Experimental Design and Mathematical Modelling Methods for the Study of Anthelmintic Resistance in UK Livestock Populations

Johnathan W. Love
Doctor of Philosophy in Statistics
Department of Mathematics & Statistics
University of Strathclyde
Glasgow, U.K.
July 2018

For Agnes, David, Hayley, Mary & in memory of George

This thesis is submitted to the University of Strathclyde for the degree of Doctor of Philosophy in the Faculty of Science.

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Johnathan William Love

Date: 25 June 2018

Acknowledgements

The author would like to thank the Engineering and Physical Sciences Research Council and the Veterinary Medicines Directorate for their funding towards this project. I am also grateful to my supervisors, Dr. Louise Kelly and Prof. Chris Robertson, and collaborators, Dr. Hannah Lester, Mr. Ian Nanjiani and Prof. Mike Taylor, for all of their support and encouragement during this project. I would also like to thank personnel at Westpoint Farm Vets and Mr. David Burden for their efforts and for the provision and recording of the field study data used in this project. I would also like to thank all of the academic staff at the University of Strathclyde for their support and, in particular, Dr. Stephen Corson, Dr. Kimberley Kavanagh, Dr. Alastair Rushworth, Mr. Ian Thurlbeck and Dr. David Young.

Abstract

The Faecal Egg Count Reduction Test (FECRT) is the most widely used field-based method for estimating anthelmintic efficacy and as an indicator of anthelmintic resistant nematodes in cattle. In this thesis, statistical aspects such as the analysis of cattle faecal egg count (FEC) data and the identification of statistically robust experimental study designs for the FECRT, are examined.

Using field study cattle FECRT data, the validity of current guidelines on parameter estimates for evaluating percentage estimates and confidence intervals (CIs), were assessed. For FECs obtained using sensitive counting techniques, percentage estimates are recommended to be evaluated using arithmetic group means. For FECs obtained by less sensitive counting techniques, the maximum likelihood estimator of zero inflated distributions is recommended when evaluating percentage estimates. It would not be recommended however, to use CIs that assume FECs to be normal, and it is therefore recommended that relevant intervals for percentage estimates be obtained using a Bootstrap or Bayesian framework.

A simulation study was conducted using Bootstrap methodology to assess the coverage probability of 95% percentile intervals, associated with different percentage estimates. The coverage was considered for scenarios involving various diagnostic sensitivities, treatment group sizes and classifications of pre-treatment group means. Very few scenarios consisted of 95% Bootstrapped percentile intervals with adequate coverage probabilities. A further simulation study was then carried out with Bayesian methodologies being employed. The accuracy of percentage estimates was examined under the scenarios described above. In the majority of scenarios: in order to obtain the most accurate percentage estimates, one would only need to adopt a paired study design involving a positive treatment group.

The following *R Shiny* prototype webpage application is available to carry out the recommended analysis of cattle FECRT data, i.e. averaging over individual-based egg count percentage reductions/changes based on the form of the *Symmetrised Percentage Change*, using our developed Bayesian methodologies: `http://outreach.mathstat.strath.ac.uk/apps/FECRT`.

Publications, presentations and software

Written publications:

- Love, J.W., L.A. Kelly, H.E. Lester, I. Nanjiani, M.A. Taylor and C. Robertson (2017). Investigating anthelmintic efficacy against gastrointestinal nematodes in cattle by considering appropriate probability distributions for faecal egg count data. *International Journal for Parasitology: Drugs and Drug Resistance*. **7**, 1-12.

- Love, J.W. and M.A. Taylor (2018). Testing for Anthelmintic Resistance in Cattle. *Vet Times. Manuscript submitted for publication.*

Oral presentations:

- Love, J.W., L.A. Kelly, H.E. Lester, I. Nanjiani, M.A. Taylor and C. Robertson (2015). Statistical exploration of Faecal Egg Count Reduction Test methods when investigating anthelmintic efficacy in cattle. In *Proceedings of the Association for Veterinary Teaching and Research Work.*

- Love, J.W., L.A. Kelly, H.E. Lester, I. Nanjiani, M.A. Taylor and C. Robertson (2016). Probability distributions of faecal egg count data and their impact on investigating anthelmintic efficacy. In *Proceedings of the British Association for Veterinary Parasitology* (Awarded the *Best Student Presentation Award*).

- Love, J.W., L.A. Kelly, I. Nanjiani, M.A. Taylor and C. Robertson (2017). Statistical methodologies and experimental designs to aid investigating anthelmintic efficacy in cattle livestock. In *Proceedings of the Royal Statistical Society 2017 International Conference* (Awarded the *Best Rapid Fire Talk Award*).

Software created as part of this project:

- Love, J.W. (2018). *Bayesian FECRT Calculator* prototype webpage application. Freely available to use from `http://outreach.mathstat.strath.ac.uk/apps/FECRT`.

Contents

1 Introduction **1**

1.1 Parasites: a brief introduction 1

1.2 Nematodes (helminthic parasites) 3

1.3 Supporting organisations of the cattle farming industry 8

1.4 Anthelmintics used for controlling helminth populations in cattle . 9

1.4.1 Alternative approaches to using anthelmintics 11

1.4.2 Anthelmintic efficacy and resistance 13

1.5 Faecal egg counts (FECs) 15

1.5.1 Faecal egg counting techniques 17

1.5.2 Faecal Egg Count Reduction Test (FECRT) 22

1.6 Overview of the field study and data 25

1.6.1 Data used as part of this PhD project 30

1.7 Aims and objectives of this project 31

2 Review of current statistical methodologies used in studies of anthelmintic efficacy **33**

2.1 Introduction 33

2.2 Experimental design considerations for the FECRT 33

2.3 Percentage estimates for efficacy 37

2.3.1 Central tendencies for representing egg count data 40
2.3.2 Maximum likelihood estimation (MLE) as a means of estimating central tendencies 42
2.4 Statistical frameworks used to derive point and interval estimates of data 43
2.4.1 Asymptotic approximation for obtaining point and interval estimates 44
2.4.2 Bootstrapping theory 47
2.4.3 The Bayesian approach 52
2.5 Obtaining intervals for estimates within the different statistical frameworks 57
2.5.1 Confidence intervals using asymptotic approximation . . . 57
2.5.2 Involving the Bootstrap 67
2.5.3 Credible intervals 72
2.5.4 Frequentism vs. Bayesianism: differences in approaches and interpretations 73
2.6 Determining resistance with the FECRT 74
2.7 Discussion 75

3 Probabilistic distributions to represent cattle FECs 77

3.1 Introduction 77
3.2 Validity of normality assumption for confidence intervals 79
3.2.1 Data transformations 79
3.2.2 Null hypothesis significance testing 80
3.2.3 Assessing normality of original and transformed FECs . . . 81
3.2.4 Results 82

3.3 Investigating distributions to be used for representing cattle FEC data . . . 83
3.3.1 Issues with discrete count data . . . 83
3.3.2 Compound distributions . . . 84
3.3.3 Fitting distributions to count data . . . 88
3.3.4 Distribution selection methods . . . 90
3.3.5 Results . . . 96
3.4 Comparing FECRT calculations: location parameters from best fitting distributions vs. arithmetic means . . . 101
3.5 Results . . . 102
3.6 Discussion . . . 106

4 Identifying a robust design of experiment via a simulation study involving Bootstrap methodology 110

4.1 Introduction . . . 110
4.2 Criteria currently used to utilise appropriate percentage estimates in anthelmintic studies: confidence intervals . . . 111
4.3 Variability of the measured responses in anthelmintic studies . . . 113
4.3.1 Variability of paired data . . . 113
4.3.2 Investigating correlations of paired FEC data . . . 116
4.3.3 Results . . . 117
4.4 Simulation studies: a general overview . . . 120
4.4.1 Number of simulations to carry out as part of a study . . . 122
4.4.2 Performance measures . . . 123
4.5 Simulating paired data: copula . . . 125
4.5.1 Gaussian copula . . . 127
4.6 Simulation study methodology . . . 129

4.6.1 Results . . . 135
4.7 Sensitivity analysis . . . 139
4.7.1 Results . . . 140
4.8 Discussion . . . 144

5 Identifying a robust design of experiment via a simulation study involving Bayesian methodology 150

5.1 Introduction . . . 150
5.2 Simulation study methodology . . . 151
5.2.1 Prior and likelihood specifications . . . 156
5.3 Results . . . 161
5.4 Prior sensitivity analysis . . . 165
5.4.1 Results . . . 165
5.5 Are there any indistinguishable RMSE values? . . . 172
5.5.1 The confidence interval for the ratio of two population RMSE values . . . 172
5.5.2 A Bootstrapped version of the confidence interval for the ratio of two population RMSE values . . . 174
5.5.3 Bootstrap approach vs. theoretical approach . . . 175
5.5.4 Results . . . 178
5.6 Discussion . . . 186

6 Applying Bayesian robust methodologies to conclude on apparent anthelmintic efficacy and resistance in UK cattle populations 197

6.1 Introduction . . . 197
6.2 Data and methodologies used . . . 198
6.3 Results . . . 202

6.3.1 Classification of treatment groups using developed Bayesian methodologies . . . 202
6.3.2 Agreement between classifications using developed Bayesian methodologies and Defra study methodologies . . . 208
6.4 Discussion . . . 217

7 Discussion, future work and an *R Shiny* prototype webpage application 225

7.1 *Bayesian FECRT Calculator* prototype webpage application . . . 234

A *R*/*RStudio* code used for simulating data 245

A.1 Code for simulating Negative Binomial data . . . 245
A.2 Code for simulating zero inflated data . . . 248

B Information relating to Bayesian simulation study 253

B.1 Generating samples from posterior distributions . . . 253
B.1.1 Gibbs sampler . . . 253
B.1.2 Metropolis-Hastings (MH) sampler . . . 255
B.2 Examples of convergence diagnostics using developed Metropolis-Hastings (MH) algorithms . . . 258
B.2.1 Trace plots and acceptance rates . . . 259
B.2.2 Autocorrelations . . . 267
B.2.3 Gelman-Rubin diagnostic . . . 270
B.3 Additional Concepts: Mean squared errors and the truncated Normal distribution . . . 280
B.3.1 Mean squared errors . . . 280
B.3.2 The truncated Normal distribution . . . 281

B.4 Measures of agreement: The Kappa (κ) and Weighted Kappa (κ_w) Statistics 283

B.4.1 The Kappa statistic (κ) 283

B.4.2 Weighted Kappa (κ_w) 286

Bibliography **288**

List of Tables

1.1 Number of animals recruited onto study and the distribution of farms used by year, adapted from Defra (2015) 25

2.1 Advantages and disadvantages for types of Bootstrapping 50

3.1 *Shapiro-Wilk* Normality test results for Day 0 and Day 14 data and the various transformations applied to these data 82

3.2 Distributions fitted and their parameters 91

3.3 Percentages of the best-fitting distributions for Day 0 data sets, categorised by the four diagnostic sensitivity groups (76 data sets in each grouping) 97

3.4 Percentages of the best-fitting distributions for Day 14 data sets, categorised by the four diagnostic sensitivity groups (76 data sets in each grouping) 98

4.1 Average Correlations of treatment groups' field study data 119

4.2 Weighted Averages of the Average Correlations 120

4.3 Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with pre-treatment group means greater than 150 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%) 136

4.4 Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%) 137

4.5 Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with pre-treatment group means less than 100 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%) 138

4.6 Standardised biases (%) of the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ for pre-treatment group mean classifications and treatment group sample sizes, based on 15EPG_McM_SCFT field study data (highlighted cells have standardised biases between $\pm$40%) 139

4.7 Standardised biases (%) of the percentage estimate $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$ for pre-treatment group mean classifications and treatment group sample sizes, based on 15EPG_McM_SCFT field study data (highlighted cells have standardised biases between $\pm$40%) 139

4.8 Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with pre-treatment group means greater than 150 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%) 141

4.9 Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%) 142

4.10 Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with pre-treatment group means less than 100 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%) 143

5.1 RMSE values of percentage estimates based on a farm with pre-treatment group means greater than 150 epg (highlighted cells have the lowest RMSE values in a given scenario) 162

5.2 RMSE values of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg (highlighted cells have the lowest RMSE values in a given scenario) 163

5.3 RMSE values of percentage estimates based on a farm with pre-treatment group means less than 100 epg (highlighted cells have the lowest RMSE values in a given scenario) 164

5.4 RMSE values of percentage estimates based on a farm with pre-treatment group means greater than 150 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario) 167

5.5 RMSE values of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario) 168

5.6 RMSE values of percentage estimates based on a farm with pre-treatment group means less than 100 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario) 169

5.7 Average standard deviation estimates of FEC data from treatment groups 170

5.8 Average standard deviation estimates associated with individual-based egg count percentage reductions/changes 171

5.9 Ratios of RMSE values obtained using theoretical calculations . . 176

5.10 Ratios of RMSE values obtained using developed Bootstrapping approach 178

5.11 Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with pre-treatment means >150 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one) 181

5.12 Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment means $\leq$ 150 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one) 183

5.13 Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with pre-treatment means <100 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one) 185

5.14 A summary of the most accurate percentage estimates to be estimated for a given diagnostic sensitivity and treatment group sample size 194

5.15 A summary of the most accurate percentage estimates (inclusive of those indistinguishable from % estimates with the lowest RMSE values) for a given scenario 195

6.1 Interpreting agreement using κ 202

6.2 Classifications of treatment groups, based on 15EPG_McM_SCFT data and utilising Bayesian methodologies with percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)$ % (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) 205

6.3 Classifications of treatment groups based on 15EPG_McM data and utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) 208

6.4 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, based on 15EPG_McM_SCFT data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) 211

6.5 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, based on 15EPG_McM_SCFT data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) 212

6.6 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, based on 15EPG_McM data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) . . 215

6.7 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, based on 15EPG_McM data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE)) . . 216

B.1 2x2 Contingency Table of hypothetical classifications between 2 Raters 283

B.2 Interpreting agreement using κ 286

List of Figures

1.1 Direct life-cycle of a Nematode, adapted from Taylor (2010a) . . 4

1.2 Life-cycle of Liver Fluke, adapted from Taylor (2010a) 7

1.3 Faecal egg counting process involving the McMaster slide 19

1.4 An outline of the Faecal Egg Count Reduction Test (FECRT) method study design, adapted from Defra (2015) 28

2.1 Distinction between *Parallel Group* and *Paired Study* Designs used for Anthelmintic Studies 36

3.1 Farm E32 Fenbendazole Day 0 FEC data with example fitted distributions and their associated AIC values. 99

3.2 Farm E32 Fenbendazole Day 14 FEC data with example fitted distributions and their associated AIC values. 100

3.3 Comparison of $100(1 - \frac{T_{14}}{C_{14}})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data. 103

3.4 Comparison of $100(1 - \frac{T_{14}}{T_0})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data. 104

3.5 Comparison of $100(1 - \frac{C_0 T_{14}}{T_0 C_{14}})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data. 105

4.1 Bootstrap Simulation Study Methodology 133

5.1 Bayesian Simulation Study Methodology 155

6.1 2012 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 203

6.2 2013 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 204

6.3 2014 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 204

6.4 2012 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 206

6.5 2013 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 206

6.6 2014 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 207

6.7 2012 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 209

6.8 2013 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 210

6.9 2014 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 210

6.10 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 212

6.11 2012 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 213

6.12 2013 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 214

6.13 2014 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 214

6.14 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green)) 216

7.1 *Bayesian FECRT Calculator* prototype webpage application . . . 235

7.2 Uploading data into the prototype webpage application 237

7.3 Selecting columns of data in drop down windows in prototype webpage application 239

7.4 *Summaries of Submitted Data* tab output in prototype webpage application 241

7.5 *Results of Bayesian Analysis* tab output in prototype webpage application 243

B.1 Trace Plots of parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) 260

B.2 Trace Plots of parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) 262

B.3 Trace Plots of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM_SCFT field study data) 264

B.4 Trace Plots of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM field study data) 266

B.5 Autocorrelation between parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) after burnin period 269

B.6 Autocorrelation between parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) after burnin period and thinning out 269

B.7 Autocorrelation between parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) after burnin period 270

B.8 Autocorrelation between parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) after burnin period and thinning out 270

B.9 Autocorrelation between parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM_SCFT field study data) after burnin period 271

B.10 Autocorrelation between parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM_SCFT field study data) after burnin period and thinning out 271

B.11 Autocorrelation between parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM field study data) after burnin period 272

B.12 Autocorrelation between parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM field study data) after burnin period and thinning out 272

B.13 Potential Scale Reduction Factors of parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) 276

B.14 Potential Scale Reduction Factors of parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) 277

B.15 Potential Scale Reduction Factors of parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM_SCFT field study data) 278

B.16 Potential Scale Reduction Factors of parameters μ and σ using simulated 100 $\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data (based on 15EPG_McM field study data) 279

Chapter 1

Introduction

1.1 Parasites: a brief introduction

Parasites are responsible for considerable morbidity and mortality across animal, plant and human health populations. A parasite is an organism either living within (endoparasitic) or living on (ectoparasitic) a particular host, which provides the parasite with the nutrients it requires (Schmidt and Roberts 1989). Some parasite species live at the expense of other parasites (hyperparasitism). For example, *Plasmodium spp.* live within mosquitos and are the main cause of Malaria in humans. This is also an example of a parasitic species that exhibits a vector life-cycle; one that involves an agent, such as water, wind or an insect to facilitate spread and in some cases, to complete their life-cycle (Lyons 1978). Hosts can be categorised as being definitive, meaning that the parasite is said to have a direct life-cycle which requires the presence of only one type of host in order to complete its life-cycle; or intermediate, meaning the parasite undergoes an indirect life-cycle where the parasite spends part of its life in another host before it parasitises its final host (Paton 1983). If a parasite requires an intermediate host, then generally it will not cause disease in this host as this would reduce the parasite's chances of survival.

