8](#)
- Box, G. E. P. (1976). Science and statistics. *J Am Statist Ass*, **71**: 791–799. [12](#)
- Brame, R. S. and Groer, P. G. (2002). Bayesian analysis of overdispersed chromosome aberration data with the negative binomial model. *Radiat Prot Dosimetry*, **102**(2): 115–9. [12](#)

- Brooks, S. and Roberts, G. (1998). Assessing convergence of Markov chain Monte Carlo algorithms. *Statistics and Computing*, **8**: 319–33. [16](#), [52](#)
- Cabaret, J. and Berrag, B. (2004). Faecal egg count reduction test for assessing anthelmintic efficacy: average versus individually based estimations. *Vet Parasitol*, **121**(1-2): 105–113. [118](#), [120](#)
- Cabaret, J., Gasnier, N. and Jacquiet, P. (1998). Faecal egg counts are representative of digestive-tract strongyle worm burdens in sheep and goats. *Parasite*, **5**(2): 137–42. [7](#)
- Chandler, K. and Love, S. (2002). Patterns of equine faecal egg counts following spring dosing with either fenbendazole or moxidectin. *Veterinary Record*, **151**(9): 269–270. [5](#), [117](#)
- Churcher, T. S., Ferguson, N. M. and Basanez, M. G. (2005). Density dependence and overdispersion in the transmission of helminth parasites. *Parasitology*, **131**(Pt 1): 121–132. [5](#), [192](#)
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*. Lawrence Erlbaum, Hillsdale, NJ, 2nd ed edition. [18](#)
- Coles, G. C. (2002). Sustainable use of anthelmintics in grazing animals. *Vet Rec*, **151**(6): 165–169. [116](#), [117](#), [119](#)
- Coles, G. C., Bauer, C., Borgsteede, F. H., Geerts, S., Klei, T. R., Taylor, M. A. and Waller, P. J. (1992). World Association for the Advancement of Veterinary Parasitology (W.A.A.V.P.) methods for the detection of anthelmintic resistance in nematodes of veterinary importance. *Vet Parasitol*, **44**(1-2): 35–44. [13](#), [117](#), [118](#), [119](#), [120](#), [121](#), [122](#), [123](#), [125](#), [128](#), [152](#), [197](#)
- Coles, G. C., Jackson, F., Pomroy, W. E., Prichard, R. K., von Samson-Himmelstjerna, G., Silvestre, A., Taylor, M. A. and Vercruysse, J. (2006). The detection of anthelmintic resistance in nematodes of veterinary importance. *Vet Parasitol*, **136**(3-4): 167–185. [114](#), [120](#), [122](#), [123](#), [139](#), [152](#), [155](#), [157](#), [195](#)
- Comer, K. C., Hillyer, M. H. and Coles, G. C. (2006). Anthelmintic use and resistance on thoroughbred training yards in the UK. *Vet Rec*, **158**(17): 596–598. [116](#), [117](#)
- Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London. [9](#), [68](#)
- Craven, J., Bjorn, H., Barnes, E. H., Henriksen, S. A. and Nansen, P. (1999). A comparison of in vitro tests and a faecal egg count reduction test in detecting anthelmintic resistance in horse strongyles. *Vet Parasitol*, **85**(1): 49–59. [118](#)
- Craven, J., Bjorn, H., Henriksen, S. A., Nansen, P., Larsen, M. and Lendal, S. (1998). Survey of anthelmintic resistance on Danish horse farms, using 5 different methods of calculating faecal egg count reduction. *Equine Vet J*, **30**(4): 289–293. [114](#), [117](#), [120](#), [122](#), [123](#), [139](#), [152](#), [195](#)
- Cringoli, G. (2006). FLOTAC, a novel apparatus for a multivalent faecal egg count technique. *Parasitologia*, **48**(3): 381–4. [19](#)
- Crofton, H. (1962). A quantitative approach to parasitism. *Parasitology*, pp. 179–193. [5](#), [6](#), [22](#), [52](#), [191](#)
- CSIRO and Australian Agricultural Council (1989). *Anthelmintic resistance : report of the working party for the Animal Health Committee of the Standing Committee on Agriculture*. CSIRO for the Standing Committee on Agriculture, Melbourne :. [119](#), [120](#), [121](#)
- Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A. and Winka, K. (2003). Comparing bootstrap and posterior probability values in the four-taxon case. *Syst Biol*, **52**(4): 477–487. [139](#)
- Dargatz, D. A., Traub-Dargatz, J. L. and Sangster, N. C. (2000). Antimicrobial and anthelmintic resistance. *Vet Clin North Am Equine Pract*, **16**(3): 515–536. [xvi](#), [2](#), [115](#), [116](#), [117](#), [118](#)

- Denwood, M. (2008). *runjags: Run Bayesian MCMC Models in the BUGS syntax from Within R*. R package version 0.9.2. [16](#), [48](#)
- Dietz, K. (1988). Density-dependence in parasite transmission dynamics. *Parasitol Today*, **4**(4): 91–7. [192](#)
- Doeschl-Wilson, A. B., Vagenas, D., Kyriazakis, I. and Bishop, S. C. (2008). Exploring the assumptions underlying genetic variation in host nematode resistance (Open Access publication). *Genet Sel Evol*, **40**(3): 241–64. [18](#)
- Döpfer, D., Kerssens, C. M., Meijer, Y. G. M., Boersema, J. H. and Eysker, M. (2004). Shedding consistency of strongyle-type eggs in Dutch boarding horses. *Vet Parasitol*, **124**(3-4): 249–258. [2](#), [5](#), [22](#), [115](#), [119](#)
- Duncan, J. (1975). Immunity to *Strongylus vulgaris* in the horse. *Equine Veterinary Journal*, **7**: 192–197. [88](#)
- Duncan, J. L., Abbott, E. M., Arundel, J. H., Eysker, M., Klei, T. R., Krecek, R. C., Lyons, E. T., Reinemeyer, C. and Slocombe, J. O. D. (2002). World association for the advancement of veterinary parasitology (WAAVP): second edition of guidelines for evaluating the efficacy of equine anthelmintics. *Vet Parasitol*, **103**(1-2): 1–18. [4](#), [18](#), [88](#), [123](#), [195](#), [197](#)
- Duncan, J. L. and Love, S. (1991). Preliminary observations on an alternative strategy for the control of horse strongyles. *Equine Vet J*, **23**(3): 226–228. [115](#)
- Duncan, P., Mackie, F., Mackay, K. and Quinn, T. (2005). Use of Statistical Process Control to Support Improved Care for Patients With Acute Myocardial Infarction Eligible for Thrombolytic Treatment: Experience From 2 Hospitals in England During 2002-2003. *Crit Pathw Cardiol*, **4**(1): 21–25. [154](#)
- Dunn, A. and Keymer, A. (1986). Factors affecting the reliability of the McMaster technique. *J Helminthol*, **60**(4): 260–2. [19](#)
- Edwards, A. (1992). *Likelihood Expanded edition*. The Johns Hopkins University Press, Baltimore. [14](#)
- Efron, B. (1979). Bootstrapping methods: Another look at the jackknife. *Ann Stat*, **7**: 1–26. [17](#), [126](#)
- Elliott, J. (1977). *Some methods for the statistical analysis of samples of benthic invertebrates*. Freshwater Biological Association, Ambleside., 2nd edition. [4](#), [5](#), [6](#), [22](#), [58](#)
- Elston, D. A., Moss, R., Boulinier, T., Arrowsmith, C. and Lambin, X. (2001). Analysis of aggregation, a worked example: numbers of ticks on red grouse chicks. *Parasitology*, **122**(Pt 5): 563–569. [5](#), [6](#), [191](#)
- Evans, M., Hastings, N. and Peacock, B. (2000). *Statistical Distributions*. Wiley, New York, third edition edition. [5](#), [25](#), [39](#), [178](#)
- Eysker, M., Bakker, J., van den Berg, M., van Doorn, D. C. K. and Ploeger, H. W. (2008). The use of age-clustered pooled faecal samples for monitoring worm control in horses. *Vet Parasitol*, **151**(2-4): 249–255. [19](#), [196](#)
- Eysker, M. and Vanmeurs, G. (1982). Seasonal pattern in the strongyle egg output of adult dairy-cows in the Netherlands. *Research In Veterinary Science*, **33**(2): 208–211. [19](#)
- Famoye, F. (1993). Restricted generalized Poisson regression-model. *Communications In Statistics - Theory and Methods*, **22**(5): 1335–1354. [7](#)
- Fisher, M. A., Jacobs, D. E., Grimshaw, W. T. and Gibbons, L. M. (1992). Prevalence of benzimidazole-resistance in equine cyathostome populations in south east England. *Vet Rec*, **130**(15): 315–318. [119](#), [120](#)

- Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics*, **11**: 182–187. [5](#), [191](#)
- Fisher, R. A. (1954). *Statistical methods for research workers*, volume no. 5. Oliver and Boyd, Edinburgh, 12th ed., rev edition. [157](#)
- Fortin, M. and DeBlois, J. (2007). Modeling Tree Recruitment with Zero-Inflated Models: The Example of Hardwood Stands in Southern Quebec, Canada. *Forest Science*, **53**: 529–539(11). [7](#), [8](#)
- Freckman, D. W. and Virginia, R. A. (1989). Plant-Feeding Nematodes in Deep-Rooting Desert Ecosystems. *Ecology*, **70**(6): 1665–1678. [2](#)
- Fulford, A. J. (1994). Dispersion and bias: can we trust geometric means? *Parasitol Today*, **10**(11): 446–8. [13](#)
- Gaba, S., Ginot, V. and Cabaret, J. (2005). Modelling macroparasite aggregation using a nematode-sheep system: the Weibull distribution as an alternative to the negative binomial distribution? *Parasitology*, **131**(Pt 3): 393–401. [7](#)
- Galvani, A. P. (2003). Immunity, antigenic heterogeneity, and aggregation of helminth parasites. *J Parasitol*, **89**(2): 232–241. [5](#)
- Gasbarre, L. C., Leighton, E. A. and Bryant, D. (1996). Reliability of a single fecal egg per gram determination as a measure of individual and herd values for trichostrongylid nematodes of cattle. *American Journal of Veterinary Research*, **57**(2): 168–171. [19](#), [187](#), [188](#)
- Gelman, A., Meng, X. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**(4): 733–760. [12](#)
- Gelman, A. and Rubin, D. (1992a). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, **7**: 457–511. [16](#), [129](#)
- Gelman, A. and Rubin, D. B. (1992b). *A single series from the Gibbs sampler provides a false sense of security*. In Bayesian Statistics 4, Oxford University Press, New York. [16](#)
- Geman, S. and Geman, D. (1990). *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. [15](#)
- Gettinby, G. (1989). Computational veterinary parasitology with an application to chemical resistance. *Vet Parasitol*, **32**(1): 57–72. [17](#)
- Gettinby, G., Soutar, A., Armour, J. and Evans, P. (1989). Anthelmintic resistance and the control of ovine ostertagiasis: a drug action model for genetic selection. *Int J Parasitol*, **19**(4): 369–76. [17](#)
- Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**(6): 1317–1339. [15](#)
- Gibson, G. J., Otten, W., Filipe, J. A. N., Cook, A., Marion, G. and Gilligan, C. A. (2006). Bayesian estimation for percolation models of disease spread in plant populations. *Statistics and Computing*, **16**(4): 391–402. [12](#)
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1998). *Markov chain Monte Carlo in practice*. Chapman and Hall, Boca Raton, Fla. [15](#), [16](#), [46](#), [82](#), [195](#)
- Gill, J. L., Ericsson, G. F. and Helland, I. S. (1986). Precision of assessing anthelmintic efficacy. *Biometrics*, **42**(4): 981–987. [19](#), [176](#)
- Goodman, S. N. (2005). Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials*, **2**(4): 282–290. [11](#), [16](#), [46](#), [52](#), [196](#)
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4): 711–732. [11](#)

- Gregory, R. D. and Woolhouse, M. E. (1993). Quantification of parasite aggregation: a simulation study. *Acta Trop*, **54**(2): 131–9. [13](#)
- Grenfell, B. T., Wilson, K., Isham, V. S., Boyd, H. E. and Dietz, K. (1995). Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology*, **111** **Suppl**: S135–51. [6](#), [52](#), [191](#)
- Harvey, S. C., Paterson, S. and Viney, M. E. (1999). Heterogeneity in the distribution of *Strongyloides ratti* infective stages among the faecal pellets of rats. *Parasitology*, **119** (Pt 2): 227–235. [5](#)
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1): 97–109. [15](#)
- Herbert, J. and Isham, V. (2000). Stochastic host-parasite interaction models. *J Math Biol*, **40**(4): 343–371. [4](#), [5](#)
- Herd, R. (1990). The changing world of worms: the rise of the cyathostomes and the decline of *S. vulgaris*. *Compend Cont Educ Pract Vet*, **12**: 732–736. [88](#)
- Herd, R. P. (1986). Epidemiology and control of equine strongylosis at Newmarket. *Equine Vet J*, **18**(6): 447–452. [19](#), [88](#), [115](#), [116](#)
- Herd, R. P., Willardson, K. L. and Gabel, A. A. (1985). Epidemiological approach to the control of horse strongyles. *Equine Vet J*, **17**(3): 202–207. [115](#), [119](#)
- Hoening, J. M. and Heisey, D. M. (2001). The Abuse of Power. *The American Statistician*, **55**(1): 19–24. [176](#), [196](#)
- Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, **14**(4): 382–401. [11](#)
- Hoste, H., Le Frileux, Y., Goudeau, C., Chartier, C., Pors, I., Broqua, C. and Bergeaud, J. P. (2002). Distribution and repeatability of nematode faecal egg counts in dairy goats: a farm survey and implications for worm control. *Res Vet Sci*, **72**(3): 211–215. [115](#), [157](#), [187](#)
- Hoste, H., Le Frileux, Y. and Pommaret, A. (2001). Distribution and repeatability of faecal egg counts and blood parameters in dairy goats naturally infected with gastrointestinal nematodes. *Res Vet Sci*, **70**(1): 57–60. [5](#)
- Hunter, G. and Quenouille, M. (1952). A statistical examination of the worm egg count sampling technique for sheep. *Journal of Helminthology*, **4**: 157–170. [19](#), [187](#), [188](#)
- Ihler, C. F. (1995). A field survey on anthelmintic resistance in equine small strongyles in Norway. *Acta Vet Scand*, **36**(1): 135–143. [5](#), [117](#), [119](#), [121](#)
- Ihler, C. F. and Bjorn, H. (1996). Use of two in vitro methods for the detection of benzimidazole resistance in equine small strongyles (*Cyathostoma* spp.). *Vet Parasitol*, **65**(1-2): 117–125. [117](#)
- Jefferys, W. H. and Berger, J. O. (1991). Sharpening Occam’s Razor on a Bayesian strop. In *Bulletin of the American Astronomical Society*, volume 23 of *Bulletin of the American Astronomical Society*. [11](#)
- Jeffreys, H. (1961). *Theory of probability*. Clarendon Press, Oxford, 3rd edition edition. [17](#)
- Jell, K., Barnes, T., Seddon, J. M., Coleman, G. T., Morton, J. and Kotze, A. (2008). The Use of Zero Inflated Modelling to Identify High and Low Gastrointestinal Nematode Egg Shedders in Groups of Horses. In *Proc. Aust. Equine Sc. Symp.*, volume 2, p. 51. [8](#)
- Joe, H. and Zhu, R. (2005). Generalized Poisson Distribution: the Property of Mixture of Poisson and Comparison with Negative Binomial Distribution. *Biometrical Journal*, **47**(2): 219 – 229. [7](#), [82](#)

- Johansen, M. V. (1989). An evaluation of techniques used for the detection of anthelmintic resistance in nematode parasites of domestic livestock. *Vet Res Commun*, **13**(6): 455–466. [119](#)
- Johnson, J. B. and Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, **19**(2): 101 – 108. [9](#), [51](#), [83](#)
- Jovani, R. and Tella, J. L. (2006). Parasite prevalence and sample size: misconceptions and solutions. *Trends Parasitol*, **22**(5): 214–8. [8](#)
- Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S. S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F., Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A. and Mäser, P. (2008). A new class of anthelmintics effective against drug-resistant nematodes. *Nature*, **452**(7184): 176–80. [117](#)
- Kao, R. R., Leathwick, D. M., Roberts, M. G. and Sutherland, I. A. (2000). Nematode parasites of sheep: a survey of epidemiological parameters and their application in a simple model. *Parasitology*, **121** (Pt 1): 85–103. [4](#), [17](#)
- Kaplan, R. M. (2002). Anthelmintic resistance in nematodes of horses. *Vet Res*, **33**(5): 491–507. [114](#), [117](#), [120](#), [122](#), [126](#), [154](#), [155](#)
- Kaplan, R. M. (2004). Drug resistance in nematodes of veterinary importance: a status report. *Trends Parasitol*, **20**(10): 477–81. [117](#)
- Kaplan, R. M., Klei, T. R., Lyons, E. T., Lester, G., Courtney, C. H., French, D. D., Tolliver, S. C., Vidyashankar, A. N. and Zhao, Y. (2004). Prevalence of anthelmintic resistant cyathostomes on horse farms. *J Am Vet Med Assoc*, **225**(6): 903–910. [117](#)
- Kass, R., Carlin, B., Gelman, A. and Neal, R. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *American Statistician*, **52**(2): 93–100. [29](#)
- Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**(430): 773–795. [11](#)
- Keymer, A. E. and Hiorns, R. W. (1986). Faecal egg counts and nematode fecundity: *Heligmosomoides polygyrus* and laboratory mice. *Parasitology*, **93** (Pt 1): 189–203. [7](#), [19](#)
- Klei, T. R. and Chapman, M. R. (1999). Immunity in equine cyathostome infections. *Vet Parasitol*, **85**(2-3): 123–133. [115](#)
- Klei, T. R. and Torbert, B. J. (1980). Efficacy of ivermectin (22,23-dihydroavermectin B1) against gastrointestinal parasites in ponies. *Am J Vet Res*, **41**(11): 1747–1750. [115](#)
- Konigova, A., Varady, M. and Corba, J. (2003). Comparison of in vitro methods and faecal egg count reduction test for the detection of benzimidazole resistance in small strongyles of horses. *Vet Res Commun*, **27**(4): 281–288. [118](#)
- Kuhnert, P., Martin, T., Mengersen, K. and Possingham, H. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics*, **16**(7): 717–747. [11](#)
- Kuzmina, T. A. and Kharchenko, V. O. (2008). Anthelmintic resistance in cyathostomins of brood horses in Ukraine and influence of anthelmintic treatments on strongylid community structure. *Vet Parasitol*, **154**(3-4): 277–288. [102](#), [111](#)
- Langrová, I. (1998). Seasonal prevalence and intensity of faecal helminth egg (larval) output in various categories of herds of horses during two grazing seasons. *Helminthologia*, **35**(1): 43–50. [19](#), [115](#)
- Lendal, S., Larsen, M. M., Bjorn, H., Craven, J., Chriel, M. and Olsen, S. N. (1998). A questionnaire survey on nematode control practices on horse farms in Denmark and the existence of risk factors for the development of anthelmintic resistance. *Vet Parasitol*, **78**(1): 49–63. [115](#)