When considering parasites, Lyons (1978) explains that they gain and exploit much from their hosts, such as regulated and maintained environments, extensive

food supplies, etc. In fact, parasites evolve in parallel with their hosts, i.e. as the hosts speciate and develop then so do the parasites becoming highly specialised to their way of life (known as co-evolution). Generally, different species of hosts - such as cattle, sheep, goats and horses - have their own species of parasites associated with them.

As well as being host-specific, most parasites are site-specific. This can effect their development and survival. For example, some parasites may develop and reproduce in areas of the body such as the intestines, which would be considered ideal for egg excretion and dispersion to occur, yet some parasites may develop and reproduce in the lungs for instance, which are further located away from where the parasite must excrete their eggs. Some host sites show less reaction to parasites than others (i.e. immunologically privileged sites, where the host is less likely to fight infection) and parasites occupying these types of sites may be relatively safe from host attack. Additionally, a parasite has to be able to resist the host's defence reactions and feed without damaging the host to the extent that it may provoke an immune response - or in some cases cause the death of the host. If the parasite is unable to become established from the first point-of-contact with the host, then the host is said to display innate (i.e. natural) immunity. In contrast, if an initial infection sparks a degree of resistance to subsequent infections then the host is said to have an acquired immunity (Lyons 1978; Schmidt and Roberts 1989; Taylor *et al.* 2007). Also, parasites reproduce in abundance to increase their survival rates and so competition for resources, within and outwith the host, becomes an issue. With this in mind, there is also the issue of the dispersal of parasites across the host populations: a parasite's location within (or out with) the host may effect this. However, the dispersal of parasites is also dependent on time, since certain parasites may only be able to infect their hosts at certain times of the year and so parasites have to synchronise their own life cycles with that of their hosts.

Parasite population sizes are related to the population size and density of the final host involved in their life-cycles. However, parasites are known to not be evenly distributed through a host population: most hosts are uninfected or contain very few parasites and very few are heavily infected (Shaw and Dobson 1995; Levecke *et al.* 2012). The variables affecting parasitic populations are complex; in order

to study such interactions mathematical population models are helpful (Paton 1983; Byrom 1990).

And so, what makes a successful parasite? Many parasitologists, as well as Lyons (1978), agree that it is simply one that is in balance with its host.

1.2 Nematodes (helminthic parasites)

The presence of helminthic parasites has been known for over 3000 years and the study of these parasites has been mainly due to their ever growing economic and animal health, welfare and production importance (Crofton 1966; Vlassoff and McKenna 1994; Corwin 1997; Molento 2009; Voort *et al.* 2013; Charlier *et al.* 2014). There are helminths that exist as free-living or involve free-living stages as part of their life-cycles, meaning that such organisms are not parasitic or involve stages where no parasitism occurs, however these types of helminths are mainly over-shadowed by the economic and social importance of the parasitic kinds. Lyons (1978) clarifies that helminths are classed into three main groups: platyhelminths (flatworms) and nematodes (roundworms) which both consist of free-living and parasitic forms, as well as the acanthocephalans (spiny-headed worms) that are known to be purely parasitic.

Nematodes typically have a cylindrical form, tapered at each end with a protective cuticle (outer-most layer). The basic nematode life-cycle is a direct life-cycle as represented in Figure 1.1 (this Figure considers a bovine host, but the life-cycle is similar for other hosts as well). The life-cycle involves an egg and various larval stages - denoted as L1, L2, L3, L4 and L5 in which L3 is usually the infective larval stage. The eggs are excreted in the faeces of the host and, within the faecal pat, develop into the L2 larval stage and thereafter to the infective third larval stage over a period of a few weeks (depending on ambient temperature). When moist conditions are present, the L3 will then migrate from the faeces onto vegetation and be ingested by the hosts. The entire parasitic phase of the life-cycle will normally take up to three weeks since the ingested larvae undergo a series of parasitic moults until they reach sexual maturity and hence the life-cycle will begin for the new generation of parasites (Crofton 1966; Schmidt and Roberts

1989; Taylor 2010a; COWS 2015b). But under certain circumstances some of the ingested L3 become arrested in their development at the early fourth larval stage for periods of up to six months in certain species; this is referred to as hypobiosis.

Taylor (2010a) and COWS (2015b) describe the epidemiology of parasitic infections as being heavily influenced by the weather and seasonal effects. In fact, the effects of climate change on the seasonality and the spread of nematodes in more recent times are being investigated (Mas-Coma *et al.* 2008; Dijk *et al.* 2009; Morgan and Dijk 2012; Skuce *et al.* 2013). In spite of this, any parasitic eggs present in the spring develop steadily to L3 and this rate of development increases as temperatures increase. Therefore, eggs deposited during April-June all reach the infective larval stage from around July, where peak numbers in worm burden occur. Progressing into autumn, with the decline of temperatures, an increasing proportion of ingested L3 become suspended at the early fourth larval stage. As a result, in the late autumn calves can play host to many of these early fourth larval staged parasites.

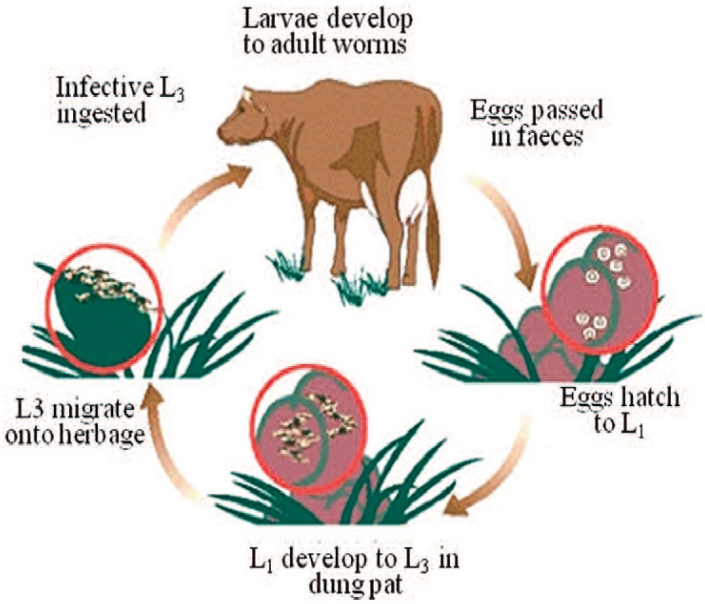

Figure 1.1: Direct life-cycle of a Nematode, adapted from Taylor (2010a)

There are many different classes of experts who study nematodes: those concerned

with the infections of humans are known as medical parasitologists, veterinary helminthologists are those concerned with nematodes infecting animals and nematologists study these kinds of infections with regards to plants. In fact, there are some organisations and reports that suggest the use of helminths as treatments (*worm therapies*) and promote their usefulness in society (R. Nuwer 2013; Worm Therapy 2018). For example, M. Jackson 2004 tells us that leeches (an example of an ectoparasite) are used in the human health industry to aid blood circulation and for healing skin grafts.

Initially, parasites of medical significance were the first to be noticed; the first written records of these types of parasites dates back to the Papyrus Ebers (circa 1500 B.C.). It was not until the 16th and 17th centuries that new anthelmintics (i.e. agents that destroy/expel parasitic worms) were introduced and, prior to this period, nematodes were considered as the carriers of diseases - rather than the infecting organisms that we think of today (Crofton 1966).

Gastro-intestinal nematodes continue to be one of the main sources of economic constraint in Great Britain (GB), and other areas of the world, with respect to the production of domesticated animals and ruminants (Byrom 1990; Falzon *et al.* 2013; Playford *et al.* 2014). For example, the *Moredun Research Institute* (Moredun) estimated that the cost of these helminths to the British sheep industry is estimated to be £84 million per annum (Moredun 2018). Helminth species belonging to the taxonomic order *Strongylida* are, in particular, of economic and medical importance.

Nematodes belonging to the *Trichostrongylidae* family are of particular interest; in spite of there being many genera and an immense number of species in this family (Schmidt and Roberts 1989). *Ostertagia spp.* are a small intestinal species known as the *brown stomach* worms and are the most economically important in comparison to others. *Teladorsagia (Ostertagia) circumcincta* are prevalent in sheep and are a major cause of intestinal infections according to Paton (1983) and Abbott *et al.* (2009), whereas *Ostertagia ostertagi* are normally found in cattle. In fact, cattle can play host to over 18 species of gastrointestinal nematodes and the disease caused by these organisms is commonly referred to as parasitic gastroenteritis (Taylor 2010a).

Another species of nematode found in GB is *Cooperia oncophora* (Taylor *et al.* 2007; Taylor 2010a), which is considered to be a mild pathogen in calves and is commonly found in young cattle during their first grazing season. It is also the *dose-limiting species* of nematodes for cattle, meaning this species identifies the dosage of anthelmintics for which 90% of efficacy occurs (Vercruysse and Rew 2002; Taylor 2012). In fact, *C. oncophora* and *O. ostertagi* can infect young cattle concurrently, in which case the impact of infection can be much higher in comparison to the impact of individual infections from each species (COWS 2015b). Other species such as *C. punctata* and *C. pectinata* are of interest even though they are less common in GB - but are still considered to be pathogenic (i.e. able to cause diseases). Worms in this genus are also considered to be the main contributor to parasitic, faecal egg counts in cattle.

Moreover, the species *Haemonchus contortus* is also found in cattle and lives in the abomasum (the fourth division of the stomach of cattle) and feeds by sucking blood. This species of nematode is known as the *Barber's pole* worm and with heavy infections, cattle will generally die. As a result from the loss of blood that this nematode can cause, cattle can suffer from anaemia and intestinal problems. Surviving cattle develop an acquired immunity. Abbott *et al.* (2009) also states that this species is widespread in sheep, in GB.

Nematodirus spp. is another genus associated with domestic livestock: *Nematodirus helvetianus* are associated with cattle and *Nematodirus battus* are associated with sheep. These species can be found worldwide (Hoberg *et al.* 1986; Bogale *et al.* 2014) but is not the most harmful of the nematodes mentioned.

Dictyocaulus vivaparus (more commonly known as lungworm or husk) is another genus found in GB, and is considered highly pathogenic. The number of recorded cases of this parasite has been on the rise in recent times according to the, formerly known as, Animal Health and Veterinary Laboratories Agency (AHVLA) and is of economic importance (AHVLA 2012, COWS 2015a). Adult worms of this species live in the animal's lungs where they produce eggs that hatch quickly. The L1 then travel up the windpipe, are swallowed and excreted in the faeces. These then mature onto pasture to the L3 infective stage and are ingested - in which case they mature to adults in the lungs. However, outbreaks of lungworm

disease are quite unpredictable; even though this parasite is found widespread across cattle herds in GB (COWS 2015a).

Fasciola hepatica (liver fluke) is a trematode of major economic significance. In fact, it is estimated that this trematode costs the United Kingdom (UK) agriculture sector £300m per year (COWS 2013). Prevalence of this infection is on the increase due to a variety of reasons such as climate change. The main difference when considering flukes is that their life-cycles depend on an intermediate host. In the case of liver flukes, this type of trematode depends on the *Galba truncatula mud snail* as shown in Figure 1.2. High risk conditions that influence the infection rates of these trematodes include warm weather and high rainfall during the summer seasons and the location of the mud snail, i.e. wet muddy areas, that could be found near or on farm.

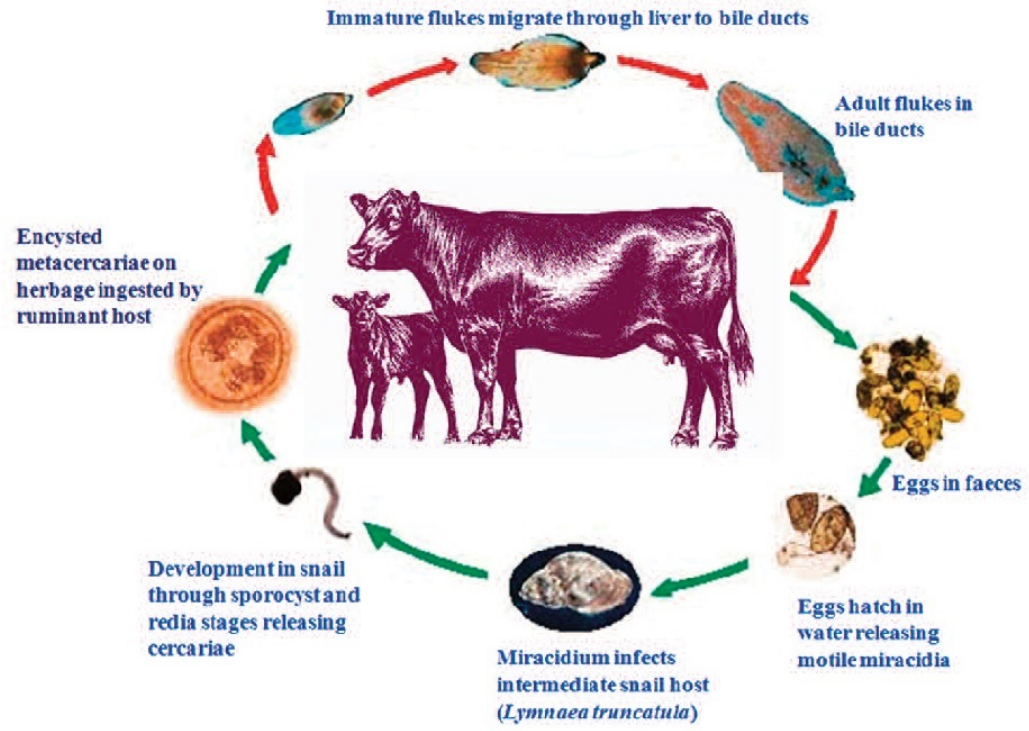

Figure 1.2: Life-cycle of Liver Fluke, adapted from Taylor (2010a)

This parasite can prove to be very harmful to the definitive host as they have the ability to cause liver damage and the destruction of tissue due to the nature

of their migration technique, once ingested by the host, and generally those with heavy infections will die.

Rumen fluke are another group of trematodes that have been found increasingly in British and Irish livestock (Taylor 2010a). Generally, rumen fluke are found as co-infections with liver fluke since they both rely on the same intermediate host.

1.3 Supporting organisations of the cattle farming industry

Government organisations are responsible for the health, welfare and associated policies related to cattle (and other animal species) - such as the Department for Environment, Food and Rural affairs (Defra) in the UK. The Veterinary Medicines Directorate (VMD) and the Animal and Plant Health agency (APHA), formerly known as the Animal Health and Veterinary Laboratories Agency (AHVLA) (APHA 2018; VMD 2018) are part of Defra and have key responsibilities for animal health and veterinary medicine. Defra invests in the rural economy, supports and develops British farming, encourages environmentally sustainable and healthy food production and aims to improve the standards of animal welfare (Defra 2018a).

The two main industries concerned with cattle are the dairy and beef industries. For the latter, the Agriculture and Horticulture Development Board (AHDB) Beef and Lamb is an organisation that exists, on behalf of levy payers, to enhance the profitability and sustainability of the English beef and lamb industry. Its main aims are to aid the beef and sheep meat supply chain to become more efficient, and to also make the industry more profitable. Additionally, AHDB Beef and Lamb carries out many research projects in animal health and welfare (AHDB Beef and Lamb 2018). Similarly, Hybu Cig Cymru (Welsh Red Meat Levy), Livestock and Meat Commission for Northern Ireland and Quality Meat Scotland are the cattle beef and lamb levy boards for Wales, Northern Ireland, and Scotland respectively (Hybu Cig Cymru 2018; Livestock and Meat Commission 2018; Quality Meat Scotland 2018).