- Levine, N. D., Mehra, K. N., Clark, D. T. and Aves, I. J. (1960). A comparison of nematode egg counting techniques for cattle and sheep feces. *Am J Vet Res*, **21**: 511–5. [19](#)
- Lichtenfels, J. (1975). Helminths of domestic equids. Illustrated keys to genera and species with emphasis on North American farms. *Proceedings of the Helminthological Society of Washington (special issue)*, **42**: 1–92. [87](#), [88](#)
- Limpert, E., Stahel, W. and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience*, **51**(5): 341–352. [25](#)
- Lind, E. O. and Christensson, D. (2009). Anthelmintic efficacy on *Parascaris equorum* in foals on Swedish studs. *Acta Vet Scand*, **51**: 45. [117](#)
- Lind, E. O., Kuzmina, T., Uggla, A., Waller, P. J. and Høglund, J. (2007). A field study on the effect of some anthelmintics on cyathostomins of horses in Sweden. *Vet Res Commun*, **31**(1): 53–65. [116](#), [117](#)
- Lloyd, S. (1998). Is anthelmintic resistance inevitable: back to basics? *Equine Vet J*, **30**(4): 280–283. [117](#)
- Lloyd, S. and Martin, S. A. (2006). Controlled trial on the effects of radionic healing and anthelmintics on faecal egg counts in horses. *Vet Rec*, **158**(21): 734–737. [116](#)
- Lord, D., Washington, S. and Ivan, J. N. (2007). Further notes on the application of zero-inflated models in highway safety. *Accid Anal Prev*, **39**(1): 53–7. [12](#), [194](#)
- Lord, D., Washington, S. P. and Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accid Anal Prev*, **37**(1): 35–46. [8](#), [59](#)
- Louis, T. A. (2005). Introduction to Bayesian methods II: fundamental concepts. *Clin Trials*, **2**(4): 291–294. [11](#)
- Love, S. (1992). Parasite-associated equine diarrhoea. *Compend Contin Educ Pract Vet*, **14**(5): 642–649. [88](#)
- Love, S. (1995). Recognising disease associated with strongyles in horses. *Compend Contin Educ Pract Vet*, **17**(14): 564–567. [2](#), [4](#), [87](#), [88](#)
- Love, S. and Duncan, J. L. (1992). The development of naturally acquired cyathostome infection in ponies. *Vet Parasitol*, **44**(1-2): 127–142. [88](#)
- Love, S., Murphy, D. and Mellor, D. (1999). Pathogenicity of cyathostome infection. *Vet Parasitol*, **85**(2-3): 113–121. [88](#)
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, **10**(4): 325–337. [15](#), [23](#), [43](#), [52](#)
- Lyons, E., Drudge, J. and Tolliver, S. (1974). Critical tests of three salts of pyrantel against internal parasites of the horse. *American Journal of Veterinary Research*, **35**: 1515–1522. [115](#), [117](#)
- MAFF (1986). *Manual of Veterinary Parasitological Laboratory Techniques*. MAFF, London, Her Majesty's Stationery Office. [18](#)
- Maingi, N., Bjorn, H., Thamsborg, S. M., Bøgh, H. O. and Nansen, P. (1996). A survey of anthelmintic resistance in nematode parasites of goats in Denmark. *Vet Parasitol*, **66**(1-2): 53–66. [120](#)
- Malan, F. S., Reinecke, R. K. and Scialdo, R. C. (1981). Anthelmintic efficacy fenbendazole paste in equines. *J S Afr Vet Assoc*, **52**(2): 127–130. [115](#)

- Marin, J., Mengersen, K. and Robert, C. (2005). *Bayesian modelling and inference on mixtures of distributions*, volume 25. In Handbook of Statistics, Elsevier-Sciences. 8
- Martin, P. J., Anderson, N. and Jarrett, R. G. (1989). Detecting benzimidazole resistance with faecal egg count reduction tests and in vitro assays. *Aust Vet J*, **66**(8): 236–240. 118, 119, 120, 122, 123, 154
- Martin, T., Wintle, B., Rhodes, J., Kuhnert, P., Field, S., Low-Choy, S., Tyre, A. and Possingham, H. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8**(11): 1235–1246. 8, 23
- Martin-Downum, K., Yazwinski, T., Tucker, C., Fincher, M., Ralph, J. and Hamilton, J. (2001). Cyathostome fecal egg count trends in horses treated with moxidectin, ivermectin or fenbendazole. *Vet Parasitol*, **101**(1): 75–79. 119, 122
- Mason, P. C., Hosking, B. C., Nottingham, R. M., Cole, D. J. W., Seewald, W., McKay, C. H., Griffiths, T. M., Kaye-Smith, B. G. and Chamberlain, B. (2009). A large-scale clinical field study to evaluate the efficacy and safety of an oral formulation of the amino-acetonitrile derivative (AAD), monepantel, in sheep in New Zealand. *N Z Vet J*, **57**(1): 3–9. 117
- Matthews, J. B. (2008). An update on cyathostomins: Anthelmintic resistance and worm control. *Equine Veterinary Education*, **20**(10): 552–560. 116
- May, R. M. and Anderson, R. M. (1978). Regulation and stability of host-parasite population interactions II. Destabilising processes. *Journal of Animal Ecology*, **47**: 248–267. 5, 22, 58, 84, 191
- May, R. M. and Anderson, R. M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. *Proc R Soc Lond B Biol Sci*, **219**(1216): 281–313. 4, 17
- McBeath, D. G., Best, J. M., Preston, N. K. and Duncan, J. L. (1978). Studies on the faecal egg output of horses after treatment with fenbendazole. *Equine Vet J*, **10**(1): 5–8. 115
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. CRC Press LLC, 2000 N.W. Corporate Blvd, Boca Raton, Florida, 33431., 2nd edition. 4, 13, 14, 50, 52
- McKenna, P. B. (1981). The diagnosis value and interpretation of faecal egg counts in sheep. *N Z Vet J*, **29**(8): 129–132. 119, 122, 157, 187
- McKenna, P. B. (1990). The detection of anthelmintic resistance by the faecal egg count reduction test: an examination of some of the factors affecting performance and interpretation. *N Z Vet J*, **38**(4): 142–147. 122
- McKenna, P. B. (2006a). A comparison of faecal egg count reduction test procedures. *N Z Vet J*, **54**(4): 202–203. 118, 122
- McKenna, P. B. (2006b). Further comparison of faecal egg count reduction test procedures: sensitivity and specificity. *N Z Vet J*, **54**(6): 365–366. 122, 152
- Meng, X.-L. (1994). Posterior Predictive p-Values. *The Annals of Statistics*, **22**(3): 1142–1160. 12
- Mercier, P., Chick, B., Alves-Branco, F. and White, C. R. (2001). Comparative efficacy, persistent effect, and treatment intervals of anthelmintic pastes in naturally infected horses. *Vet Parasitol*, **99**(1): 29–39. 4, 88, 114, 115, 117, 120
- Mes, T. H. M. (2003). Technical variability and required sample size of helminth egg isolation procedures. *Vet Parasitol*, **115**(4): 311–20. 19, 157, 196
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, **21**(6): 1087–1092. 15

- Miaou, S. and Lord, D. (2003). Modeling traffic crash flow relationships for intersections - Dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Statistical Methods and Modeling and Safety Data, Analysis, and Evaluation*, **1840**: 31–40. [9](#), [12](#), [51](#)
- Millar, R. B. (2009). Comparison of Hierarchical Bayesian Models for Overdispersed Count Data using DIC and Bayes' Factors. *Biometrics*. [9](#), [11](#), [83](#)
- Miller, C. M., Waghorn, T. S., Leathwick, D. M. and Gilmour, M. L. (2006). How repeatable is a faecal egg count reduction test? *N Z Vet J*, **54**(6): 323–328. [114](#), [119](#), [120](#), [153](#), [155](#), [157](#), [187](#)
- Miller, W. (1953). The general problem of parasitic infestation in horses. *Vet Record*, **65**(14): 213–217. [119](#)
- Mines, J. J. (1977). Modifications of the McMaster worm egg counting method. *Aust Vet J*, **53**(7): 342–3. [4](#)
- Molento, M. B. (2009). Parasite control in the age of drug resistance and changing agricultural practices. *Vet Parasitol*, **163**(3): 229–34. [2](#)
- Mooney, C. Z. and Duval, R. D. (1993). *Bootstrapping: a nonparametric approach to statistical inference*. Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320. [17](#), [126](#)
- Morgan, E. R., Cavill, L., Curry, G. E., Wood, R. M. and Mitchell, E. S. E. (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep. *Vet Parasitol*, **131**(1-2): 79–87. [5](#), [22](#), [123](#), [187](#)
- Morrison, D. A. (2004). Technical variability and required sample size of helminth egg isolation procedures: revisited. *Parasitol Res*, **94**(5): 361–366. [19](#), [157](#)
- Morton, A. P., Whitby, M., McLaws, M. L., Dobson, A., McElwain, S., Looke, D., Stackelroth, J. and Sartor, A. (2001). The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *J Qual Clin Pract*, **21**(4): 112–117. [154](#)
- Mwalili, S. (). zicounts: Counts data models: zero-inflation as well as interval censored. *R package version 114*. [53](#)
- Neyman, J. and Pearson, E. S. (1928). On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I. *Biometrika*, **20A**(1/2): 175–240. [9](#)
- Nielsen, M. K., Haaning, N. and Olsen, S. N. (2006a). Strongyle egg shedding consistency in horses on farms using selective therapy in Denmark. *Vet Parasitol*, **135**(3-4): 333–335. [2](#), [115](#), [117](#), [126](#)
- Nielsen, M. K., Monrad, J. and Olsen, S. N. (2006b). Prescription-only anthelmintics—a questionnaire survey of strategies for surveillance and control of equine strongyles in Denmark. *Vet Parasitol*, **135**(1): 47–55. [88](#)
- Nødtvedt, A., Dohoo, I., Sanchez, J., Conboy, G., DesCôteaux, L., Keefe, G., Leslie, K. and Campbell, J. (2002). The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows. *Can J Vet Res*, **66**(4): 249–257. [4](#), [8](#), [22](#), [51](#), [59](#)
- Ogbourne, C. P. (1971). Variations in the fecundity of strongylid worms of the horse. *Parasitology*, **63**(2): 289–98. [19](#)
- Ogbourne, C. P. (1976). The prevalence, relative abundance and site distribution of nematodes of the subfamily *Cyathostominae* in horses killed in Britain. *J Helminthol*, **50**(3): 203–214. [115](#)
- O'Meara, B. and Mulcahy, G. (2002). A survey of helminth control practices in equine establishments in Ireland. *Vet Parasitol*, **109**(1-2): 101–110. [115](#)

- Osterman Lind, E., Hoglund, J., Ljungstrom, B. L., Nilsson, O. and Uggla, A. (1999). A field survey on the distribution of strongyle infections of horses in Sweden and factors affecting faecal egg counts. *Equine Vet J*, **31**(1): 68–72. [87](#), [88](#), [119](#), [122](#)
- Pacala, S. W. and Dobson, A. P. (1988). The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation. *Parasitology*, **96** (Pt 1): 197–210. [6](#), [52](#), [191](#)
- Parnell, I. W. (1962). Observations on the seasonal variations in the worm burdens of young sheep in southern Western Australia. *J Helminthol*, **36**: 161–88. [19](#)
- Pascoe, R. J., Wilson, T. J. and Coles, G. C. (1999). Nematode control in eventer horses. *Vet Rec*, **145**(7): 200–201. [117](#)
- Paterson, S. and Lello, J. (2003). Mixed models: getting the best use of parasitological data. *Trends Parasitol*, **19**(8): 370–5. [8](#)
- Peek, N., Goud, R. and Abu-Hanna, A. (2008). Application of statistical process control methods to monitor guideline adherence: a case study. *AMIA Annu Symp Proc*, pp. 581–585. [154](#)
- Peng, Y. and Zhang, J. (2008). Identifiability of a mixture cure frailty model. *Statistics & Probability Letters*, **78**(16): 2604–2608. [8](#)
- Plummer, M. (2008). *Just Another Gibbs Sampler (JAGS)*. [15](#), [23](#), [43](#)
- Pook, J. F., Power, M. L., Sangster, N. C., Hodgson, J. L. and Hodgson, D. R. (2002). Evaluation of tests for anthelmintic resistance in cyathostomes. *Vet Parasitol*, **106**(4): 331–343. [115](#), [117](#), [118](#), [119](#), [122](#), [152](#)
- Presidente, P. (1985). *Methods for the detection of resistance to anthelmintics*. In: Anderson, N. and Waller, P.J. (editors) Resistance in Nematodes to Anthelmintic Drugs, CSIRO Australia and Australian Wool Corporation Technical Publication, new South Wales, Australia. [121](#)
- Presland, S. L., Morgan, E. R. and Coles, G. C. (2005). Counting nematode eggs in equine faecal samples. *Vet Rec*, **156**(7): 208–210. [19](#), [188](#), [196](#)
- Prichard, R. K., Hall, C. A., Kelly, J. D., Martin, I. C. and Donald, A. D. (1980). The problem of anthelmintic resistance in nematodes. *Aust Vet J*, **56**(5): 239–51. [2](#), [116](#), [119](#)
- Quenouille, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series. *Biometrics*, **5**(2): 162–4. [6](#)
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. [14](#), [26](#), [52](#)
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, **25**: 111–163. [10](#), [17](#)
- Raftery, A. E. and Lewis, S. M. (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. In *Practical Markov Chain Monte Carlo* (WR Gilks, DJ Spiegelhalter and S Richardson, eds) London, UK: Chapman and Hall. [16](#), [129](#)
- Reid, J. F. and Armour, J. (1972). Seasonal fluctuations and inhibited development of gastro-intestinal nematodes of sheep. *Res Vet Sci*, **13**(3): 225–9. [19](#)
- Reid, S. W., Mair, T. S., Hillyer, M. H. and Love, S. (1995). Epidemiological risk factors associated with a diagnosis of clinical cyathostomiasis in the horse. *Equine Vet J*, **27**(2): 127–30. [14](#)
- Reinders, R. D., Jonge, R. D. and Evers, E. G. (2003). A statistical method to determine whether micro-organisms are randomly distributed in a food matrix, applied to coliforms and *Escherichia coli* O157 in minced beef. *Food Microbiology*, **20**(3): 297 – 303. [7](#)

- Reinemeyer, C. R. (2009). Diagnosis and control of anthelmintic-resistant *Parascaris equorum*. *Parasites & Vectors*, **2**. 117
- Reinemeyer, C. R., Smith, S. A., Gabel, A. A. and Herd, R. P. (1984). The prevalence and intensity of internal parasites of horses in the U.S.A. *Vet Parasitol*, **15**(1): 75–83. 88
- Reinemeyer, C. R., Smith, S. A., Gabel, A. A. and Herd, R. P. (1986). Observations on the population dynamics of five cyathostome nematode species of horses in northern USA. *Equine Vet J*, **18**(2): 121–124. 88, 119
- Reinke, W. A. (1991). Applicability of industrial sampling techniques to epidemiologic investigations: examination of an underutilized resource. *Am J Epidemiol*, **134**(10): 1222–1232. 153
- Ripley, B. D. (1987). *Stochastic simulation*. Wiley, New York. 15
- Rissanen, J. (1978). Modeling By Shortest Data Description. *Automatica*, **14**: 465–471. 9
- Rodrigues-Motta, M., Gianola, D., Heringstad, B., Rosa, G. J. M. and Chang, Y. M. (2007). A zero-inflated poisson model for genetic analysis of the number of mastitis cases in Norwegian Red cows. *J Dairy Sci*, **90**(11): 5306–5315. 8
- Ros, M., SanCristobal, M., Dupont-Nivet, M. and Mallard, J. (2002). Comparison of maximum likelihood, bootstrap and MCMC methods in a structural model for heterogeneous variances. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, **Session 17**. 139
- Rossanigo, C. E. and Gruner, L. (1991). Accuracy of two methods for counting eggs of sheep nematode parasites. *Vet Parasitol*, **39**(1-2): 115–21. 19, 188
- Sangster, N. (1999a). Pharmacology of anthelmintic resistance in cyathostomes: will it occur with the avermectins / milbemycins? *Veterinary Parasitology*, **85**: 189–204. 116
- Sangster, N. C. (1999b). Anthelmintic resistance: past, present and future. *Int J Parasitol*, **29**(1): 115–24; discussion 137–8. 2, 116
- SAS (2003). *SAS statistical software*. SAS INSTITUTE INC, Cary, NC, USA. 52
- Schougaard, H. and Nielsen, M. K. (2007). Apparent ivermectin resistance of *Parascaris equorum* in foals in Denmark. *Vet Rec*, **160**(13): 439–40. 117
- Schumacher, J., Livesey, L., DeGraves, F., Blagburn, B., Ziska, S., Caldwell, M. and Brock, K. (2009). Efficacy of moxidectin against cyathostomins after long-term use in a large herd of draught horses with a high stocking density. *Vet Rec*, **164**(21): 652–4. 134
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**(2): 461–464. 9, 51
- Shaw, D. J. and Dobson, A. P. (1995). Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology*, **111 Suppl**: S111–27. 4
- Shaw, D. J., Grenfell, B. T. and Dobson, A. P. (1998). Patterns of macroparasite aggregation in wildlife host populations. *Parasitology*, **117** (Pt 6): 597–610. 5, 6, 22, 59, 84, 126, 191
- Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, **86**(2): 420–8. 157
- Sinniah, B. (1982). Daily egg production of *Ascaris lumbricoides*: the distribution of eggs in the faeces and the variability of egg counts. *Parasitology*, **84**(1): 167–175. 4, 5, 18, 187
- Smeal, M. G., Hotson, I. K., Mylrea, P. J., Jackson, A. R., Campbell, N. J. and Kirton, H. C. (1977). Studies on nematode infections of beef cattle in New South Wales. *Aust Vet J*, **53**(12): 566–73. 18, 19