Equivalently, the AHDB Dairy exists for British dairy farmers, funded by levy payers, in order to improve the sustainability of the British dairy industry by providing evidence-based information on issues such as animal health and welfare (AHDB Dairy 2018). These levy boards are also part of the AHDB, which is an impartial board to make these industries competitive and sustainable through evidence-based advice, information and activities (AHDB 2018).

With respect to parasitism, in 2010 the Control of Worms Sustainably (COWS) initiative was set up in order to provide evidence-based information to the cattle industry in relation to the sustainable control of parasites in cattle herds (COWS 2018). There are several other equivalent forms of these initiatives for other livestock species - such as Sustainable Control of Parasites in Sheep (SCOPS 2018).

There are also organisations that exist in order to ensure the safeguarding of medicines and vaccinations for cattle. The Responsible Use of Medicines in Agriculture Alliance (RUMA) is a growing coalition of organisations, that has been set up to review and provide guidance on the use of medicines in all livestock (RUMA 2018). In fact, they have established practical strategies to promote the correct use of vaccines in the dairy and beef cattle industries and also highlight the responsibilities of the dairy and beef farmers with respect to the health, welfare and productivity of their herds (RUMA 2007). Further, the National Office of Animal Health (NOAH) represents the UK animal medicines industry and aims to promote the benefits of safe, effective use and quality of medicines for animals (NOAH 2018a).

In summation, there are many organisations that consider the welfare of cattle to be of high significance and will continue to do so for the foreseeable future.

1.4 Anthelmintics used for controlling helminth populations in cattle

By definition, *Cattle* is the generic term that describes the many different classifications of the domesticated bovine species. For male members of the species,

they can be classed as bulls (non-castrated) or steers (castrated). In addition, the term oxen is given to those castrated males that are kept for draft purposes, i.e. an animal trained to perform tasks. For the female members of the species, a young female that has not given birth to a calf (a term given to young cattle of either sex) and is also under three years of age is known as a heifer. A heifer, after giving birth, is referred to as a cow.

As mentioned in Section 1.2, anthelmintics (also known as wormers) are chemical agents used for the control of helminths. Many products are available in GB, and worldwide, and for cattle most are marketed for both treatment and prevention of helminthoses. Broad-spectrum anthelmintics that are licensed for use in cattle are discussed by Taylor (2010a) and COWS (2014), and can be categorised into one of three classes: *benzimidazoles, imidazothiazoles* and *macrocyclic lactones*. These classes of anthelmintics are referred to as the BZ, LV and ML groups respectively.

The BZ group of anthelmintics were first made available on the cattle anthelmintic market, with their discovery being made in the 1960s (Yadav and Singh 2011). Following from this, much progress was made in producing LV anthelmintics (McKellar and Jackson 2004). Afterwards, in the 1980s, ML anthelmintics were made available for use in cattle and now dominate the anthelmintic market (Campbell *et al.* 1983; Vercruysse and Rew 2002; Omura 2008; Taylor 2010a).

According to COWS (2014), anthelmintics can be administered in many ways (on the basis of an animals weight): drench (the animal swallows the anthelmintic), injection or by topical application (pour-on). Anthelmintics can also be delivered through means of a bolus device; where a relatively large quantity of anthelmintic is swallowed (with use of a specially designed dosing gun) and sits in the forestomach and is delivered prophylactically. Boluses can be categorised in two ways: either the bolus releases the drug constantly over a period of time (sustained release); or the drug is released at certain time intervals (pulse released).

Taylor (2010a) tell us that the BZ class of anthelmintics (also known as white drenches) are effective against nematodes and their eggs, i.e. they are also ovicidal, and are often efficacious against trematodes (COWS 2014). After receiving the anthelmintic, the BZ anthelmintic passes into the rumen (the first division of

the stomach) of the cattle and this acts as a reservoir allowing gradual release of the drug.

The LV class of anthelmintics are commonly known as yellow drenches (SCOPS 2016). According to Taylor (2010a) and Yadav and Singh (2011), LV anthelmintics paralyse the worms leading to their expulsion. The drugs that are found in this class of anthelmintics are rapidly absorbed and excreted within 24 hours. As a result, it is not essential to maintain high concentrations of these types of drugs for prolonged periods of time and it is worth noting that this class of anthelmintics are not ovicidal.

ML anthelmintics (known as clear drenches) are regarded as the most popular among the three classes of anthelmintics mentioned (Vercruysse and Rew 2002; Omura 2008; Taylor 2010a). After administration the compounds for this class of anthelmintic are stored in fat tissue from where they are then slowly released. Due to popularity, ML anthelmintics are available in a wide range of forms, including injectable and pour-on formulations.

Not all anthelmintics are efficacious against hypobiotic fourth-stage larvae. According to Taylor (2010a), at recommended dose rates, BZ and ML classes of anthelmintics are more active against hypobiotic larvae and are also ovicidal compared to the LV class of anthelmintics, resulting in BZ and ML classes being more widely used (Levecke *et al.* 2012).

Information on recommended dose rates and anthelmintics can be found courtesy of NOAH, which provides a compendium of data sheets for animal medicines for veterinarians (NOAH 2018b). As well as this, SCOPS provides information on different classes of anthelmintics mentioned (SCOPS 2016) and COWS provides information on anthelmintics specifically used for cattle (COWS 2014).

1.4.1 Alternative approaches to using anthelmintics

According to Stear *et al.* (2007), control measures for parasitic, nematode infections rely on anthelmintic treatments; but are threatened by the development of resistance in parasite populations. Resistance is defined to be the genetically-transmitted ability of a parasite to tolerate a normally effective dose of an an-

thelmintic (Bishop and Stear 2003; Taylor 2010a).

Alternative methods for controlling worms have been suggested without relying on efficacious anthelmintics. One of these is grazing management schemes (Waller 2006; Stear *et al.* 2007). The objective of these schemes involves maximising the use of available pasture for livestock grazing and the level of parasites *in-refugia*, i.e. the non-parasitic stages of the nematode life-cycle, whilst at the same time minimising the number of infective L3 larvae that have the potential to infect grazing animals.

For instance, the number of animals on a particular area of pasture for a specific period of time (i.e. stocking density) could be considered, to the extent where farmers reduce the number of animals in a given area. As a result, fewer animals being infected with egg-laying nematodes would occur and could lead to lower levels of pasture contamination (Stromberg 1997). In addition, farmers could also have different livestock species present on their pastures to reduce contamination, since most nematodes are host-specific; though, mixed infections are the most pathogenic according to Bishop and Stear (2003). There is also the option of rotational grazing, where pasture is divided into sections that are sequentially grazed on. The logic here is that in the absence of a grazing host in certain sections; infective larvae cannot parasitise and would eventually die by the time animals return to a section of pasture. However, the length of life-cycles and weather conditions would have to be considered in order for this to be a viable option for a grazing management scheme (Callinan *et al.* 1982).

Outwith grazing management schemes, there is a promising option of using biological control methods, such as introducing predators of parasitic nematodes to pastures. For example, the nematode-trapping fungus *Duddingtomia flagrans*, has the ability to reduce the number of infective L3 on pasture and reduce the intensity of infection (Wolstrup *et al.* 1994; Waller *et al.* 2004; Waller 2006; Stear *et al.* 2007).

There are also various ways in which to control nematode infections by enhancing the immune response of the host through nutritional supplementation, by developing new vaccinations for cattle and through the utilisation of genetic variation in hosts to become more resistant towards nematodes. However, Stear *et al.*

(2007) also explains that not one control method could be recommended to the extent of excluding another; as well as no single method being recommended to all types of farms. As a result, there could be scope for a potential *combination* of the control methods to be explored and implemented in the future.

1.4.2 Anthelmintic efficacy and resistance

Anthelmintic resistance is defined to be the heritable ability of a parasite to tolerate a normally effective dose of an anthelmintic (Bishop and Stear 2003; Taylor 2010a). Parasites are considered resistant if they survive exposure to the standard, recommended dose of the anthelmintic and this ability to survive is passed on to their offspring. According to Taylor (2010a) and Gilleard and Beech (2007), anthelmintic resistance is now accepted as a pre-adaptive phenomenon, in that the alleles (different forms of a gene at a specific position on a chromosome) that give rise to resistance already exist in the worm population, before they have even been exposed to the anthelmintic. On the other hand, natural selection keeps the resistance alleles at a very low frequency - in the absence of anthelmintics - since these alleles make the helminthic parasites carrying them less fit for survival than the susceptible (those that can be affected by the anthelmintic). However in the case of when an anthelmintic is continually used, resistant nematodes gain a survival advantage as they reproduce at higher rates (in comparison with the susceptible nematodes present) and their numbers increase.

Worldwide, the numbers of cattle herds thought to have been exposed to anthelmintic resistant helminths are not as alarming as the numbers for sheep flocks (Sangster 1999; Kaplan 2004; Wolstenholme *et al.* 2004; Waller 1997; COWS 2015b) though resistance has been widely reported in Australia, New Zealand, parts of Europe and in some parts of the United States of America (Waghorn *et al.* 2006; Demeler *et al.* 2009; El-Abdellati *et al.* 2010; Edmonds *et al.* 2010; Sutherland and Leathwick 2011). Although there have been no widespread reports of resistant helminths in cattle in the UK, sporadic cases have been reported in the dose-limiting species, *C. oncophora* (Stafford and Coles 1999; Sargison *et al.* 2009). Indeed, the true representation of resistance is difficult to assess mainly due to inconsistencies in treatment dose administrations, faecal sample collection

and handling methods, faecal egg counting techniques used, associated experimental designs (Taylor 2012) and the lack of robust methods for determining anthelmintic resistance under field conditions i.e. the lack of field data supported by controlled slaughter studies, or the availability of validated molecular and in-vitro methods for cattle nematodes

Efficacy can be defined as a quantitative measure of the effectiveness of a drug intended to produce a desired effect (Vidyashankar *et al.* 2012). A fully effective anthelmintic is expected to reduce faecal egg counts (FECs), i.e. the measured response of anthelmintic studies, to zero after administration of the anthelmintic. The most reliable method for determining anthelmintic efficacy is the controlled anthelmintic efficacy test, whereby animals are artificially infected, treated, then slaughtered and worm burden counts performed (Powers *et al.* 1982), but these are not practicable in the field. It is common to assume that any apparent lack of efficacy is due to anthelmintic resistance – but this apparent resistance can be the result of anthelmintic failure due to other factors, most commonly under-dosing due to inaccurate estimation of bodyweight (Taylor *et al.* 2002).

The most common method used to investigate anthelmintic resistance is the Faecal Egg Count Reduction Test (FECRT) (Coles *et al.* 1992; Coles *et al.* 2006) which will be further explained in Section 1.5.2. However, this test has not been validated against slaughter studies and the European Medicines Agency (EMA) regards this test as an estimation of efficacy, and not confirmation of resistance (EMA 2014). True resistance must be confirmed through laboratory slaughter studies, potentially supported by molecular level studies, or methods such as egg hatch tests (Vidyashankar *et al.* 2012).

There are many factors that can influence the anthelmintic resistance status of nematodes. For example the size of *in-refugia* populations should be considered, as these stages are not exposed to any anthelmintic treatment (Wak 2001). In general, the larger the *in-refugia* population is, in comparison to the anthelmintic-exposed population; the more slowly resistance will develop since unexposed populations are not affected by the anthelmintics.

The more frequent anthelmintic treatments are given to cattle, the faster anthelmintic resistance is likely to develop in nematode populations. Again, Taylor

(2010a) explains that the underlying principle for anthelmintic resistance is due to treatments giving the resistant nematodes a survival advantage over the susceptible nematodes, since resistant worms have a longer period of time to reproduce. As a result, the number of resistant nematodes increases along with the number of susceptible worms decreasing and so a high pasture contamination of resistant worms occurs. In fact, if the pasture is highly infective, and the cattle are highly susceptible, re-infection occurs quickly and selection for resistance is minimised, due to susceptible worms being re-established, explains Taylor (2010a).

Anthelmintic dosing rates can also affect the resistance status of helminths in cattle herds, since low dose rates are likely to result in a low number of responses of the desired effect of the anthelmintic, and add to the selection pressure for anthelmintic resistance when resistance is in the early stages of development. Under-dosing of cattle with anthelmintics occurs frequently due to farmers under-estimating weights, faulty dosing equipment, misleading dose calculations or adverse weather conditions effecting treatments at the time of administration.

Other resistance selection pressures can stem from previous parasitic exposures and concurrent diseases. However there are some methods, such as rotation of anthelmintics that are used to delay the onset of resistance in helminths. According to Taylor (2010a), it was considered in the past that the gradual rotation of anthelmintics of different classes should be adopted by farmers - particularly for sheep worm control practices. This has however proved to be more challenging carry out in cattle herds, mainly due to the widespread use of ML anthelmintics on UK cattle farms. If this method can be adopted, the emphasis shifts to using a different class of anthelmintic - as it is natural selection that would be expected to reduce the prevalence of resistant nematode populations.

1.5 Faecal egg counts (FECs)

Counting eggs in faeces gives an indirect measure of worm burden present in cattle herds (and other livestock) since experimental studies have shown there is a weak, positive correlation between FEC data and actual worm burden (Eysker and Ploeger 2000). These counts, usually reported as the number of worm eggs

per gram (epg) of faeces, can aid in the decision-making process regarding whether or not anthelmintic treatments are necessary to use, to be safely delayed in using or not to be used for cattle herds. Monitoring FECs (i.e. FEC screenings) can also direct the use of anthelmintics to be better timed and more efficient. In cases where anthelmintics are used excessively, FEC monitoring can provide farmers with information to reduce the use of anthelmintics, in order to decrease the selection pressure to resistance (Taylor 2010a). This is known as targeted strategic treatment.

Counting parasitic eggs, does however, present many interpretational issues. For instance, a high parasitic egg count in faeces may be regarded as an indication of a high worm burden, but takes no consideration of the fact that species of nematode vary in their fecundity and pathogenicity. Also, faecal egg production varies throughout the year and is greatly influenced by a number of factors including levels of parasite challenge (which in turn is influenced by seasonal weather patterns) and the development of protective immunity. Also, FECs cannot distinguish between certain species of cattle nematode, whose eggs look very similar morphologically. As a result, according to Taylor (2010a) larval culture and differentiation (hatching the eggs found in a sample of faeces and identifying the larvae) are often carried out in order to find out whether worms of one particular species dominate FECs or not. Normally, 50 or 100 larvae are counted and the proportion of each species is reported. These results are best used as a general indication of the species' present, rather than an exact determination of the proportion of the FECs contributed by each species.

On the other hand, a low parasitic egg count cannot be associated with a low worm burden since low egg counts do not take account of parasites being immature or in a hypobiotic state. Furthermore, as cattle grow older they can develop an acquired immunity that will effectively reduce nematode fecundity and this results in egg counts being an unreliable indicator of the magnitude of worm burdens present (COWS 2015b). The only way of estimating worm burden to a higher degree of accuracy than what egg counts can provide is through a post-mortem examination of cattle (i.e. slaughter trials).

1.5.1 Faecal egg counting techniques

Egg counts can be determined by a variety of methods. The methods most commonly used worldwide are those counting processes involving the *McMaster* technique. There have been many modifications of this technique since it was first established (Gordon and Whitlock 1939; Whitlock 1948) and these offer different egg detection limits, i.e. diagnostic sensitivities typically ranging from 15 to 100 epg according to the Ministry of Agriculture, Fisheries and Food (MAFF) (MAFF 1986).

Most counting processes, where the McMaster slide is present (Figure 1.3), involve weighing out Fgrams of faeces, say, into a test tube and adding water to the faeces (this is known as the dilution stage). Afterwards, the test tube is shaken to break up the faecal matter. The solution is screened through a sieve, and the resultant filtrate centrifuged. The supernatent is discarded and a volume of Vml salt solution is added to the tube so as to break up the pellet (this is regarded as the flotation stage). Using the resultant solution, the McMaster slide is filled and examined microscopically to count the number of eggs present under each grid, or chamber. The distribution of eggs would be expected to follow a Poisson distribution, assuming eggs are well mixed within the faecal sample (Denwood 2010).