- Smith, A. F. M. and Spiegelhalter, D. J. (1980). Bayes factors and choice criteria for linear-models. *Journal of the Royal Statistical Society Series B-Methodological*, **42**(2): 213–220. [10](#), [11](#)
- Smith, G. (1990). A mathematical model for the evolutions of anthelmintic resistance in a direct life cycle nematode parasite. *Int J Parasitol*, **20**(7): 913–21. [17](#)
- Smith, G. and Grenfell, B. T. (1994). Modelling of parasite populations: gastrointestinal nematode models. *Vet Parasitol*, **54**(1-3): 127–43. [17](#)
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. v. d. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **64**(4): 583–639. [9](#), [11](#)
- Sréter, T., Molnar, V. and Kassai, T. (1994). The distribution of nematode egg counts and larval counts in grazing sheep and their implications for parasite control. *Int J Parasitol*, **24**(1): 103–108. [2](#), [5](#), [50](#)
- Stanish, W. M. and Taylor, N. (1983). Estimation of the Intraclass Correlation Coefficient for the Analysis of Covariance Model. *The American Statistician*, **37**(3): 221–224. [157](#)
- Stear, M., Abuagob, O., Benothman, M., Bishop, S., Innocent, G., Kerr, A. and Mitchell, S. (2006). Variation among faecal egg counts following natural nematode infection in Scottish Blackface lambs. *Parasitology*, **132**(2): 275–80. [7](#)
- Stear, M., Bairden, K., Duncan, J., Gettinby, G., McKellar, Q., Murray, M. and Wallace, D. (1995a). The distribution of faecal nematode egg counts in Scottish Blackface lambs following natural, predominantly *Ostertagia circumcincta* infection. *Parasitology*, **110**(5): 573–81. [6](#)
- Stear, M. J., Bairden, K., Bishop, S. C., Gettinby, G., McKellar, Q. A., Park, M., Strain, S. and Wallace, D. S. (1998). The processes influencing the distribution of parasitic nematodes among naturally infected lambs. *Parasitology*, **117** (Pt 2): 165–71. [2](#), [89](#)
- Stear, M. J., Bairden, K., Innocent, G. T., Mitchell, S., Strain, S. and Bishop, S. C. (2004). The relationship between IgA activity against 4th-stage larvae and density-dependent effects on the number of 4th-stage larvae of *Teladorsagia circumcincta* in naturally infected sheep. *Parasitology*, **129**(Pt 3): 363–369. [14](#), [50](#)
- Stear, M. J. and Bishop, S. C. (1999). The curvilinear relationship between worm length and fecundity of *Teladorsagia circumcincta*. *Int J Parasitol*, **29**(5): 777–80. [7](#), [18](#)
- Stear, M. J., Bishop, S. C., Duncan, J. L., McKellar, Q. A. and Murray, M. (1995b). The repeatability of faecal egg counts, peripheral eosinophil counts, and plasma pepsinogen concentrations during deliberate infections with *Ostertagia circumcincta*. *Int J Parasitol*, **25**(3): 375–380. [119](#), [171](#)
- Stoll, N. (1930). On methods of counting nematode ova in sheep dung. *Parasitology*, **22**: 116–136. [18](#)
- Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society Series B (Methodological)*, **41**(2): 276–278. [11](#)
- Streftaris, G. and Gibson, G. (2004). Bayesian analysis of experimental epidemics of foot-and-mouth disease. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**(1544): 1111–1117. [12](#)
- Tarigo-Martinie, J. L., Wyatt, A. R. and Kaplan, R. M. (2001). Prevalence and clinical implications of anthelmintic resistance in cyathostomes of horses. *J Am Vet Med Assoc*, **218**(12): 1957–60. [117](#)
- Thomas, A., Hara, B. O., Ligges, U. and Sturtz, S. (2006). Making BUGS Open. *R News*, **6**: 12–17. [15](#), [43](#)
- Thorburn, W. (1918). The Myth of Occam’s Razor. *Mind*, **27**: 345–353. [9](#), [51](#)

- Thygesen, H. H. and Zwinderman, A. H. (2006). Modeling Sage data with a truncated gamma-Poisson model. *BMC Bioinformatics*, **7**: 157. [6](#)
- Tierney, L. (1994). Markov-Chains for Exploring Posterior Distributions. *Annals of Statistics*, **22**(4): 1701–1728. [15](#)
- Toft, N., Innocent, G. T., Gettinby, G. and Reid, S. W. J. (2007). Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. *Prev Vet Med*, **79**(2-4): 244–256. [16](#), [52](#)
- Toft, N., Innocent, G. T., Mellor, D. J. and Reid, S. W. J. (2006). The Gamma-Poisson model as a statistical method to determine if micro-organisms are randomly distributed in a food matrix. *Food Microbiol*, **23**(1): 90–94. [6](#), [59](#)
- Torgerson, P. R., Schnyder, M. and Hertzberg, H. (2005). Detection of anthelmintic resistance: a comparison of mathematical techniques. *Vet Parasitol*, **128**(3-4): 291–298. [xvi](#), [4](#), [5](#), [14](#), [18](#), [22](#), [50](#), [52](#), [114](#), [120](#), [191](#)
- Traversa, D., von Samson-Himmelstjerna, G., Demeler, J., Milillo, P., Schuermann, S., Barnes, H., Otranto, D., Perrucci, S., di Regalbono, A. F., Beraldo, P., Boeckh, A. and Cobb, R. (2009). Anthelmintic resistance in cyathostomin populations from horse yards in Italy, United Kingdom and Germany. *Parasites & Vectors*, **2**. [117](#)
- Trawford, A. F., Burden, F. and Hodgkinson, J. (2005). Suspected moxidectin resistance in cyathostomins in two donkey herds at the Donkey Sanctuary, UK. *Proceedings of the 20th International Conference of the World Association for the Advancement of Veterinary Parasitology, Christchurch, October 16 to 20*, p. 196. [117](#)
- Uhlinger, C. (1990). Effects of three anthelmintic schedules on the incidence of colic in horses. *Equine Veterinary Journal*, **22**(4): 251–254. [88](#)
- Uhlinger, C. (1991). Equine small strongyles; epidemiology, pathology and control. *Compend Contin Educ Pract Vet*, **13**: 863–869. [2](#), [4](#), [88](#), [114](#), [115](#), [119](#)
- Uhlinger, C. A. (1993). Uses of Fecal Egg Count Data in Equine Practice. *Compendium On Continuing Education For the Practicing Veterinarian*, **15**(5): 742–749. [2](#), [4](#), [5](#), [18](#), [19](#), [88](#), [114](#), [115](#), [119](#), [123](#), [152](#), [155](#), [187](#), [189](#)
- Urquhart, G., Armour, J., Duncan, J., Dunn, A. and Jennings, F. (1996). *Veterinary Parasitology*. Black Well Science Ltd. Oxford and London, UK, 2nd edition. [2](#), [3](#), [4](#), [87](#), [88](#), [89](#), [115](#), [119](#)
- van de Ven, R. (1993). Estimating the shape parameter of the negative binomial distribution. *Journal of Statistical Computation and Simulation*, **46**: 111–123. [14](#)
- van de Ven, R. and Weber, N. (1999). On the asymptotic behaviour of estimators of the shape parameter of the negative binomial distribution. *ISI 99 52nd Session Bulletin of the International Statistical Institute, pp 419-420 Edita Ltd*. [14](#)
- Vanimisetti, H. B., Andrew, S. L., Zajac, A. M. and Notter, D. R. (2004). Inheritance of fecal egg count and packed cell volume and their relationship with production traits in sheep infected with *Haemonchus contortus*. *J Anim Sci*, **82**(6): 1602–11. [19](#)
- Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*. Springer, Berlin, fourth edition. [5](#), [58](#)
- Vercruysse, J., Holdsworth, P., Letonja, T., Barth, D., Conder, G., Hamamoto, K. and Okano, K. (2001). International harmonisation of anthelmintic efficacy guidelines. *Vet Parasitol*, **96**(3): 171–193. [122](#)

- Vercruysse, J., Holdsworth, P., Letonja, T., Conder, G., Hamamoto, K., Okano, K. and Rehbein, S. (2002). International harmonisation of anthelmintic efficacy guidelines (Part 2). *Vet Parasitol*, **103**(4): 277–297. [123](#)
- Vidyashankar, A. and Kaplan, R. (2005). A simulation approach for evaluating methods used to identify treatment effects in studies of anthelmintic resistance. *Proceedings of the 20th International Conference of the World Association for the Advancement of Veterinary Parasitology*. [118](#), [120](#), [122](#), [154](#)
- Vidyashankar, A. N., Kaplan, R. M. and Chan, S. (2007). Statistical approach to measure the efficacy of anthelmintic treatment on horse farms. *Parasitology*, **134**(Pt.14): 2027–2039. [4](#), [17](#), [18](#), [126](#), [133](#), [153](#)
- Vizard, A. L. and Wallace, R. J. (1987). A simplified faecal egg count reduction test. *Aust Vet J*, **64**(4): 109–111. [122](#)
- von Samson-Himmelstjerna, G., Fritzen, B., Demeler, J., Schurmann, S., Rohn, K., Schnieder, T. and Epe, C. (2007). Cases of reduced cyathostomin egg-reappearance period and failure of *Parascaris equorum* egg count reduction following ivermectin treatment as well as survey on pyrantel efficacy on German horse farms. *Vet Parasitol*, **144**(1-2): 74–80. [154](#)
- von Samson-Himmelstjerna, G., von Witzendorff, C., Sievers, G. and Schnieder, T. (2002). Comparative use of faecal egg count reduction test, egg hatch assay and beta-tubulin codon 200 genotyping in small strongyles (cyathostominae) before and after benzimidazole treatment. *Vet Parasitol*, **108**(3): 227–235. [117](#), [118](#)
- Vose, D. (2004). *ModelAssist for @Risk*. Risk Thinking, Ltd. [5](#), [58](#)
- Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, **57**(2): 307–333. [9](#)
- Walker, M., Hall, A., Anderson, R. M. and Basáñez, M.-G. (2009). Density-dependent effects on the weight of female *Ascaris lumbricoides* infections of humans and its impact on patterns of egg production. *Parasit Vectors*, **2**(1): 11. [8](#), [59](#)
- Waller, P. J. (1997). Anthelmintic resistance. *Vet Parasitol*, **72**(3-4): 391–405. [117](#)
- Warnick, L. D. (1992). Daily variability of equine fecal strongyle egg counts. *Cornell Vet*, **82**(4): 453–463. [87](#), [119](#)
- Warton, D. I. (2005). Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics*, **16**(3): 275–289. [12](#)
- Wasserman (2000). Bayesian Model Selection and Model Averaging. *J Math Psychol*, **44**(1): 92–107. [11](#), [17](#)
- Whitlock, H. V. and Gordon, H. M. (1939). A new technique for counting nematode eggs in sheep faeces. *J Court Sci Indust Res Aust*, **12**: 50–52. [18](#)
- Williams, B. G. and Dye, C. (1994). Maximum likelihood for parasitologists. *Parasitol Today*, **10**(12): 489–93. [14](#)
- Williams, C. B. (1947). The Logarithmic Series and Its Application to Biological Problems. *The Journal of Ecology*, **34**(2): 253–272. [6](#)
- Williamson, G. D. and Weatherby Hudson, G. (1999). A monitoring system for detecting aberrations in public health surveillance reports. *Stat Med*, **18**(23): 3283–3298. [154](#)
- Wilson, K., Grenfell, B. and Shaw, D. (1996). Analysis of aggregated parasite distributions: a comparison of methods. *Functional Ecology*, **10**: 592–601. [4](#), [13](#), [14](#), [18](#), [50](#), [52](#), [122](#), [191](#)