Variations of this method involve differences between weights of the faeces used, flotation times and the specific gravity of the solutions involved (these solutions are used to break up the composition of the eggs and faeces in the pellet to ensure eggs are well mixed within the faecal sample), presense/absence of a centrifugation stage (this removes fine particles and colouring and eases the identification of eggs) and the number of McMaster counting chambers and grids that are used in the counting process (MAFF 1986; Taylor 2010b; Lester and Matthews 2013).

The McMaster slide features two chambers and grids, and the volume of faecal suspension used is determined by the McMaster slide itself, Cml say, by counting different partitions of the chambers and grids of the slide. Since both the ruled grids and each chamber are precisely measured, eggs can be examined under one or both grids, under one chamber or under the total area of both chambers. If one

grid is used, then C takes the value of 0.15ml but if two grids are used then the area used to count eggs is doubled and so C takes the value of 0.3ml. Similarly, if one chamber is used then C takes the value of 0.5ml (i.e. we increase the area to be examined and should be able to identify more eggs in comparison with the 0.15ml when using only one grid) but if two chambers are used as part of the counting process then, similarly, C doubles in value to 1.0ml. These potential values of C contribute to determining the *limit of detection* used as part of the counting process (MAFF 1986; Taylor 2010b).

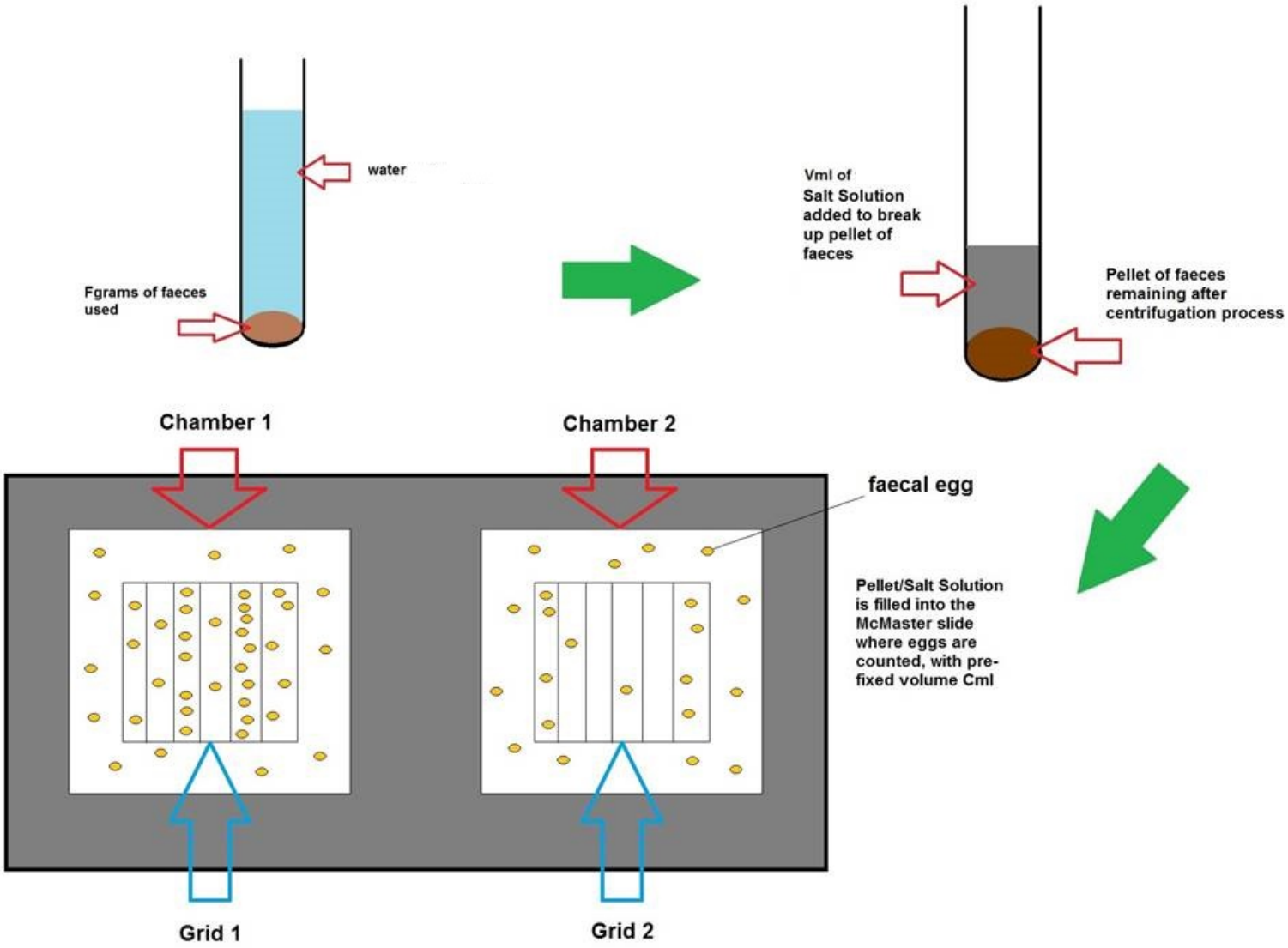

Figure 1.3: Faecal egg counting process involving the McMaster slide

The limit of detection (i.e. diagnostic sensitivity) is essentially a multiplication factor; where the number of eggs present on the McMaster slide is multiplied by this value in order to represent the number of eggs per gram of faeces that have been sampled. This diagnostic sensitivity is calculated as:

$$\frac{V}{FC} \text{ eggs per gram} \tag{1.1}$$

where F, V and C are defined as before. For example, if one were to use 10ml water and 2g of faeces to obtain the faecal suspension and used two chambers on the McMaster slide when counting faecal eggs (so C would equal the value of 1ml), then we would derive a limit of detection of 5 epg using equation (1.1).

According to Lester and Matthews (2013), if the egg detection limit is high (for example 30 epg, 50 epg, etc.) then the counting technique will not be particularly sensitive to changes in egg abundance below or around this limit and so zero eggs seemingly found may not necessarily correspond to no eggs being present; this is more likely to mean that the counting technique is not sufficiently sensitive to be able to detect any eggs that could actually be present. As a result, *false/excess* zeros could be obtained. Hence, a lower-valued limit of detection corresponds to an increased precision in counting faecal eggs and this can be achieved by decreasing the volume Vml (the dilution) or by increasing the grams of faeces used, Fgrams, or by increasing the value of Cml - for example, using more chambers and grids on the McMaster slide to count eggs (Cringoli *et al.* 2004). However, Morgan *et al.* (2005) and Lester and Matthews (2013) tell us that increasing the precision is counteracted by the increased effort, e.g. from trained operators counting the eggs, associated costs of the slides and by the diminishing returns, if more than four chambers are counted.

The standard McMaster technique in the veterinary and parasitology communities uses a limit of detection of 50 epg and this is recommended for featuring in counting processes by published guidelines (Coles *et al.* 1992; Coles *et al.* 2006). However, Coles *et al.* (2006) tell us that an agreed standard method for counting faecal eggs is warranted, as there are several methods and variations on

the standard McMaster technique that result in different diagnostic sensitivities being achieved such as 30 epg, 15 epg and 10 epg (MAFF 1986; Taylor 2010b). For example, the *Parasep* system is also a helminth egg filtration kit that uses the McMaster counting chamber and can produce sensitivities such as 10 epg.

It is also possible to obtain a 1 epg diagnostic sensitivity through means of a *Sensitive Centrifugal Flotation* technique (MAFF 1986). Using a technique with a 1 epg limit of detection means that any eggs that are not detected are indeed not present in the used faecal sample and this is highly suitable for anthelmintic studies when low egg counts may be present in faeces obtained. In fact, some research institutes have invested much time into developing techniques that are accurate in detecting not just the high number of egg counts that may be present; but also possible low numbers (Christie and Jackson 1982; Lester and Matthews 2013).

Another counting method for monitoring faecal eggs is the *FLOTAC* technique, which involves the FLOTAC apparatus that has been designed to carry out flotation in a centrifuge and carries a limit of detection of 1 epg (Cringoli 2006) - but specialist equipment and training are required to carry out this technique. However, the suitability of this technique for anthelmintic studies in cattle is questionable, since very few studies involving cattle FEC data have been used as part of on-going validation testing and comparisons with existing counting processes in recent times (Cringoli *et al.* 2010; Rinaldi *et al.* 2010; Rinaldi *et al.* 2011; Levecke *et al.* 2012).

FECPAK is another development with respect to parasitic egg counting. It is a commercial attempt to sell all materials (in a small case) required to count faecal eggs on farm by the farmers themselves or by their farm vet (Techion Group Ltd, Technology in Action 2018). However, McCoy *et al.* (2005) tell us that in a study examining *on-farm* FECPAK and *in-lab* FECPAK results there dramatic differences in the number of eggs counted, with the results indicating that farmers were over-estimating their counts.

Evidently there are many methods available for monitoring and measuring FECs. However, it is worth noting that different diagnostic sensitivities being used when measuring FECs can result in different interpretations with respect to apparent

anthelmintic efficacy, since the higher the multiplication factor used; the higher the number of eggs thought to be present per gram of faeces.

1.5.2 Faecal Egg Count Reduction Test (FECRT)

The *Faecal Egg Count Reduction Test* (FECRT) is the most widely used field-based method for investigating suspected anthelmintic resistance and estimating apparent anthelmintic efficacy (EMA 2014), but has not been validated against slaughter studies in cattle. The test traditionally involves evaluating the arithmetic group means of FECs (i.e. central tendency and maximum likelihood estimator of a Poisson or Negative Binomial distribution) and calculating the percentage reduction in FECs and the corresponding 95% confidence intervals for a treated (i.e. positive treatment) and untreated control (i.e. negative control) group of animals - usually based on the faecal samples collected on Day 14, post-treatment. As a result, the percentage estimate is calculated as:

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%, \tag{1.2}$$

where T_{14} and C_{14} are the arithmetic group sample mean FECs collected on Day 14 (post-treatment), from the treatment and control groups respectively.

The methodology for carrying out a FECRT, described above, is recommended as part of produced guidelines from the World Association for the Advancement in Veterinary Parasitology (WAAVP) (Coles *et al.* 1992; Coles *et al.* 2006) and are widely accepted. In them, WAAVP recommends the use of the *McMaster* technique with a limit of detection of 50 epg but also refers to the use of the McMaster technique where 15 epg can be adopted as the diagnostic sensitivity for the utilised counting techniques.

A negative control group is present in order to allow for natural changes in FECs during the test period. The sample size to be used in each treatment group was originally 10 animals per treatment group according to Coles *et al.* (1992), but these group sizes were updated to 15 animals based on a review of the guidelines

(Coles *et al.* 2006). If individual cattle FECs are initially recorded as a value below 100 epg, then Coles *et al.* (2006) recommend that a faecal egg counting technique with a lower limit of detection be used.

The corresponding 95% confidence interval for the percentage estimate (1.2) is

$$100\left(1-\frac{T_{14}}{C_{14}}\exp\left(\pm 2.048\sqrt{\frac{s^2_{t.eos}}{n_{treat}T^2_{14}}+\frac{s^2_{c.eos}}{n_{control}C^2_{14}}}\right)\right)\% \tag{1.3}$$

where, $s^2_{t.eos}$, n_{treat}, $s^2_{c.eos}$ and $n_{control}$ represent the sample variances and group sizes for the treated and control groups at Day 14, respectively (Coles *et al.* 1992). More on the statistical exploration and derivation of the form of the confidence interval (1.3) are discussed in Section 2.5.1.

According to Coles *et al.* (1992), anthelmintic resistance is confirmed if the percentage estimate value (1.2) is less than 95% and the lower confidence limit of the confidence interval (1.3) is less than 90%. If only one of these criteria are met; anthelmintic resistance is suspected.

These guidelines have evoked much discussion in the veterinary and parasitological worlds. For instance, the use of more sensitive faecal egg counting techniques have featured in many anthelmintic studies (Cabaret and Berrag 2004; Levecke *et al.* 2012; Lester *et al.* 2013) other than what the WAAVP guidelines recommend. There is also a lack of consensus in communications with respect to how percentage estimates and confidence intervals should be evaluated as part of a FECRT; with alternative experimental designs and statistical frameworks and approaches featuring in communications (Dash *et al.* 1988; Torgerson *et al.* 2005; Dobson *et al.* 2009; Lyndal-Murphy *et al.* 2010; Dobson *et al.* 2012; Levecke *et al.* 2012; Lyndal-Murphy *et al.* 2014).

WAAVP primarily produced these guidelines for anthelmintic investigations involving sheep, but were extended to consider other livestock species, such as other ruminants and equines. Many studies involving sheep have been conducted (Dobson and Barnes 1995; Cabaret and Berrag 2004; Taylor *et al.* 2009; Barrere *et al.* 2013; Falzon *et al.* 2013) since communications such as Coles *et al.* (1992),

Jackson and Coop (2000), Wolstenholme *et al.* (2004) and Abbott *et al.* (2009) warned that widespread anthelmintic resistance in sheep now poses problems to sheep farmers and the livestock industry. It is worth mentioning that at the farm level, obtaining FEC data from sheep is considered easier (due to large flock numbers and the fact that they are smaller ruminants in comparison to others) and occurs at a reduced cost in comparison to other livestock species. As a result, there is an abundance of egg count data concerning sheep.

For horses, according to Kaplan (2002) not much progress has been made in the experimental design or data analysis of the FECRT on horse farms; in fact Vidyashankar *et al.* (2007) and McKenna (2006) tell us that issues are present in the study design and analysis of FEC data from horses due to large numbers of horses being detected with zero egg counts and few numbers of horses being present in herds available to test. However, in recent times studies involving FEC data from horses have been carried out and are growing in terms of availability (Dobson *et al.* 2012; Lester *et al.* 2013; Relf *et al.* 2014; Stratford *et al.* 2014).

It is also worth noting that there has been an increase in the use of data that has been simulated through known and accepted distributions (these will be discussed in Chapter 3) and the utilisation of *Bootstrapping* and *Monte Carlo* methods (both of which will be explained later in Chapter 2), featuring in many communications as a means of obtaining and concluding on FEC data for different livestock species (Morgan *et al.* 2005; Learmount *et al.* 2006; Dobson *et al.* 2009; Denwood 2010; Levecke *et al.* 2012; Calvete and Uriarte 2013; Lyndal-Murphy *et al.* 2014; Wang *et al.* 2017). Though it is an approach with minimal costs and effort to carry out; simulations could be thought of as being *idealistic* and do require validation, usually from multiple data sources in order to achieve a degree of confidence in their representation of data that could be obtained in the *real world*.

Generally, research and obtaining cattle FEC data has been limited in the past, but the number of anthelmintic investigations being conducted is growing in recent times (Fiel *et al.* 2001; Lyndal-Murphy *et al.* 2010; Dobson *et al.* 2012; Levecke *et al.* 2012; Geurden *et al.* 2015; George *et al.* 2017).

1.6 Overview of the field study and data

The cattle FEC and FECRT data available for use in this project were obtained as a result of the *VM0503* Defra project (Defra 2015; Defra 2018b).

The Defra project began on 1st September 2011 and was completed on 28 February 2015. During the project *Westpoint Farm Vets* (WFV) conducted field studies on cattle over three and a half years, i.e. over three grazing seasons, in different parts of England. The main objectives of this project were to conclude on practical methods for the early detection of anthelmintic resistance in cattle in England and to evaluate how efficacious commonly used anthelmintics were in reducing levels of worm egg output. Farms (either beef or dairy) that participated in the field studies were clients of and approached by WFV. Farms that had adequate handling facilities and that had not yet treated their first year grazing cattle prior to turn-out to pasture were selected for the field studies, on the basis that cattle in their first grazing season had not developed their immunities towards cattle nematode species.

Table 1.1 highlights the details of the number of animals enrolled onto the study and the distributions (i.e. locations) of farms used each year. It is worth noting that in 2011, the Defra project was in its infancy and only five treatment groups (540 cattle) were recruited and all of the cattle were located in the South East of England. The project began mid-season and none of the groups of animals reached the threshold for treatment (described below) and thus no FECRTs were performed.

Year	2011	2012	2013	2014
Number of groups (animals) enrolled for monitoring	5 (540)	56 (~1,800)	41 (~1,900)	40 (~1,950)
Farm distribution	SE England	SE, Midlands, Cumbria	SE, SW; Midlands, Cumbria	SE, SW; Midlands, Cumbria

Table 1.1: Number of animals recruited onto study and the distribution of farms used by year, adapted from Defra (2015)

Composite group faecal samples were collected approximately every two weeks from farms (i.e. routine FEC screening) until the group mean FEC reached a high enough level (>150 epg) to conduct a FECRT. This threshold was chosen as it

was unlikely to be high enough to cause clinical disease in individual animals, but still high enough for a robust FECRT assessment (Coles *et al.* 1992; Coles *et al.* 2006). These FEC screenings were carried out with ten cattle being sampled per 40 cattle in a group, where possible, and approximately 50g of faeces were retrieved from each individual animal. Composite samples were prepared by taking 0.3g of fresh faeces from each of the ten animals, and each 3g composite was then examined using the Modified McMaster technique with a diagnostic sensitivity of 15 epg (MAFF 1986).

Once the group mean FEC reached the pre-determined threshold level (or less for some groups which had clinical lungworm outbreaks), a FECRT was then conducted. Cattle at Day 0 were systematically allocated to treatment groups as they came through the cattle crush and faecal samples were collected prior to treatment administration, from all animals involved. Fresh faecal samples were collected from all animals, placed into zip-lock bags, labelled with the individual ear tag numbers and refrigerated. Cattle in the treatment groups were dosed based on the individual body weights (kg), measured using either weightape or by electronic weigh scales, where available, using dose rates based on 10kg increments (ML) or 13kg increments (BZ). All cattle were returned to the same pastures so that they were subject to the same parasite challenge. Further faecal samples were collected 14 days post-treatment. The control animals which were not treated on Day 0 were treated after obtaining faecal samples on Day 14. Blinding of the laboratory technicians was maintained during faecal egg counting. However, in the years 2013 and 2014, on a small number of farms, faecal samples were collected on Days 0, 14 and 21 (in these cases, control groups were treated on Day 21 post-treatment after faecal samples were collected, as opposed to the usual Day 14 post-treatment). The Day 21 samples were collected in order to investigate whether or not egg counts increased between Days 14 and 21 post-treatment; indicating a period of temporary suspension in egg laying following treatment with the anthelmintic used, rather than truly expelling/killing worms present.

As well as egg counts, faecal samples from treated and untreated cattle (pre- and post-treatment) had worm eggs hatched and 100 individual, larvae cultured to third-stage larvae in order to identify the species of nematodes present (since

it is challenging to identify species' of nematode by inspection of their eggs). The species' proportions were evaluated before and after treatment to see if one particular species of nematode survived treatment, as this can occur if they are becoming resistant to an anthelmintic.

On farm treatments were with products either from the BZ and ML class of anthelmintics, based on farm history and previous anthelmintic use. From the BZ group, an oral drench product containing fenbendazole (Panacur 10% Oral Solution TM, MSD Animal Health, 7.5mg fenbendazole/kg bodyweight); and from the ML group, doramectin injection (Dectomax Injection for Cattle and Sheep, Elanco Animal Health Ltd, 200 mcg doramectin/kg bodyweight), doramectin pour-on (Dectomax Pour-On for Cattle, Elanco Animal Health, Ltd, 500mcg/kg bodyweight), ivermectin injection (Ivomec Classic Injection for Cattle and Sheep, Merial Animal Health, Ltd., 200mcg/kg bodyweight) and ivermectin pour on (Ivomec Classic Pour-On for Cattle, 500mcg/kg bodyweight) featured as part of this study and report (Defra 2015). A negative or a positive control group (i.e. a control group that is not exposed to the direct treatment of interest, but is exposed to some other treatment that is known to produce the expected effect) was used on all pastures, excluding those where pour-on products were used due to the likelihood of cross-contamination of negative controls with pour-on products. Treatment groups varied in size on farms throughout the study, with some farms also having more than one positive treatment group enrolled into the FECRT.

An outline of the FECRT method study design described above is given in Figure 1.4.

With respect to the egg counting techniques adopted for the duration of the project, the following methods were used in order to investigate any methodologies that could be recommended to use for the early detection of anthelmintic resistance in cattle:

- Modified McMaster method involving 30 epg limit of detection;
- Modified McMaster method involving 15 epg limit of detection;
- Modified McMaster method involving 10 epg limit of detection;

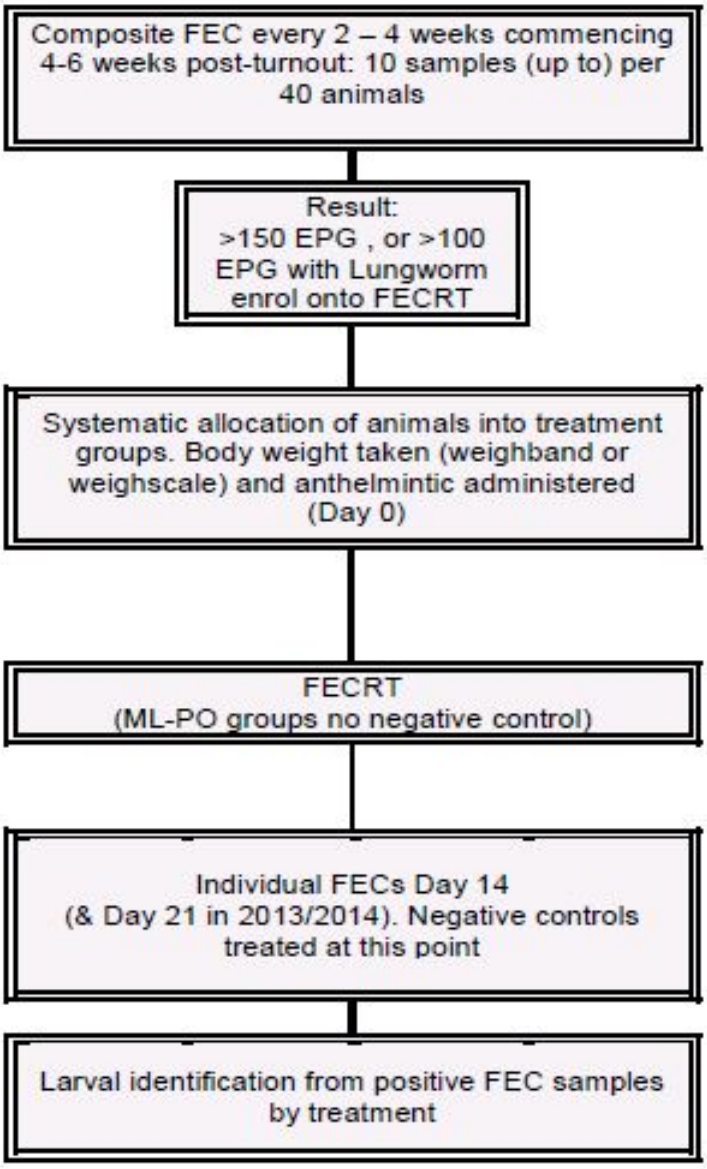

Figure 1.4: An outline of the Faecal Egg Count Reduction Test (FECRT) method study design, adapted from Defra (2015)

- Sensitive-Centrifugal Flotation Technique (SCFT) with a limit of detection of 1 epg (*Primary counting technique used in all field studies that involved following the procedure of the Modified McMaster Improved counting technique but eggs floated to a coverslip following centrifugation, re-suspension and re-centrifugation* (MAFF 1986));

- FLOTAC system using a limit of detection of 1 epg;
- and a Parasep system using 30 epg and 15 epg limits of detection.

Any individual faecal samples that were detected as having below 120 epg (in 2012) or 60 epg (in 2013 and 2014), with respect to a 15 epg diagnostic sensitivity, were re-analysed with a technique that had a lower limit of detection, i.e. the SCFT. The change in threshold between the years resulted from a project review.

Regarding the field studies as a whole, slight modifications were made each year, i.e. target minimum group size, presence of negative control group, inclusion of Day 21 FEC post-treatment. In 2014, following a statistical review of the data obtained from 2012 and 2013, negative control groups did not feature as part of the field studies for that year in order to increase available treatment group sizes. Overall however, the majority of the studies tried to follow a parallel group design, where possible.

As part of the Defra project, statistical analyses were also carried on the FECRT data. In accordance, with the recommended WAAVP guidelines (Coles *et al.* 1992; Coles *et al.* 2006) any cattle for which FECs were not obtained either on Day 0, or Day 14, were removed and not included in the analysis. Any cattle that were mis-dosed (as recorded by the veterinarian at the time of treatment), e.g. anthelmintic was rejected by animals after administration or animal movement caused only a partial dose being received, were also excluded from this analysis. Under both circumstances animals received a re-dosing with a full dose as per standard veterinary practice. Where duplicate samples were obtained through the experimental process, the first set of FECs recorded was used as part of the analysis in order to ensure independence between the egg counts obtained per animal.

In total, 113 treatment groups were considered for the statistical analyses: 25 negative control groups, 19 and 8 doramectin injectable and pour-on treatment groups respectively, 27 and 16 ivermectin injectable and pour-on treatment groups respectively and 18 fenbendazole treatment groups. Overall, 2,762 cattle met the criteria for inclusion of the statistical analyses.

The percentage reduction (1.2) was calculated

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%,$$

as well as the following percentage estimate (1.4) (which will be described in more detail in Chapter 2):

$$100\left(1-\frac{T_{14}}{T_0}\right)\%, \tag{1.4}$$

where T_0 is the arithmetic group sample mean of FECs obtained at baseline, i.e. on Day 0, from a positive treatment group, were evaluated along with 95% confidence intervals of a similar form given by the interval (1.3), to conclude on apparent efficacy/resistance in cattle herds.

For further information on the results of this analysis we refer the reader to Defra (2015), but we will consider and present some of these results later in Chapter 6.

1.6.1 Data used as part of this PhD project

We use the term data to describe each of one of the variations of the sets of egg counts that were obtained from the first and second McMaster chambers which featured as part of the Modified McMaster technique adopted, using a diagnostic sensitivity of 30 epg (hereby referred to as 30EPG_McM1, 30EPG_McM2 data, respectively). The average of the two chambers was also considered, resulting in data sets with egg counts being obtained using a diagnostic sensitivity of 15 epg (hereby referred to as 15EPG_McM counts). A hybrid set of FEC data was also considered, which involved counts obtained using the SCFT with a 1 epg sensitivity, as well as the other 15EPG_McM counts that were greater than or equal to the thresholds mentioned in Section 1.6 (hereby referred to as 15EPG_McM_SCFT data). As a result, there were four possible sets of data produced for each individual treatment group. Seventy six treatment groups were considered for analysis purposes as part of this project, based on the Day 0 and

Day 14 FEC data obtained by the counting techniques considered above. As a result, a total of 304 data sets were considered for analysis, i.e. 76 data sets were considered for each diagnostic sensitivity grouping.

From this subset of data and with the inclusion criteria adopted described in Section 1.6, 12 groups of cattle who received a fenbendazole treatment (coded as FBZ) were considered. From the ML group of anthelmintics: 19 groups of cattle who received a doramectin injection (coded as DectoInj), 8 groups of cattle who received a doramectin pour-on formulation (coded as DectoPouron), 15 ivermectin injection and 7 ivermectin pour on (coded as IvmInj and IvmPouron, respectively) treatment groups of cattle were also considered. Based on the 61 positive treatment groups present in this subset of data, the median group size for positive treatment groups was 27 cattle, with group sizes ranging between 12-61 cattle. For the 15 negative control groups present in this study, the median group size was 18 cattle, with group sizes ranging between 12-54 cattle. Overall, 2,501 cattle were sampled during the FECRTs over this subset of 52 farms, and of these, 2,175 animals results were used for analysis purposes of this project that satisfied the inclusion criteria highlighted in Section 1.6.

All analyses were carried out using *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.) and any statistical tests which feature as part of this project were carried out at a 5% significance level.

1.7 Aims and objectives of this project

The overall aim of this project is to improve the analysis of cattle FEC data and, subsequently, identify appropriate experimental designs that could be carried out and utilised as part of improving the FECRT. This is achieved using various statistical methodologies/frameworks, which will be able to provide robust statistics, in order to conclude on the apparent anthelmintic efficacy/resistance status' of cattle herds.

In Chapter 2, a literature review is presented. This review highlights the different types of statistical analyses and frameworks and examines the various exper-

imental designs that have been adopted as part of conducting FECRT studies for different livestock species.

Following the review, an investigation of the statistical validity of current guidelines on constructing percentage estimates and confidence intervals as part of the FECRT is undertaken (Chapter 3). Namely, the validity of the asymptotic assumption of normality of cattle FEC data, on which recommended confidence intervals are based, is investigated and various discrete probability distributions, such as compound distributions other than the Negative Binomial, are fitted to cattle FEC data. The latter is conducted in order to determine the most appropriate distributions, and by extension location parameter estimates/central tendencies, for representation. Based on the results, recommendations of possible alternative calculations are given.

The conclusions from Chapter 3, along with available cattle FEC field study data, are used in simulation studies that identify robust percentage estimates and, by extension, design of experiments within Bootstrapping and Bayesian frameworks, and are described in Chapters 4 and 5, respectively.

Furthermore, the robust methodologies identified in Chapter 5 are adopted and applied to available field study FEC data. A breakdown of the classifications of the apparent anthelmintic efficacy/resistance status' of cattle herds and measures of agreement between the classifications using these methodologies used and those carried out as part of the Defra project, are highlighted in Chapter 6.

Finally, general discussion and ideas about future work are described in Chapter 7, along with a robust field-based *FECRT calculator*, i.e. a *R Shiny* prototype webpage application being presented.

Chapter 2

Review of current statistical methodologies used in studies of anthelmintic efficacy

2.1 Introduction

In this Chapter, a review of the literature is presented, focusing on the experimental design considerations, various statistical calculations (i.e. different percentage estimates that can be considered) and the statistical frameworks for which interval estimation can be carried out for the FECRT. With respect to statistical frameworks, confidence intervals using asymptotic approximations for the most commonly used percentage estimates were derived and presented as part of this project. Bootstrap and Bayesian methodologies are also discussed here.

2.2 Experimental design considerations for the FECRT

According to Winer *et al.* (1991), *experimental design* is concerned with reducing and controlling variability in ways which make statistical theory applicable

to decisions made about nature. An experimental design simply describes how an experiment is to be executed and in designing the structure, decisions must be made regarding the variables and experimental material of interest, such as measured responses, objective of the study being conducted and the experimental units under observation. For example, in the case of FECRT studies, FECs would be considered the measured responses, the objective of these studies would be to investigate the apparent efficacy/resistance status of the herds being considered and the experimental units would be the animals sampled.

The types of experiment considered for the FECRT are currently recommended (Coles *et al.* 1992; Coles *et al.* 2006) to involve two treatment groups: one positive treatment group which receives the treatment/intervention under consideration and one negative control group which does not receive the treatment/intervention and these form as part of a *parallel group* study design; where only counts from end of study are included in calculations (end of study usually being Day 14 after administration of treatment/intervention). The use of a negative control group in this case is to allow for natural changes in FECs during the test period, between baseline and end of study. By only obtaining post-treatment counts, this can result in saved labour and time at the farm level. However, no information is used about the groups of animals involved from the baseline of study when investigating apparent efficacy in this case.

Some communications also recommend the use of baseline counts of the positive treatment and negative control groups (Dash *et al.* 1988; Torgerson *et al.* 2005; Lyndal-Murphy *et al.* 2010; Lyndal-Murphy *et al.* 2014). As a result, all of the available information that can be obtained from this type of experiment can be used. However, interpreting the changes in the negative control group could be challenging due to the different sources of variation. Indeed, Vidyashankar *et al.* (2007) and Vidyashankar *et al.* (2012) tell us that there are numerous sources of variability in FEC data that can impact on the interpretation of efficacy. Some of these sources include the overdispersed distribution of parasites in the hosts (causing differences in pre-treatment counts between animals on farm), differences in infection intensities between farms (causing differences in counts between farms), the variability in different samples from the same animals, overall health and well-being of animals which impacts on the pharmacokinetics and pharmacody-

namics, spatial and temporal differences between farms due to the location of the farms and time of sampling, respectively, and differences in age, breed and sex of animals both on and between farms (we refer the reader to Vidyashankar *et al.* (2007) and Vidyashankar *et al.* (2012) for further information).