- Wilson, K. and Grenfell, B. T. (1997). Generalized linear modelling for parasitologists. *Parasitology Today*, **13**(1): 33 – 38. [4](#), [5](#), [6](#), [13](#), [14](#), [59](#), [84](#), [191](#)
- Winter, M. D. (2002). *Nematodirus battus* 50 years on—a realistic vaccine candidate? *Trends Parasitol*, **18**(7): 298–301. [89](#)
- Woolhouse, M. E. (1992). A theoretical framework for the immunoepidemiology of helminth infection. *Parasite Immunol*, **14**(6): 563–578. [4](#), [17](#)
- Xiang, L., Lee, A. H., Yau, K. K. W. and McLachlan, G. J. (2007). A score test for overdispersion in zero-inflated poisson mixed regression model. *Stat Med*, **26**(7): 1608–1622. [59](#)
- Yergeau, E., Bokhorst, S., Huiskes, A. H. L., Boschker, H. T. S., Aerts, R. and Kowalchuk, G. A. (2007). Size and structure of bacterial, fungal and nematode communities along an Antarctic environmental gradient. *FEMS Microbiol Ecol*, **59**(2): 436–451. [2](#)
- Young, K. E., Garza, V., Snowden, K., Dobson, R. J., Powell, D. and Craig, T. M. (1999). Parasite diversity and anthelmintic resistance in two herds of horses. *Vet Parasitol*, **85**(2-3): 205–214. [88](#), [117](#)

APPENDIX A

Model specifications

Model specifications

A.1 Faecal egg count models

```
for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * meancount[row];
probpos[row] ~ dbern(prob);
meancount[row] ~ dgamma(a, b);
}
mean <- a / b

# Priors
prob ~ dunif(0,1);
a <- 1/ia;
ia ~ dunif(0.0001,100);
b ~ dunif(0.000001,10000);
```

Model 2.1: - Independent shape and scale parameters

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row] * mean;
probpos[row] ~ dbern(prob);
gamma[row] ~ dgamma(a, a);
}

# Priors
mean ~ dunif(0.001,1000);
a <- 1/ia;
ia ~ dunif(0.0001,100);
prob ~ dunif(0,1);

```

Model 2.2: - Gamma distribution centred on one

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
probpos[row] ~ dbern(prob);
gamma[row] ~ dgamma(a, b);
}
b <- a / mean;

# Priors
mean ~ dunif(0.001,1000);
a <- 1/ia;
ia ~ dunif(0.0001,100);
prob ~ dunif(0,1);

```

Model 2.3: - Gamma distribution centred on mean

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * mean * gamma[row];
probpos[row] ~ dbern(ind.prob[row]);
gamma[row] ~ dgamma(a, b);
}
prob <- mean(ind.prob[]);
b <- a / mean;

# Priors
mean ~ dunif(0.001,1000);
a <- 1/ia;
ia ~ dunif(0.0001,100);
for(row in 1 : N){
ind.prob[row] ~ dunif(0,1);
}

```

Model 2.4: - Independent zero-inflation for each count

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row] * mean;
probpos[row] ~ dbern(prob);
gamma[row] ~ dgamma(a, a);
}

# Priors
mean ~ dunif(0.001,1000);
a <- 1/ia;
ia <- exp(log.ia);
log.ia ~ dunif(-9.21,4.6);
prob ~ dunif(0,1);

```

Model 2.2a: - Gamma distribution centred on one with log prior on variability

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
probpos[row] ~ dbern(prob);
gamma[row] ~ dgamma(a, b);
}
b <- a / mean;

# Priors
mean ~ dunif(0.001,1000);
a <- 1/ia;
ia <- exp(log.ia);
log.ia ~ dunif(-9.21,4.6);
prob ~ dunif(0,1);

```

Model 2.3a: - Gamma distribution centred on mean with log prior on variability

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
gamma[row] ~ dlnorm(lmu, lprec);
probpos[row] ~ dbern(prob);
}

lprec <- 1 / lsd^2;
meanl <- log(0.001) - ((lsd^2)/2);
meanu <- log(1000) - ((lsd^2)/2);

# Priors
lmu ~ dunif(meanl,meanu);
prob ~ dunif(0,1);
lsd ~ dunif(0.01,2.149);
}

```

Model 2.5: - Prior distribution for log standard deviation and log mean

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
gamma[row] ~ dlnorm(lmu, lprec);
probpos[row] ~ dbern(prob);
}

lvar <- log((sd/mean)^2 + 1);
lprec <- 1 / lvar;
lmu <- log(mean) - ((lvar) / 2)

# Priors
mean ~ dunif(0.001,1000);
prob ~ dunif(0,1);
sd ~ dunif(0.000001,10000);
}

```

Model 2.6: - Prior distribution for standard deviation and mean

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
gamma[row] ~ dlnorm(lmu, lprec);
probpos[row] ~ dbern(prob);
}

lprec <- 1 / lsd^2;
lmu <- log(mean) - ((lsd^2) / 2)

# Priors
mean ~ dunif(0.001,1000);
prob ~ dunif(0,1);
lsd ~ dunif(0.01,2.149);
}

```

Model 2.7: - Prior distribution for log standard deviation and mean

```

for(row in 1 : N){
Count[row] ~ dpois(lambda[row]);
lambda[row] <- probpos[row] * gamma[row];
gamma[row] ~ dlnorm(lmu, lprec);
probpos[row] ~ dbern(prob);
}

lprec <- 1 / lsd^2;
lsd <- exp(llsd);
meanl <- log(0.001) - ((lsd^2)/2);
meanu <- log(1000) - ((lsd^2)/2);

# Priors
lmu ~ dunif(meanl,meanu);
prob ~ dunif(0,1);
llsd ~ dunif(-4.61,0.81);
}

```

Model 2.8: - Prior distribution for log of the log standard deviation and log mean

A.2 Faecal egg count reduction test models

```

for(row in 1:N){
Pre[row] ~ dpois(xpre.lambda[row])
Post[row] ~ dpois(xpost.lambda[row])
xpre.lambda[row] <- probpos[row] * pre.mean * pre.gamma[row]
xpost.lambda[row] <- probpos[row] * post.mean * post.gamma[row]
pre.gamma[row] ~ dgamma(pre.disp, pre.disp)T(10^-200,)
post.gamma[row] ~ dgamma(post.disp, post.disp)T(10^-200,)
delta.mean[row] <- post.mean * post.gamma[row] / pre.mean * pre.gamma[row]
probpos[row] ~ dbern(prob)
}

pre.disp <- 1 / ia
post.mean <- pre.mean * efficacy
post.disp <- pre.disp * delta.disp

pre.mean ~ dunif(0.001, 1000)
prob ~ dunif(0,1)
ia <- exp(logia)
logia ~ dunif(-9.21,4.6)
efficacy ~ dbeta(1,1)
delta.disp ~ dlnorm(0, 0.01)T(0.001,1000)

```

Model 5.1: - FECRT model based on two (zero-inflated) gamma-Poisson distributions with scaling mean and shape parameters deriving efficacy and change in variability

```

for(row in 1:N){
Pre[row] ~ dpois(xpre.lambda[row])
Post[row] ~ dpois(xpost.lambda[row])
xpre.lambda[row] <- probpos[row] * pre.mean[row] * pre.gamma[row]
xpost.lambda[row] <- probpos[row] * post.mean[row] * post.gamma[row]
pre.gamma[row] ~ dgamma(pre.disp, pre.disp)T(10^-200,)
post.gamma[row] ~ dgamma(post.disp, post.disp)T(10^-200,)
pre.mean[row] <- group.mean * animal.gamma[row]
animal.gamma[row] ~ dgamma(animal.disp, animal.disp)T(10^-200,)
post.mean[row] <- pre.mean[row] * delta.mean[row]
ind.delta[row] <- (post.mean * post.gamma[row]) / (pre.mean * pre.gamma[row])
probpos[row] ~ dbern(prob)
delta.mean[row] <- efficacy
}

pre.disp <- 1 / ia
post.disp <- pre.disp * delta.disp
animal.disp <- 1 / iaa

group.mean ~ dunif(0.001, 1000)
prob ~ dunif(0,1)
ia <- exp(logia)
logia ~ dunif(-9.21,4.6)
iaa <- exp(logiaa)
logiaa ~ dunif(-9.21,4.6)
efficacy ~ dbeta(1,1)
delta.disp ~ dlnorm(0, 0.01)T(0.001,1000)