There are some communications, however, that conduct anthelmintic studies by only considering the pre- and post-treatment counts of a positive treatment group, particularly in the instance where it may not be possible to have a negative control group (Cabaret and Berrag 2004; Lyndal-Murphy *et al.* 2010; Levecke *et al.* 2012; Lester *et al.* 2013; Lyndal-Murphy *et al.* 2014; Stratford *et al.* 2014; George *et al.* 2017). In fact, Vidyashankar *et al.* (2007) tell us that this type of design of experiment is adopted for anthelmintic studies in horses; mainly due to the fact that too few horses are usually available to contribute to two treatment groups on one farm. This set up of experimental design ensures that all animals are treated. In the instance of considering counts from both Day 0 and Day 14 from a treatment group, this would form the basis of a *paired study* design or, as it is formally known, a *paired comparison* study since repeated measures are involved from the same subjects. Gardiner and Gettinby (1998) tell us that this design allows us to account for an extraneous source of variation. Any differences detected between FECs obtained between Days 0 and 14, could be due to, for example, the effects of the treatment, variation between animals on each day or a mixture of both. A paired study design allows us to account for the variation between animals because each animal is considered as their own control, thus any differences between the measured responses over time are more likely to be caused by the effects of the treatment administered. It is worth noting that the paired study design focuses on one treatment group being examined, this usually being the positive treatment group, whereas a parallel group design is an extension that involves two treatment groups under paired study design regimes being used in calculations (where a negative control group would usually be considered as well). The distinction between the parallel group design and paired study design can be highlighted in Figure 2.1.

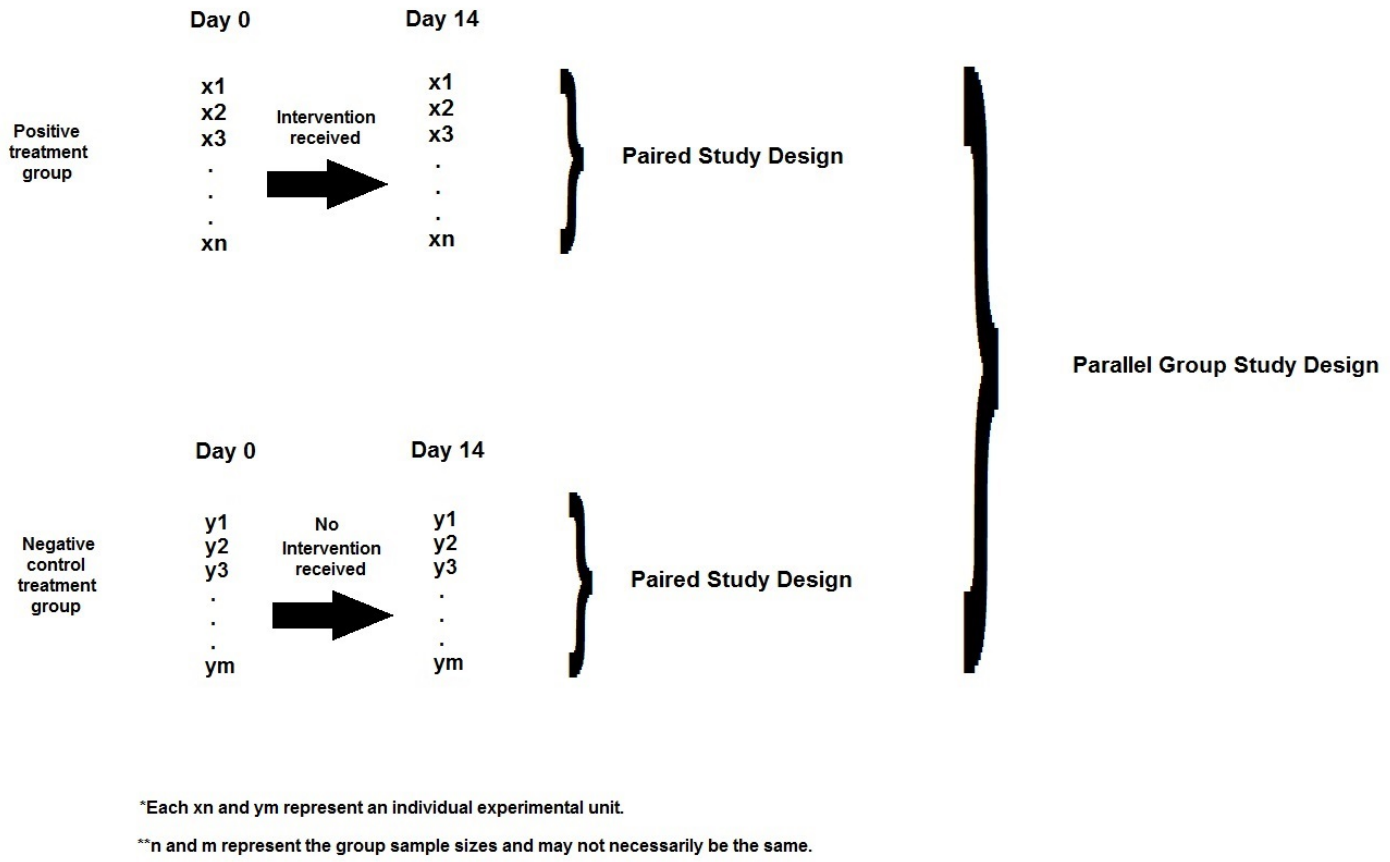

Figure 2.1: Distinction between *Parallel Group* and *Paired Study* Designs used for Anthelmintic Studies

Gardiner and Gettinby (1998) also tell us that the development of experimental design principles is generally attributed to Sir Ronald Fisher (Fisher 1970) and terms such as statistics, experimental design, treatment effect, randomisation, etc. were to become recognised with the planning, collecting and analysis of data. In fact, the term *experiment* is open to a very broad interpretation and covers any type of study, trial or investigation where data are to be collected and assessed. With respect to variation, Gardiner and Gettinby (1998) also explain that data can be thought of intuitively as a combination of variation which is controlled through the experimental process and variation which can be considered as *random*, i.e. *experimental error*. Indeed, experimental design is considered the cornerstone of good statistical practice and has been internationally adopted for the effective development of new processes, products and investigations (Gardiner and Gettinby 1998).

The question, however, of what design of experiment, and particularly what percentage estimate, should be used to ensure robust statistical calculations are carried out as part of a FECRT remains unanswered. With respect to anthelmintic studies for cattle, this is mainly due to limited research being carried out in this species as well as a lack of consensus being present towards the definition of a design being *robust* for these investigations (Cabaret and Berrag 2004). Despite the differences discussed above, as agreed by many researchers, there is a strong need to investigate which design is appropriate to determine and classify apparent anthelmintic efficacy in a robust manner with respect to statistical calculations being carried out as part of the FECRT (Vidyashankar *et al.* 2007; Vidyashankar *et al.* 2012; Lyndal-Murphy *et al.* 2014).

2.3 Percentage estimates for efficacy

To investigate apparent anthelmintic efficacy, the FECRT is currently recommended to evaluate a percentage estimate (based on the ratio of arithmetic means from the treated and untreated groups of animals) and a 95% confidence intervals (CIs) for these estimates, using faecal samples collected on Day 14 post-treatment (Coles *et al.* 1992) as part of a parallel group design. This estimate is shown in

(2.1):

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%, \tag{2.1}$$

where T_{14} and C_{14} are the arithmetic group sample mean FECs collected on Day 14, after baseline of study, from the treatment and control groups respectively.

In this Section, we consider the generalisation of the above percentage estimate and other formulations that are explored in the communications considered.

The percentage estimate (2.1), is a specific case of the generalisation in (2.2):

$$100\left(1-\frac{T_i}{C_i}\right)\%, \tag{2.2}$$

where T_i and C_i are the arithmetic group sample mean FECs collected on Day i, after baseline of study, from the treatment and control groups, respectively (Pepper *et al.* 2003; McKenna 2006; Barrere *et al.* 2013; Falzon *et al.* 2013).

Another commonly used percentage estimate observed in the literature, which involves the ratio of arithmetic means of baseline and end of study FECs from a positive treatment group of animals (i.e. from a paired study design) is percentage estimate (2.3):

$$100\left(1-\frac{T_i}{T_0}\right)\%, \tag{2.3}$$

where T_i and T_0 are defined as the arithmetic group sample mean FECs at the end of study, on Day i, and baseline, on Day 0, for the treatment group, respectively (Kochapakdee *et al.* 1995; McKenna 2006; Lyndal-Murphy *et al.* 2010; Dobson *et al.* 2012; Levecke *et al.* 2012; Vidyashankar *et al.* 2012; Lester *et al.* 2013; Lyndal-Murphy *et al.* 2014; Relf *et al.* 2014; George *et al.* 2017).

The final, most commonly used percentage estimate in communications, which involves baseline and end of study arithmetic mean FECs from positive treatment and negative control groups, is given in (2.4):

$$100\left(1-\frac{\left(\frac{T_i}{T_0}\right)}{\left(\frac{C_i}{C_0}\right)}\right)\% = 100\left(1-\frac{C_0T_i}{T_0C_i}\right)\% \tag{2.4}$$

where T_i, T_0, C_i, C_0, are defined as the arithmetic group sample mean FECs at the end of study, on Day i, and baseline on Day 0, for the treatment and control groups, respectively (Dash *et al.* 1988; Torgerson *et al.* 2005; McKenna 2006; Taylor *et al.* 2009; Lyndal-Murphy *et al.* 2010; Dobson *et al.* 2012; Lyndal-Murphy *et al.* 2014).

The *Symmetrised Percentage Change* (SPC) (Berry and Ayers 2006), given in (2.5):

$$100\left(\frac{T_0-T_i}{T_0+T_i}\right)\% \tag{2.5}$$

could also be considered as an estimate for apparent anthelmintic efficacy, where T_0 and T_i are defined as before. If this percentage estimate were to be considered, we would be able take account of any increases in FECs between baseline and end of study as well as any zero valued FECs at baseline or end of study. The SPC, according to Berry and Ayers (2006), is also bounded ($\pm 100\%$) and having a bounded range means that the influence of outliers is greatly reduced. It is often the case with FEC data that a small number of individual animals will be shedding high numbers of eggs in their faeces, and thus be outliers.

It is also worth noting that values obtained for the most popular percentage estimates are not necessarily the same as those obtained using the SPC. For example, if $T_0 = 100$ and $T_{14} = 5$, then this would correspond to a 95% reduction, but the $SPC = 90.5\%$. Also, if $T_0 = 100$ and $T_{14} = 10$, then this would correspond to a 90% reduction, but the $SPC = 81.8\%$ (these would be the threshold values to

be adopted to correspond to those recommended by WAAVP).

The approach of averaging over individual-based egg count percentage reductions/changes has also been considered, but not much research has been dedicated to this concept (Cabaret and Berrag 2004). For instance, in the case of considering percentage estimate (2.3), we would have (2.6) as an average:

$$\frac{\sum_{j}^{n_{treat}} \left[100\left(1-\frac{T_{i,j}}{T_{0,j}}\right)\%\right]}{n_{treat}}, \tag{2.6}$$

where $T_{0,j}$ and $T_{i,j}$ are, respectively, Day 0 and Day i FECs from host j, from a total n_{treat} hosts from the positive treatment group. In fact, each host would serve as its own control. To consider this type of approach, we would be required to work with FECs from individual animals at baseline and end of study. Cabaret and Berrag (2004) tell us that the individual-based egg count percentage reductions/changes of the form (2.6) presented lower values and more reliable evaluations given certain conditions than the average-based egg count percentage reductions/estimates in most cases.

We could also extend this approach to consider the individual-based egg count percentage reductions/changes in the case of the SPC (2.5), that is (2.7):

$$\frac{\sum_{j}^{n_{treat}} \left[100\left(\frac{T_{0,j}-T_{i,j}}{T_{0,j}+T_{i,j}}\right)\%\right]}{n_{treat}}. \tag{2.7}$$

2.3.1 Central tendencies for representing egg count data

The most commonly used percentage estimates in the literature, i.e. percentage estimates (2.2), (2.3) and (2.4), are based on arithmetic group means of treatment group FECs. These are the central tendencies/maximum likelihood estimates from a Poisson or a Negative Binomial distribution, which FEC data are assumed to follow since most hosts are uninfected with parasites (or contain very few parasites) and very few hosts are heavily infected; hence parasite populations

are known to be statistically aggregated (Shaw and Dobson 1995; Levecke *et al.* 2012; Wang *et al.* 2017).

These are not the only central tendencies that have been considered in the literature, however. According to Presidente (1985), Wood *et al.* (1995), Smothers *et al.* (1999) and Vercruysse *et al.* (2001), geometric group means could also be used when forming percentage estimates, on the grounds that we could gain a more accurate representation of the distribution of nematode populations within a herd of animals and could also gain a more accurate degree of apparent efficacy of an anthelmintic. Upton and Cook (2011) defines the geometric mean, g as

$$g = \left(\prod_{i}^{n} x_i \right)^{\frac{1}{n}}$$

for a set of positive observations $x_1, x_2, ..., x_n$. However, Dobson *et al.* (2009) concluded that efficacy estimated from using arithmetic means using percentage estimate (2.2) provided consistent, unbiased results, i.e. the expected value of efficacy was close to the true value of efficacy, in comparison to the use of geometric means which consistently produced biased results. Additionally, Dash *et al.* (1988) advocates the use of arithmetic means as opposed to geometric means on the basis that geometric means underestimate total egg outputs and different transformations being used on egg count data, before evaluating the geometric mean (otherwise the geometric mean becomes zero in the presence of zero egg count data being present) make it difficult to draw comparisons and conclusions.

Transformations of FEC data, such as $log_{10}(\cdot)$, $ln(\cdot)$, *square-root* and *arcsine* transformations, have been considered by several authors in order to correct the usual skewness present in FEC data. The transformed data have then been used to calculate arithmetic means (Fulford 1994; Pook *et al.* 2002; Mejia *et al.* 2003; Torgerson *et al.* 2005; Vidyashankar *et al.* 2007; Dobson *et al.* 2009). Transformations such as $ln(x + a)$, where $a > 0$, have also been considered to take account of egg counts with a value of zero (Torgerson *et al.* 2005). More on

data transformations will be discussed in Section 3.2.1.

However, it could be meaningful to investigate other representative distributions and their relevant central tendencies, i.e. investigate the validity of assuming FEC data follow Poisson/Negative Binomial distributions, and hence observe whether it is valid to use arithmetic means in percentage estimate calculations.

2.3.2 Maximum likelihood estimation (MLE) as a means of estimating central tendencies

Maximum likelihood estimation (MLE) is the most commonly used technique for estimating parameters of probability distributions (Rice 2007), such as central tendencies. It is based on the *Likelihood Principle*, which asserts that if a probabilistic model is claimed to describe the behaviour of observed data, then all the relevant information about the model (available from the data) is contained in the joint probability density function (pdf) or probability mass function (pmf), assuming the model is correct.

Suppose a random variable X had pmf/pdf $f_X(x;\theta)$ on some domain Ω for X, where θ is a parameter of the model whose value we do not know but wish to estimate. Suppose further that we observe a simple, random sample of size n, where our observations are independent of one another and follow the identical pdf with the same value of θ. Then the joint pdf of our vector of observations $\mathbf{X} = (X_1, X_2, ..., X_n)$ is simply

$$f_{\mathbf{X}}(x,\theta) = f_X(x_1;\theta)f_X(x_2;\theta)...f_X(x_n;\theta)$$

on the domain Ω^n. This is a legitimate multivariate pmf/pdf of $\mathbf{X}$. It is worth noting however, that the unknown parameter θ plays an essential role and, in general, it too will belong to some particular domain, i.e. the parameter space T. In using the maximum likelihood method, instead of considering $f_{\mathbf{X}}(x,\theta)$ as a

function of $x_1, x_2, ...x_n$ in which θ is a fixed but unknown constant, we regard $x_1, x_2, ...x_n$ as values which are fixed once the data have been collected, and allow θ to vary across the parameter space T. When we think of the joint density function as a function of θ in this way, it is re-named as the *Likelihood Function*.

We write the likelihood function as given in (2.8):

$$L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f_X(x_i; \theta), \tag{2.8}$$

where the x_i are the different observed data values. The likelihood function does not have a direct probabilistic interpretation in terms of the parameter space, but the most plausible values for θ are those which make the likelihood function as large as possible, i.e. make the observed data values *most likely*, and this principle is used for choosing an estimator $\hat{\theta}$ for θ. This is regarded as the *Maximum Likelihood Principle*.

It is worth noting that, according to Rice (2007), it is the *log-likelihood* function which is most often used in practice, i.e. by using equation (2.8) we consider $l(\theta) = ln(L(\theta))$. Working with the log-likelihood function often simplifies the algebra involved in deriving estimators plus any value of θ which maximises $L(\theta)$ will also maximise $l(\theta)$, since the $ln(\cdot)$ is monotonically increasing function. Finally, if a probability model contains several unknown parameters, then the maximum likelihood method can be used to estimate these parameters *simultaneously*.

2.4 Statistical frameworks used to derive point and interval estimates of data

The point and interval estimates associated with a data set can be estimated in a variety of ways. In this Section, we consider the three main statistical frameworks from which we can obtain point and interval estimates of data, namely the asymptotic approximation, Bootstrap and Bayesian approaches.

2.4.1 Asymptotic approximation for obtaining point and interval estimates

2.4.1.1 Calculation of expected values and variance

Suppose we have a random variable X with some distribution $f(x)$ and we know the expected value, $E[X]$, and variance of X, $Var[X]$. Let the random variable Y be defined as $Y = g[X]$ for some real-valued function $g(\cdot)$. How do we calculate $E[Y]$ and $Var[Y]$? To solve this problem, in the case of finding an analytical solution, we would evaluate either

$$E[Y] = \sum_j g(x_j)P(X = x_j)$$

for a discrete random variable X, or

$$E[Y] = \int_{-\infty}^{\infty} g(x_j)f(x_j)$$

for a continuous random variable X, where x_j, for $j = 1, .., n$ are our data.

The $Var[Y]$ is defined as

$$\begin{aligned} Var[Y] &= E[Y^2] - (E[Y])^2 \\ &= E[g(X)^2] - (E[g(X)])^2. \end{aligned}$$

However, for most real world problems, the conclusion of an analytical solution is intractable and as a result, we can look to approximation approaches to aid us in point and interval estimation, such as the *Delta Method*.

2.4.1.2 The Delta method

The main idea behind the Delta method is that we can use a linear approximation of $g(\cdot)$, near the maximum likelihood estimator (mle) $\hat{\theta}$, for the population parameter θ, of a random variable X this gives approximations for $E[Y]$ and $Var[Y]$ (Hosmer *et al.* 2008) for Y defined as in Section 2.4.1.1.

Let $g : \Theta \to \mathcal{R}$, where Θ is the parameter space and suppose that $g(\cdot)$ is differentiable for all $\theta \in \Theta$ and $g'(\theta) \neq 0$. If we use a first-order Taylor series approximation around θ:

$$Y = g(X) \approx g(\theta) + g'(\theta)(X - \theta)$$

and so to first order

$$\begin{aligned} E[Y] &\approx E[g(\theta) + g'(\theta)(X - \theta)] \\ &= E[g(\theta)] + E[g'(\theta)(X - \theta)] \\ &= g(\theta), \ \textit{since } E[(X - \theta)] = 0. \end{aligned} \tag{2.9}$$

Furthermore, we obtain the first-order approximation for $Var[Y]$

$$\begin{aligned} Var[Y] &\approx Var[g(\theta)] + Var[g'(\theta)(X - \theta)] \\ &= 0 + [Var[X] - Var[\theta]](g'(\theta))^2 \\ &= (g'(\theta))^2 Var[X]. \end{aligned} \tag{2.10}$$

How good the approximations (2.9) and (2.10) are, depends on how *non-linear* $g(\cdot)$ is in the neighbourhood of θ and also on the size of $Var[X]$.

For the remainder of this Chapter, we will let $g(\cdot)$ equate to the natural logarithmic function $ln(\cdot)$, since, if we were to consider the logarithm of a ratio of two

parameters, then this is equivalent to considering the difference of the logarithms of the two parameters. As a result, by using approximations (2.9) and (2.10) we can obtain the following approximations (2.11) and (2.12):

$$E[Y] \approx ln(\theta) \tag{2.11}$$

and

$$Var[Y] \approx \frac{Var[X]}{\theta^2}. \tag{2.12}$$

2.4.1.3 Extension of the Central Limit Theorem

Another useful property which we will make use of centralises around convergence of estimators. Suppose that a sequence of random variables X_n converges to a random variable X, if there exists N such that for all $n > N$, X and X_n have the same cumulative distribution function. Further, assume

$$\hat{\theta} \xrightarrow{d} \mathcal{N}\left(\theta, \frac{\sigma^2}{n}\right)$$

where $\xrightarrow{d}$ represents convergence in distribution, which implies that $\hat{\theta}$ will be centred at θ with variance $\frac{\sigma^2}{n}$ for a sufficiently large sample size n and $\mathcal{N}$ represents the normal distribution with mean θ and variance $\frac{\sigma^2}{n}$. If we consider a differentiable and continuous function $g(\cdot)$ for all θ in the parameter space Θ, then we have the following result (Hosmer *et al.* 2008; Greene 2012):

$$g(\hat{\theta}) \xrightarrow{d} \mathcal{N}\left(g(\theta), \frac{\sigma^2(g'(\theta))^2}{n}\right). \tag{2.13}$$

If we let $g(\cdot)$ equate to the natural logarithmic function $ln(\cdot)$ again, then the convergence property (2.13) becomes

$$ln(\hat{\theta}) \xrightarrow{d} \mathcal{N}\left(ln(\theta), \frac{\sigma^2}{n\theta^2}\right). \tag{2.14}$$

In essence, equation (2.14) tells us that the sampling distribution of $ln(\hat{\theta})$ will follow a normal distribution with mean $ln(\theta)$ and variance $\frac{\sigma^2}{n\theta^2}$. This is similar to the Central Limit Theorem (Upton and Cook 2011; Rumsey 2016), which states that in spite of the distribution of the population, the sampling distribution of the sample mean, $\bar{x}$ say, will follow a normal distribution with mean θ and variance $\frac{\sigma^2}{n}$, for sufficiently large n.

One might ask how sufficiently large n has to be in order to ensure the sampling distribution of the sample mean satisfies this condition? The answer to this question depends on the distribution of the population from which the samples are taken from. If we consider a population that has finite mean θ and variance σ^2 and is normally distributed then the distribution of $\bar{x}$ must be normally distributed, irrespective of the size of n. In other words, if the population distribution is normal then the sampling distribution of $\bar{x}$ is normal. However as the population becomes of a more non-normal nature, a larger n is needed to guarantee convergence (Rumsey 2016). Upton and Cook (2011) tell us that as long as a sample size of at least 30 is obtained, then the sampling distribution of $\bar{x}$ can be assumed to be normal.

2.4.2 Bootstrapping theory

The use of Bootstrapping methods has proved to be popular in the veterinary and parasitological communities in different animal studies (Cabaret and Berrag 2004; Vidyashankar *et al.* 2007; Traversa *et al.* 2009; Vidyashankar *et al.* 2012; Lester and Matthews 2013; Lester *et al.* 2013; Wang *et al.* 2017) in order to obtain point and interval estimates for percentage estimates. In this Section, we describe the various Bootstrapping techniques that can be used to obtain these.

Bootstrapping is a computer-intensive technique which involves re-sampling from observed data. Essentially, there are two types of Bootstrapping: parametric and

non-parametric. Parametric Bootstrapping involves re-sampling from observed data under the assumption of a well-defined probability distribution. Conversely, non-parametric Bootstrapping does not require a probability distribution to be assumed. The theory proceeds as follows: suppose we have a set of n random measurements of some characteristic of the population of interest (in our case this would be FECs) and we want to estimate some parameter of the distribution. Bootstrapping theory tells us that the true distribution of FECs, f say, can be reasonably approximated by the distribution of sampled observations, f^* say. As a result the larger the value of n the more reasonable this assumption. Based on this assumption, the theory continues by constructing n random samples (with replacement) from f^* and calculating the statistic of interest from the generated sample. This process is repeated a large number of times until a stable distribution of the statistic of interest is obtained. This distribution is then the uncertainty/sampling distribution of the statistic of interest for which we can obtain point estimates and confidence intervals.

The following give algorithmic steps associated with the parametric and non-parametric Bootstrapping procedures.

The Non-parametric Bootstrap:

1. *Collect a data set of measurements with sample size n:* $\underline{X} = \{x_1, \ldots, x_n\}$.
2. *Generate B Bootstrap data sets* $\{x_1^*, \ldots, x_B^*\}$, *where each* x_i^* *(for* $1 \leq i \leq B$*) is a generated dataset of size n, i.e. a replication/iteration, featuring randomly sampled values (with replacement) from* $\underline{X}$.
3. *For each Bootstrap data set* $\{x_1^*, \ldots, x_B^*\}$ *calculate the required statistic of interest,* $\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$ *respectively.*

The Parametric Bootstrap:

1. *Collect a data set of measurements with sample size n:* $\underline{X} = \{x_1, \ldots, x_n\}$.
2. *Determine the parameter(s) of the distribution that best fits the data from the known distribution, such as the maximum likelihood estimator(s).*
3. *Define the known distribution using, say, the maximum likelihood estimate(s) of the parameter(s).*
4. *Generate B Bootstrap data sets* $\{x_1^*, \ldots, x_B^*\}$*, where each* x_i^* *(for* $1 \leq i \leq B$*) is a generated data set of size n, i.e. a replication/iteration, featuring randomly sampled values (with replacement) from the fitted distribution.*
5. *For each Bootstrap data set* $\{x_1^*, \ldots, x_B^*\}$ *calculate the required statistic of interest which estimate the parameter(s) defined earlier,* $\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$ *respectively.*

Both techniques offer their own benefits and costs as described in Table 2.1:

	Non-parametric Bootstrap	Parametric Bootstrap
Advantages	– Simple to apply. – General and robust method of setting confidence intervals.	– General and robust method of setting confidence intervals. – measurements need not be identically and independently distributed.
Disadvantages	– Measurements need to be identically and independently distributed. – In complex applications, it is often unclear what the unit of re-sampling should be. – Generally only asymptotically exact, i.e. when $B, n \to \infty$.	– Parametric model must be assumed. – Generally only asymptotically exact, i.e. when $B, n \to \infty$.

Table 2.1: Advantages and disadvantages for types of Bootstrapping

The Bootstrap approximates the sampling distribution with three sources of approximation error: simulation, statistical and specification. The simulation error refers to the fact that we are using a finite number of replications to stand for the *whole* sampling distribution. Statistical error refers to the distributions of the simulated data estimates and the distribution of the original data set not being exactly the same. Finally, the specification error refers to the idea that any specified distribution, i.e. the subjectivity of the modeller when specifying the distribution, might not be suitable given the original and any simulated data sets.

With bootstrapping methods, there are usually three statistical features which

are of interest:

Distribution of Uncertainty:

When considering the approaches described above, the distribution of the estimate $\hat{\theta}$ can be denoted as $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$, and this represents the uncertainty about the true value of the population parameter θ.

Variance and Standard Error:

According to Efron and Tibshirani (1993), from the sampling distribution $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$ we are able to derive approximations for the variance and standard error for the population parameter θ namely:

$$\begin{aligned} Var[\theta] &\approx Var[\hat{\theta}] \\ &\approx Var[dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}] \end{aligned}$$

where $Var[dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}]$ is the sample variance estimate of the sampling distribution of $\hat{\theta}$.

The standard error associated with θ, $se[\theta]$ can be approximated in a similar way:

$$\begin{aligned} se[\theta] &\approx se[\hat{\theta}] \\ &\approx sd[dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}] \end{aligned}$$

where $sd[dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}]$ is the sample standard deviation estimate of the sampling distribution of $\hat{\theta}$.

Bias:

Bias can be defined as the difference between the population parameter θ and the expected value of the estimator $\hat{\theta}$. This bias may occur due to simulation, statistical and/or specification errors occurring. It can be written and approximated

as follows:

$$Bias = E[\hat{\theta}] - \theta \tag{2.15}$$
$$\approx E[dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}] - \hat{\theta}$$
$$\approx \hat{\theta^*} - \hat{\theta} \tag{2.16}$$

where $\hat{\theta^*}$ is the average of $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$.

The most common reason as to why we would be interested in estimating the bias (2.15) of $\hat{\theta}$ is to enable correction for it. To do this, let approximation (2.16) be denoted as b. An obvious *bias-corrected estimator* would be as follows:

$$\begin{aligned} \theta^- &= \hat{\theta} - b \\ &= \hat{\theta} - (\hat{\theta^*} - \hat{\theta}) \\ &= 2\hat{\theta} - \hat{\theta^*} \end{aligned} \tag{2.17}$$

However, Efron and Tibshirani (1993) cautions us that this straight-forward bias correction (2.17) may not be reliable in practice - for instance correcting the bias may result in a larger standard error being evaluated. The notion of bias correction will be used when constructing confidence intervals in Section 2.5.2.

2.4.3 The Bayesian approach

Bayesian statistics involve estimating parameters of interest using available data. The classical approach to statistics (often known as the *frequentist* approach) assumes that the parameters of interest are fixed, but have known values to be estimated; whereas the Bayesian approach assumes the parameters are not fixed but have some fixed, unknown distribution. Thus by using the Bayesian approach, this leads us to working with a distribution for the parameters of interest and is the basis for subsequent inference within the Bayesian paradigm (Lee 2004; Rice

2007; Gelman *et al.* 2013).

In the discrete sense and according to Rice (2007), *Bayes' Theorem* can be thought of in terms of basic probability laws. Let A and $B_1, \ldots, B_n$ be events, such that the B_i are disjoint, the union of the B_i's forms the sample space and $P(B) > 0$ for all i where P denotes the probability of the event occurring. Then based on the law of conditional probability (which we do not state here):

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

In the continuous sense, Bayes' Theorem states in (2.18):

$$\pi(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)}, \tag{2.18}$$

where $\pi(\theta|\mathbf{x})$ is the posterior distribution of the parameters $\theta = \{\theta_1, \ldots, \theta_m\}$ given the data $x = \{x_1, \ldots, x_n\}$, $f(x|\theta)$ is the probability of observing the data x under different parameter values θ (this is known as the likelihood), $p(\theta)$ is the prior distribution of the parameters and $f(x)$ is a normalisation constant so that the posterior distribution is a valid probability density function. In the case of discrete random variables, the probability density functions are replaced with probability mass functions. In fact, Bayes' Theorem is often written as

$$\pi(\theta|x) \propto f(x|\theta)p(\theta),$$

Intuitively, Bayes' Theorem focuses on our prior beliefs about an experiment before we observe any data (the *prior*); we then conduct the experiment and observe data (the *likelihood*) and as a result we update our prior beliefs on the grounds of the observed data (the *posterior*), where the updating procedure is

just Bayes' Theorem (2.18).

We may, however, be interested in a particular parameter, θ_1 say, and we can make inference on this parameter from the posterior distribution given in (2.18) by the *marginal* (posterior) distribution, which is defined as

$$\pi(\theta_1|x) = \int \pi(\theta|x) d\theta_2, \ldots, d\theta_m \tag{2.19}$$

(the integration featured in (2.19) is analytically intractable for many real problems but we are able to estimate such integrals).

A Bayesian approach to analysing data offers many benefits as well. For instance, the usual normality assumption within statistical models can typically be removed and unrealistic assumptions and simplifications can be avoided when considering data. The more appealing advantage of this approach is the idea of being able to incorporate external information into an analysis, via prior specification. This concept is, however, controversial, since classical approaches to statistics involve any analyses being objective and based purely on the observed data.

To commence the analysis, a prior distribution is specified, for each parameter of interest, independent of the observed data. Essentially, there are two types of prior distribution that can be considered: *informative* and *non-informative* priors. The former involves aiming to reflect the available information, independently of the data being studied. Conversely, a non-informative prior distribution reflects not having any prior information concerning the parameters of interest. Examples of this type of prior include the Uniform, Jeffrey's and Hierarchical. One can also adopt the use of *conjugate* priors, where the prior for the parameter(s) are such that the corresponding posterior distribution for them is from the same family (irrespective of sample size and data observed).

It is also worth mentioning that the dependence of the posterior distribution on the prior distribution should always be assessed through a prior sensitivity analysis to confirm whether or not the posterior distribution is *data-driven* (i.e. a posterior distribution that is insensitive to the choice of prior) or *prior-driven*

(i.e. a posterior distribution that is sensitive to the choice of prior), and this is typically carried out by considering a range of prior specifications and comparing the posterior distributions obtained. In the presence of reasonably informative data there should be little prior sensitivity, though extreme prior sensitivity often points to problems such as parameter redundancy or overly restrictive prior assumptions.

The posterior distribution is the most appropriate summary for the parameters involved, and is often summarised by point estimates (such as central tendency, spread and correlation) and credible intervals for these estimates, of which we will discuss in Section 2.5.3. To obtain these estimates of interest, we require integration of the posterior density, e.g. for the specified value

$$E_\pi(\theta) = \int \theta \pi(\theta|x) d\theta.$$

Integration can be undertaken using Monte Carlo methods and this is known as *Monte Carlo* integration (Rice 2007) and Expectations of this sort can be estimated by drawing samples $\theta_1, \ldots, \theta_m$ from the posterior distribution of θ (i.e. obtaining the sampling distribution of θ) and then calculating the sample mean/median of these values - this method of integration is also necessary to derive the marginal distribution (2.19). In a similar way, we would also be able to obtain estimates of spread, such as the variance.