```

Model 5.2: - FECRT model based on compound gamma-Poisson distributions separating within and between animal variability

```

for(herd in 1:n.herds){
for(horse in 1:n.horses[herd]){
Pre[horse,herd] ~ dpois(xpre.lambda[horse,herd])
Post[horse,herd] ~ dpois(xpost.lambda[horse,herd])
xpre.lambda[horse,herd] <- pre.mean[horse,herd] * probpos[horse,herd]
  * pre.gamma[horse,herd]
xpost.lambda[horse,herd] <- post.mean[horse,herd] * probpos[horse,herd]
  * post.gamma[horse,herd]
pre.gamma[horse,herd] ~ dgamma(pre.disp[herd], pre.disp[herd])T(10^-200,)
post.gamma[horse,herd] ~ dgamma(post.disp[herd], post.disp[herd])T(10^-200,)
pre.mean[horse,herd] <- group.mean[herd] * animal.gamma[horse,herd]
animal.gamma[horse,herd] ~ dgamma(animal.disp[herd], animal.disp[herd])
  T(10^-200,)
post.mean[horse,herd] <- pre.mean[horse,herd] * delta.mean[horse,herd]
probpos[horse,herd] ~ dbern(prob[herd])
delta.mean[horse,herd] <- efficacy[herd]
}

pre.disp[herd] <- 1 / ia[herd]
animal.disp[herd] <- 1 / iaa[herd]
post.disp[herd] <- pre.disp[herd] * delta.disp[herd]

# Priors
group.mean[herd] ~ dunif(0.1, 1000)
ia[herd] <- exp(logia[herd])
logia[herd] ~ dunif(-9.21,4.6);
iaa[herd] <- exp(logiaa[herd])
logiaa[herd] ~ dunif(-9.21,4.6);
prob[herd] ~ dunif(0,1)

efficacy[herd] ~ dgamma(dma, dmb)
delta.disp[herd] ~ dgamma(dda,ddb)
}

dmb <- dma / mean.efficacy
dma <- 1 / exp(logdmia)
logdmia ~ dunif(-4.6,4.6)
mean.efficacy ~ dbeta(1,1)
ddb <- dda / mean.deltadisp
dda <- 1 / exp(logddia)
logddia ~ dunif(-4.6,4.6)
mean.deltadisp ~ dlnorm(0, 0.01)

```

Model 5.3: - FECRT model based on compound gamma-Poisson distributions separating within and between animal variability, adapted for meta-population analysis

A.3 Models for quantifying variability of FEC

```

for(i in 1:9){
adj.replicate.shape[i] <- 1/(mcmasters.cv^2 + ((i^0.5/i)*replicate.cv)^2 +
  (mcmasters.cv^2 * ((i^0.5/i)*replicate.cv)^2))
}

for(animal in 1:Nani){
ani.mean[animal] ~ dunif(0,100)
for(day in 1:Ndays){
day.mean[animal,day] ~ dgamma(day.shape, day.shape/ani.mean[animal])
for(sample in 1:Nsamples[animal,day]){
sample.mean[animal,day,sample]
  ~ dgamma(sample.shape, sample.shape/day.mean[animal,day])
for(replicate in 1:Nreplicates[animal,day,sample]){
for(repeat in 1:Nrepeats[animal,day,sample,replicate]){
Count[animal,day,sample,replicate,repeat]
  ~ dpois(replicate.lambda[animal,day,sample,replicate])
}
replicate.lambda[animal,day,sample,replicate]
  ~ dgamma(adj.replicate.shape[Grams[animal,day,sample,replicate]],
  adj.replicate.shape[Grams[animal,day,sample,replicate]]
  /sample.mean[animal,day,sample])
}
}
}
}

replicate.logia ~ dunif(-9.21,4.6); replicate.shape <- 1/exp(replicate.logia)
replicate.cv <- 1/replicate.shape^0.5
mcmasters.logia ~ dunif(-9.21,4.6); mcmasters.shape <- 1/exp(mcmasters.logia)
mcmasters.cv <- 1/mcmasters.shape^0.5
sample.logia ~ dunif(-9.21,4.6); sample.shape <- 1/exp(sample.logia)
sample.cv <- 1/sample.shape^0.5
day.logia ~ dunif(-9.21,4.6); day.shape <- 1/exp(day.logia)
day.cv <- 1/day.shape^0.5

```

Model 6.1: - Variability quantification model using a hierarchical compound gamma distribution with log-uniform priors on the inverse shape parameter

```

for(i in 1:9){
adj.replicate.shape[i] <- 1/(mcmasters.cv^2 + ((i^0.5/i)*replicate.cv)^2
  + (mcmasters.cv^2 * ((i^0.5/i)*replicate.cv)^2))
adj.sample.shape[i] <- 1 / (((1/adj.replicate.shape[i]) + (1/sample.shape)
  + ((1/adj.replicate.shape[i]) * (1/sample.shape))))
adj.day.shape[i] <- 1 / (((1/adj.sample.shape[i]) + (1/day.shape)
  + ((1/adj.sample.shape[i]) * (1/day.shape))))
}

for(i in 1:Days){
for(j in 1:Day.sublength[i]){
for(k in 1:Day.repeats[i,j]){
Day.count[i,j,k] ~ dpois(day.lambda[i,j])}
day.lambda[i,j] ~ dgamma(adj.day.shape[Day.grams[i,j]],
  adj.day.shape[Day.grams[i,j]]/day.mean[i])}
day.mean[i] ~ dunif(0,100)
# Doubled up records from Day.count are NA
}

for(i in 1:Samples){
for(j in 1:Sample.sublength[i]){
for(k in 1: Sample.repeats[i,j]){
Sample.count[i,j,k] ~ dpois(sample.lambda[i,j])}
sample.lambda[i,j] ~ dgamma(adj.sample.shape[Sample.grams[i,j]],
  adj.sample.shape[Sample.grams[i,j]]/sample.mean[i])}
sample.mean[i] ~ dgamma(day.shape, day.shape/day.mean[Sample.daytrack[i]])
# Doubled up records from Sample.count are NA
}

for(i in 1:Replicates){
for(j in 1:Replicate.sublength[i]){
for(k in 1:Replicate.repeats[i,j]){
Replicate.count[i,j,k] ~ dpois(replicate.lambda[i,j])}
replicate.lambda[i,j] ~ dgamma(adj.replicate.shape[Replicate.grams[i,j]],
  adj.replicate.shape[Replicate.grams[i,j]]/replicate.mean[i])}
replicate.mean[i] ~ dgamma(sample.shape,
  sample.shape/sample.mean[Replicate.sampletrack[i]])
}

replicate.logia ~ dunif(-9.21,4.6); replicate.shape <-
  1/exp(replicate.logia); replicate.cv <- 1/replicate.shape^0.5
mcmasters.logia ~ dunif(-9.21,4.6); mcmasters.shape <-
  1/exp(mcmasters.logia); mcmasters.cv <- 1/mcmasters.shape^0.5
sample.logia ~ dunif(-9.21,4.6); sample.shape <-
  1/exp(sample.logia); sample.cv <- 1/sample.shape^0.5
day.logia ~ dunif(-9.21,4.6); day.shape <- 1/exp(day.logia)
day.cv <- 1/day.shape^0.5

```

Model 6.2: - Variability quantification model using separate non-hierarchical gamma distributions with log-uniform priors on the inverse shape parameter

```

for(animal in 1:Nani){
ani.mean[animal] ~ dunif(0,1000)
for(day in 1:Ndays){
day.mean[animal,day] ~ dgamma(day.a[animal], day.a[animal]/ani.mean[animal])
for(sample in 1:Nsamples[animal,day]){
sample.mean[animal,day,sample] ~ dgamma(sample.a[animal], sample.a[animal]
/day.mean[animal,day])
for(replicate in 1:Nreplicates[animal,day,sample]){
replicate.mean[animal,day,sample,replicate]
~ dgamma(adj.replicate.a[animal,day,sample,replicate],
(adj.replicate.a[animal,day,sample,replicate]
/sample.mean[animal,day,sample]))
adj.replicate.a[animal,day,sample,replicate] <- 1 / ((1/(replicate.a[animal]
*Grams[animal,day,sample,replicate])) + (1/(mcmasters.a)
+ ((1/(replicate.a[animal]*Grams[animal,day,sample,replicate]))
*(1/(mcmasters.a))))))
for(repeat in 1:Nrepeats[animal,day,sample,replicate]){
Count[animal,day,sample,replicate,repeat]
~ dpois(replicate.mean[animal,day,sample,replicate])
}
}
}
}

day.a[animal] <- 1 / exp(i.day.a[animal])
sample.a[animal] <- 1 / exp(i.sample.a[animal])
replicate.a[animal] <- 1 / exp(i.replicate.a[animal])
i.day.a[animal] ~ dunif(-9.21,4.6)
i.sample.a[animal] ~ dunif(-9.21,4.6)
i.replicate.a[animal] ~ dunif(-9.21,4.6)
day.cv[animal] <- 1/day.a[animal]^0.5
sample.cv[animal] <- 1/sample.a[animal]^0.5
replicate.cv[animal] <- 1/replicate.a[animal]^0.5
}

mcmasters.a <- 1 / exp(i.mcmasters.a)
i.mcmasters.a ~ dunif(-9.21,4.6)
mcmasters.cv <- 1/mcmasters.a^0.5

```

Model 6.3: - Multiple animal variability quantification model using a hierarchical compound gamma distribution with log-uniform priors on the inverse shape parameter

```

for(animal in 1:Nani){
ani.mean[animal] ~ dunif(0,1000)
for(day in 1:Ndays){
day.mean[animal,day] ~ dgamma(day.a[animal], day.a[animal]/ani.mean[animal])
for(sample in 1:Nsamples[animal,day]){
sample.mean[animal,day,sample] ~ dgamma(sample.a[animal], sample.a[animal]
/day.mean[animal,day])
for(replicate in 1:Nreplicates[animal,day,sample]){
replicate.mean[animal,day,sample,replicate]
~ dgamma(adj.replicate.a[animal,day,sample,replicate],
(adj.replicate.a[animal,day,sample,replicate]
/sample.mean[animal,day,sample]))
adj.replicate.a[animal,day,sample,replicate] <- 1 / ((1/(replicate.a[animal]
*Grams[animal,day,sample,replicate])) + (1/(mcmasters.a)
+ ((1/(replicate.a[animal]*Grams[animal,day,sample,replicate]))
*(1/(mcmasters.a))))))
for(repeat in 1:Nrepeats[animal,day,sample,replicate]){
Count[animal,day,sample,replicate,repeat]
~ dpois(replicate.mean[animal,day,sample,replicate])
}
}
}
}

day.cv[animal] ~ dunif(0.001,100)
day.a[animal] <- 1/day.cv[animal]^2
sample.cv[animal] ~ dunif(0.001,100)
sample.a[animal] <- 1/sample.cv[animal]^2
replicate.cv[animal] ~ dunif(0.001,100)
replicate.a[animal] <- 1/replicate.cv[animal]^2
}

mcmasters.cv ~ dunif(0.001,100)
mcmasters.a <- 1/mcmasters.cv^2

```

Model 6.3a: - Multiple animal variability quantification model using a hierarchical compound gamma distribution with uniform priors on the *cv*

```

for(animal in 1:Nani){
ani.mean[animal] ~ dunif(0,1000)
for(day in 1:Ndays){
day.mean[animal,day] ~ dgamma(day.a[animal], day.a[animal]/ani.mean[animal])
for(sample in 1:Nsamples[animal,day]){
sample.mean[animal,day,sample] ~ dgamma(sample.a[animal], sample.a[animal]
/day.mean[animal,day])
for(replicate in 1:Nreplicates[animal,day,sample]){
replicate.mean[animal,day,sample,replicate]
~ dgamma(adj.replicate.a[animal,day,sample,replicate],
(adj.replicate.a[animal,day,sample,replicate]
/sample.mean[animal,day,sample]))
adj.replicate.a[animal,day,sample,replicate] <- 1 / (((1/(replicate.a
*Grams[animal,day,sample,replicate])) + (1/(mcmasters.a)
+ ((1/(replicate.a*Grams[animal,day,sample,replicate]))
*(1/(mcmasters.a))))))
for(repeat in 1:Nrepeats[animal,day,sample,replicate]){
Count[animal,day,sample,replicate,repeat]
~ dpois(replicate.mean[animal,day,sample,replicate])
}
}
}
}

day.a[animal] <- 1 / exp(i.day.a[animal])
sample.a[animal] <- 1 / exp(i.sample.a[animal])
i.day.a[animal] ~ dunif(-9.21,4.6)
i.sample.a[animal] ~ dunif(-9.21,4.6)
day.cv[animal] <- 1/day.a[animal]^0.5
sample.cv[animal] <- 1/sample.a[animal]^0.5
}

replicate.a <- 1 / exp(i.replicate.a)
i.replicate.a ~ dunif(-9.21,4.6)
replicate.cv <- 1/replicate.a^0.5
mcmasters.a <- 1 / exp(i.mcmasters.a)
i.mcmasters.a ~ dunif(-9.21,4.6)
mcmasters.cv <- 1/mcmasters.a^0.5

```

Model 6.4: - Multiple animal variability quantification model using a hierarchical compound gamma distribution with log-uniform priors on the inverse shape parameter and replicate.cv fixed between animals

APPENDIX B

Derivation of the *cv* of a compound gamma distribution

Derivation of the cv of a compound gamma distribution

In several places within this thesis, the effective cv of a distribution consisting of the combination of two gamma distributions has been calculated from the variability of each individual distribution using the following formula:

$$cv_{total} = \sqrt{(cv_1^2 + cv_2^2 + cv_1^2 \times cv_2^2)}$$

Where cv_1 and cv_2 are the cv of each independent distribution, and cv_{total} is the effective cv of the compound distribution.

In this Appendix, the derivation of this formula from the combination of two gamma distributions is given (thanks to Louise Matthews for assistance with the work shown here). Throughout, the gamma distributions will be defined in terms of the mean, μ , and variance, σ^2 , which are related to the shape, α , and scale, β , parameters as follows:

$$\begin{aligned}\mu &= \alpha\beta \\ \sigma^2 &= \alpha\beta^2\end{aligned}$$

Let X represent a random variate sampled from a gamma distribution with probability density function $p(X = x) = f(x; \mu, \sigma^2)$. Further, if μ is a sample from a further gamma distribution with pdf $p(M = \mu) = g(\mu; \mu_{pop}, \sigma^2_{pop})$ then the combined pdf is given by:

$$p(X = x) = \int f(x; \mu, \sigma^2) g(\mu; \mu_{pop}, \sigma^2_{pop}) d\mu$$

For simplicity, this can be rewritten as:

$$pdf(x) = \int f(x; \mu) g(\mu) d\mu$$

The objects of interest are the expectation (mean or $E[X]$) and expected variance ($Var[X]$) of the compound distribution represented by $pdf(x)$. The expectation is given by:

$$\begin{aligned}
 E[X] &= \int x \, pdf(x) \, dx \\
 &= \int x \int f(x\mu) g(\mu) \, d\mu \, dx \\
 &= \int g(\mu) \int x f(x\mu) \, dx \, d\mu
 \end{aligned}$$

The expression $\int x f(x\mu) \, dx$ is equal to the mean of x , which by definition is μ . Therefore:

$$\begin{aligned}
 E[X] &= \int g(\mu) \left(\int x f(x\mu) \, dx \right) \, d\mu \\
 &= \int g(\mu) \mu \, d\mu \\
 &= \mu_{pop}
 \end{aligned}$$

The expected variance of $pdf(x)$ is given by:

$$\begin{aligned}
 Var[X] &= E[(X - \bar{x})^2] \\
 &= E[X^2] - (E[x])^2 \\
 &= \int x^2 \, pdf(x) \, dx - \mu_{pop}^2 \\
 &= \int x^2 \int f(x\mu) g(\mu) \, d\mu \, dx - \mu_{pop}^2 \\
 &= \int g(\mu) \left(\int x^2 f(x\mu) \, dx \right) \, d\mu - \mu_{pop}^2 \\
 &= \int g(\mu) (Var(f) + \mu^2) \, d\mu - \mu_{pop}^2
 \end{aligned}$$

As $Var(f) = \alpha\beta^2$ and $E(f) = \alpha\beta$; $Var(f)$ can be substituted by $\mu^2 \div \alpha$. Therefore:

$$\begin{aligned}
 Var[x] &= \int g(\mu) (Var(f) + \mu^2) \, d\mu - \mu_{pop}^2 \\
 &= \int g(\mu) \left(\frac{\mu^2}{\alpha} + \mu^2 \right) \, d\mu - \mu_{pop}^2 \\
 &= \left(1 + \frac{1}{\alpha} \right) \int g(\mu) \mu^2 \, d\mu - \mu_{pop}^2
 \end{aligned}$$

Substituting $\int g(\mu) \mu^2 \, d\mu = Var[g(\mu)] + E[g(\mu)]^2 = Var[g(\mu)] + \mu_{pop}^2$ gives:

$$\begin{aligned}
 Var[x] &= \left(1 + \frac{1}{\alpha}\right) \int g(\mu) \mu^2 d\mu - \mu_{pop}^2 \\
 &= \left(1 + \frac{1}{\alpha}\right) (Var[g(\mu)] + \mu_{pop}^2) - \mu_{pop}^2 \\
 &= \left(1 + \frac{1}{\alpha}\right) Var[g(\mu)] + \left(1 + \frac{1}{\alpha}\right) \mu_{pop}^2 - \mu_{pop}^2 \\
 &= \left(1 + \frac{1}{\alpha}\right) Var[g(\mu)] + \left(\frac{1}{\alpha}\right) \mu_{pop}^2
 \end{aligned}$$

In addition, substituting $Var[g(\mu)] = \frac{\mu_{pop}^2}{\alpha_{pop}}$ and $cv = \frac{\sigma}{\mu} = \frac{\sqrt{(\alpha\beta^2)}}{\alpha\beta} = \frac{\sqrt{\alpha}}{\alpha} = \frac{1}{\sqrt{\alpha}}$ gives:

$$\begin{aligned}
 Var[x] &= \left(1 + \frac{1}{\alpha}\right) Var[g(\mu)] + \left(\frac{1}{\alpha}\right) \mu_{pop}^2 \\
 &= \left(1 + \frac{1}{\alpha}\right) \times \frac{\mu_{pop}^2}{\alpha_{pop}} + \left(\frac{1}{\alpha}\right) \mu_{pop}^2 \\
 &= (1 + cv^2) \times \mu_{pop}^2 \times cv_{pop}^2 + cv^2 \mu_{pop}^2 \\
 &= \mu_{pop}^2 \times cv_{pop}^2 + \mu_{pop}^2 \times cv_{pop}^2 \times cv^2 + cv^2 \times \mu_{pop}^2
 \end{aligned}$$

Finally, since $cv_{total} = \frac{\sigma}{\mu_{pop}}$, the total effective cv is given by:

$$\begin{aligned}
 cv_{total} &= \frac{\sqrt{(\mu_{pop}^2 \times cv_{pop}^2 + \mu_{pop}^2 \times cv_{pop}^2 \times cv^2 + cv^2 \times \mu_{pop}^2)}}{\mu_{pop}} \\
 &= \frac{\sqrt{(cv_{pop}^2 + cv^2 + cv_{pop}^2 \times cv^2)} \times \sqrt{\mu_{pop}^2}}{\mu_{pop}} \\
 &= \sqrt{(cv_{pop}^2 + cv^2 + cv_{pop}^2 \times cv^2)}
 \end{aligned}$$

This can be re-written with $cv_{pop} = cv_2$ and $cv = cv_1$ as:

$$cv_{total} = \sqrt{(cv_1^2 + cv_2^2 + cv_1^2 \times cv_2^2)}$$

This derivation applies to the combination of any two gamma distributions, and can be extended to multiple gamma distributions by substituting cv_2 for $\sqrt{(cv_2^2 + cv_3^2 + cv_2^2 \times cv_3^2)}$ etc. It should be noted that this equation was derived for the specific case of combining two gamma distributions, and despite using a measure of variation that can easily be calculated for many distributions is strictly applicable to the gamma distribution alone. However, the calculated value should also be approximately correct for a distribution with similar shape, such as the lognormal distribution with high mean and relatively small variance.