The question of how to generate these samples from the posterior arises. One way of doing this is to construct a *Markov Chain*. This is a stochastic sequence of numbers where each value in the sequence depends only upon the last and the starting point of the process is arbitrary. At time t, the state of the chain is updated from θ_t to θ_{t+1} using some stochastic process. As a result, such a Markov Chain - given certain conditions - should converge to a *stationary* distribution, i.e. if we run the chain for a long enough time then the generated values will have a certain probability distribution, which would be the posterior. Thus, we would have carried out a Markov Chain Monte Carlo (MCMC) process.

There are several ways of constructing these chains, for example *Gibbs* sampling

and the *Metropolis-Hastings* (MH) algorithms, and these sampling methods are described in Appendix B.1. However, it is worth mentioning that an assessment of convergence should always be carried out to observe indications that the distribution of parameters has reached convergence to the stationary distribution. this can be done by examining relevant *trace plots*, *acceptance rates* (essentially the percentage of unique values the chain has) and *autocorrelations* of chains and one could also consider deriving the *Brooks-Gelman-Rubin* Statistic. These methods of assessment are described in further detail (and have been utilised as part of work presented in Chapter 5) in Appendix B.2. It is worth noting however, that even though convergence is guaranteed mathematically, there is no way of being able to *prove* if a chain has converged to the stationary distribution - these assessments can only provide an *indication* of lack of convergence and how efficient we are being when sampling estimates from the stationary distribution. On that note, sometimes a *burnin-period* - a discarding of so many initial samples - is required in order to try and achieve sampling from the converged distribution (Rice 2007). As well as this, one can also consider *thinning-out* the resulting sample of estimates, i.e. consider every n-th sample, which can aid in avoiding autocorrelation bias of Markov chain samples and can help to save storage and memory.

The main appeal of the MCMC method for obtaining posterior estimates is that the updating procedure remains relatively simple; no matter how complex the posterior distribution may be. Though a computational skill set may be required to carry out analysis in the Bayesian context; software packages such as *Just Another Gibbs Sampler* (JAGS), *Bayesian inference Using Gibbs Sampling* (OpenBUGS or WinBUGS) and *R/RStudio* can be used to ease the implementation of computations (Lunn *et al.* 2000; M. Plummer 2008; CRAN 2018).

Research into the application of Bayesian methods when conducting anthelmintic studies has featured in communications, but mainly using equine FEC data (Denwood 2010; Denwood *et al.* 2010; Vidyashankar *et al.* 2012; Lester and Matthews 2013; Matthews 2014). Matthews (2014) tells us that MCMC methods are being suggested to account for the highly aggregated distribution that appears when dealing with equine FEC data. However, Bayesian approaches are being employed in more recent sheep and cattle studies (Denwood *et al.* 2008; Dobson

et al. 2012; Busin *et al.* 2013; Geurden *et al.* 2015; Wang *et al.* 2017).

As mentioned in Section 1.5.2, the amount of egg count data being obtained from horses is growing in terms of availability; yet most results being derived using MCMC methods are based on simulated horse egg count data sampled from different distributions - with only some verification of real-world data being conducted (Denwood *et al.* 2008; Denwood 2010; Denwood *et al.* 2010). Matthews (2014) also tells us that a limitation to adopting MCMC methods in analysing FEC data is the ability to use advanced statistical programmes, such as JAGS, BUGS and *R/RStudio* discussed earlier in this Section, which the layperson may not be familiar with using. On the other hand, attempts for those with a lack of statistical training to carry out these advanced statistical methods have been made in the form of producing *R/RStudio* packages such as *eggCounts* and *bayescount* (Denwood 2010; Michaela and Furrer 2014; Torgerson *et al.* 2014; Denwood 2015).

In contrast, not much research has been invested in adopting Bayesian methods for the analysis of cattle FEC data (Geurden *et al.* 2015; Wang *et al.* 2017). This could be due to the limited research and data that has been carried out and made available to date. It may thus be worth exploring this further.

2.5 Obtaining intervals for estimates within the different statistical frameworks

2.5.1 Confidence intervals using asymptotic approximation

Many communications obtain the corresponding confidence intervals for percentage estimates which assume a normal distribution under the logarithmic transformation as discussed in Section 2.4.1.3 (Coles *et al.* 1992; Lyndal-Murphy *et al.* 2010; Levecke *et al.* 2012; Lyndal-Murphy *et al.* 2014).

In this Section, original derivations of confidence intervals, for some of the more commonly used percentage estimates, i.e. (2.2), (2.3) and (2.4), were carried out as part of this project and make use of the asymptotic approximation presented

in Section 2.4.1.3. Any statistical properties that should be considered when deriving the confidence intervals are also discussed.

2.5.1.1 Percentage estimate derivation 1: post-treatment counts (negative control group of cattle present)

As mentioned in Section 1.5.2, the percentage reduction recommended by Coles *et al.* (1992) and Coles *et al.* (2006) as part of a parallel group design is percentage estimate (1.2):

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%,$$

and has a corresponding 95% confidence interval (1.3):

$$100\left(1-\frac{T_{14}}{C_{14}}\exp\left(\pm 2.048\sqrt{\frac{s^2_{t.eos}}{n_{treat}T^2_{14}}+\frac{s^2_{c.eos}}{n_{control}C^2_{14}}}\right)\right)\%$$

where, $s^2_{t.eos}$, n_{treat}, $s^2_{c.eos}$ and $n_{control}$ represent the sample variances and group sizes for the treated and control groups at the end of study, respectively. The value of 2.048 comes from the Student's t-distribution with $n_{treat}+n_{control}-2$ degrees of freedom (where n_{treat} and $n_{control}$ are both equal to 15, giving rise to 28 degrees of freedom) with an upper-tail probability $\frac{\alpha}{2}$ for a significance level $\alpha\%$.

Again, the percentage estimate (1.2), is a specific case of the generalised percentage estimate (2.2). Now to obtain the corresponding confidence interval for this generalised percentage estimate, let $\mu_{c.eos}$ and $\mu_{t.eos}$ represent the population mean FECs at end of study for the control and treatment groups, respectively.

Now we make no assumptions about the distribution of the FECs at end of study for either the positive treatment group or negative control group, which we define as our random variables, $X_{t.eos}$ and $X_{c.eos}$, respectively. By letting

$Y = ln(X) = ln\left(\frac{X_{t.eos}}{X_{c.eos}}\right)$ then we can use the approximation (2.11) in order to obtain:

$$\begin{aligned} E[Y] &= E\left[ln\left(\frac{X_{t.eos}}{X_{c.eos}}\right)\right] \\ &= E[ln(X_{t.eos}) - ln(X_{c.eos})] \\ &= E[ln(X_{t.eos})] - E[ln(X_{c.eos})] \\ &\approx ln(\mu_{t.eos}) - ln(\mu_{c.eos}) \\ &\approx ln(T_i) - ln(C_i) \\ &= ln\left(\frac{T_i}{C_i}\right) \end{aligned}$$

where T_i and C_i are defined as the sample mean FECs at the end of study, on Day i, for the treatment and control groups, respectively.

Also, we can use approximation (2.12), derived earlier, to obtain

$$\begin{aligned} Var[Y] &= Var\left[ln\left(\frac{X_{t.eos}}{X_{c.eos}}\right)\right] \\ &= Var[ln(X_{t.eos}) - ln(X_{c.eos})] \\ &= Var[ln(X_{t.eos})] + Var[ln(X_{c.eos})] - 2Cov[ln(X_{t.eos}), ln(X_{c.eos})] \\ &\approx \frac{Var[X_{t.eos}]}{\mu^2_{t.eos}} + \frac{Var[X_{c.eos}]}{\mu^2_{c.eos}} - 2Cov[ln(X_{t.eos}), ln(X_{c.eos})] \\ &\approx \frac{s^2_{t.eos}}{T_i^2} + \frac{s^2_{c.eos}}{C_i^2} - 2Cov[ln(X_{t.eos}), ln(X_{c.eos})] \end{aligned} \tag{2.20}$$

where $s^2_{t.eos}$ and $s^2_{c.eos}$ represent the sample variances for the treated and untreated control groups, respectively, and $Cov[ln(X_{t.eos}), ln(X_{c.eos})]$ represents the co-variance between the ln-transformed, end of study counts of the positive treatment and negative control groups. In fact, the co-variance between two random variables, A and B say, can be defined as:

$$\begin{aligned} Cov[A,B] &= E[AB] - E[A]E[B] \\ &= \rho(A,B)\sqrt{Var[A]Var[B]}. \end{aligned}$$

where ρ is defined as the product-moment correlation coefficient.

It is worth noting that before *ln*-transforming counts, zero-valued counts would have to be handled in an appropriate manner in order to conclude on the covariance.

Some communications have assumed independence between the counts obtained from the control and treatment groups (Coles *et al.* 1992; Lyndal-Murphy *et al.* 2014), so approximation (2.20) becomes:

$$Var[Y] \approx \frac{s^2_{t.eos}}{T_i^2} + \frac{s^2_{c.eos}}{C_i^2},$$

This assumption is likely to be fair as different animals are involved between treatment and control groups, though non-independence might arise when data from many farms are combined to be studied as animals within a farm are more likely to be similar than animals in different farms.

In order to obtain a $100(1-\alpha)\%$ confidence interval, where $0 < \alpha < 1$, we extend the asymptotic result (2.14) to consider the difference between two population means, which in our case is essentially the natural logarithm of the ratio of two population means. The following interval (2.21) is thus obtained

$$ln\left(\frac{T_i}{C_i}\right) \pm Z_{(\frac{\alpha}{2})}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2} + \frac{s^2_{c.eos}}{n_{control}C_i^2}} \tag{2.21}$$

where $Z_{(\frac{\alpha}{2})}$ is the $\frac{\alpha}{2}$ percentile of the standard normal distribution (this is for a large sample approximation) and n_{treat} and $n_{control}$ are the number of animals in the treatment and control groups, respectively.

Now for the percentage estimate (2.2), the $100(1-\alpha)\%$ confidence interval can be defined as

$$100\left(1-\frac{T_i}{C_i}\exp\left(\pm Z_{(\frac{\alpha}{2})}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2}+\frac{s^2_{c.eos}}{n_{control}C_i^2}}\right)\right)\%.$$

In fact, we can obtain the following $100(1-\alpha)\%$ confidence interval

$$100\left(1-\frac{T_i}{C_i}\exp\left(\pm t_{(n_{treat}+n_{control}-2)}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2}+\frac{s^2_{c.eos}}{n_{control}C_i^2}}\right)\right)\%. \qquad (2.22)$$

where $t_{(n_{treat}+n_{control}-2)}$ is the $\frac{\alpha}{2}$ upper-tail probability for a Student's t-distribution with $n_{treat}+n_{control}-2$ degrees of freedom. The Student's t-distribution is used under the assumption of a normal distribution but for smaller sample sizes, such as those less than 30, the Student's t-distribution is more conservative as the limits are wider in comparison with the standard normal distribution being used in confidence interval (2.21),which produces more useful limits in large sample results. However, it is worth noting that if a percentage estimate of 100% was to be observed, then confidence interval (2.22) would not be defined since $T_i = 0$, which would be the case if an anthelmintic was 100% effective.

For the confidence interval (2.22) to match the confidence interval (1.3), we simply set n_{treat} and $n_{control}$ equal to the value of 15, in which case $t_{(n_{treat}+n_{control}-2)} = 2.048$. However, the value of 2.048 should not be regarded as fixed, i.e. $t_{(n_{treat}+n_{control}-2)}$ should change value for different values of the group sizes n_{treat} and $n_{control}$.

2.5.1.2 Percentage estimate derivation 2: pre- and post-treatment counts (treated groups of cattle involved only)

To derive the corresponding confidence interval for percentage estimate (2.3): consider the population means $\mu_{t.base}$ and $\mu_{t.eos}$ for the baseline and end of study FECs of a positive treated group, respectively. Our random variables are defined as the FECs at baseline and end of study for the positive treatment group, $X_{t.base}$ and $X_{t.eos}$, respectively, and let $Y = ln(X) = ln\left(\frac{X_{t.eos}}{X_{t.base}}\right)$.

In a similar way to deriving the approximations in Section 2.5.1.1, from approximations (2.11) and (2.12), we can derive the following:

$$\begin{aligned} E[Y] &= E\left[ln\left(\frac{X_{t.eos}}{X_{t.base}}\right)\right] \\ &= E[ln(X_{t.eos}) - ln(X_{t.base})] \\ &= E[ln(X_{t.eos})] - E[ln(X_{t.base})] \\ &\approx ln(\mu_{t.eos}) - ln(\mu_{t.base}) \\ &\approx ln(T_i) - ln(T_0) \\ &= ln\left(\frac{T_i}{T_0}\right) \end{aligned}$$

where T_0 and T_i are the sample means of the baseline and end of study FECs of a positive treated group, respectively.

Additionally,

$$\begin{aligned} Var[Y] &= Var\left[ln\left(\frac{X_{t.eos}}{X_{t.base}}\right)\right] \\ &= Var[ln(X_{t.eos}) - ln(X_{t.base})] \\ &= Var[ln(X_{t.eos})] + Var[ln(X_{t.base})] - 2Cov[ln(X_{t.eos}), ln(X_{t.base})] \end{aligned}$$

where $Cov[ln(X_{t.base}), ln(X_{t.base})]$ represents the co-variance between the ln-transformed counts at baseline and end of study for the positive treatment group.

It would not be reasonable to assume that there is independence between the counts at baseline and end of study, since faecal samples obtained on these days would be from the same experimental unit, i.e. the individual members of the cattle herd (in the case of assuming independence, we would have $Cov[ln(X_{t.base}), ln(X_{t.eos})] = 0$). As a result, we obtain the following approximation by using asymptotic approximation (2.12):

$$\begin{aligned} Var[Y] &\approx \frac{Var[X_{t.eos}]}{\mu^2_{t.eos}} + \frac{Var[X_{t.base}]}{\mu^2_{t.base}} - 2Cov[ln(X_{t.eos}), ln(X_{t.base})] \\ &\approx \frac{s^2_{t.eos}}{T_i^2} + \frac{s^2_{t.base}}{T_0^2} - 2Cov[ln(X_{t.eos}), ln(X_{t.base})]. \end{aligned}$$

where $s^2_{t.eos}$ and $s^2_{t.base}$ represent the sample variances for the treated group at end of study and baseline, respectively.

Therefore, the $100(1-\alpha)\%$ confidence interval can be defined using the asymptotic approximation (2.14):

$$100\left(1 - \frac{T_i}{T_0}\exp\left(\pm Z_{(\frac{\alpha}{2})}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2} + \frac{s^2_{t.base}}{n_{treat}T_0^2} - \frac{2Cov[ln(X_{t.base}), ln(X_{t.eos})]}{n_{treat}}}\right)\right)\%$$

where n_{treat} is the number of animals in the treatment group.

For finite n_{treat} we have the following $100(1-\alpha)\%$ confidence interval:

$$100\left(1 - \frac{T_i}{T_0}\exp\left(\pm t_{(n_{treat}-1)}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2} + \frac{s^2_{t.base}}{n_{treat}T_0^2} - \frac{2Cov[ln(X_{t.base}), ln(X_{t.eos})]}{n_{treat}}}\right)\right)\%. \tag{2.23}$$

where $t_{(n_{treat}-1)}$ is the $\frac{\alpha}{2}$ upper-tail probability for a Student's t-distribution with

$n_{treat} - 1$ degrees of freedom. Again, it is worth noting that if a percentage estimate of 100% was to be observed, then confidence interval (2.23) would not be defined since $T_i = 0$.

Mood *et al.* (1913) tell us that an approximation for the variance of a ratio of random variables is

$$\begin{aligned} Var[Y] &= Var\left[\frac{X_{t.eos}}{X_{t.base}}\right] \\ &\approx \left(\frac{\mu_{t.eos}}{\mu_{t.base}}\right)^2 \left(\frac{s^2_{t.eos}}{\mu^2_{t.eos}} + \frac{s^2_{t.base}}{\mu^2_{t.base}} - \frac{2Cov[X_{t.eos}, X_{t.base}]}{\mu_{t.eos}\mu_{t.base}}\right) \end{aligned} \tag{2.24}$$

and this resembles the form of variance that some communications present for this efficacy description (Lyndal-Murphy *et al.* 2014). However, to completely match the formulae in these communications, we would require

$$\left(\frac{\mu_{t.eos}}{\mu_{t.base}}\right)^2 = 1 \Leftrightarrow \mu_{t.eos} = \mu_{t.base}, \text{ since } \mu_{t.eos}, \mu_{t.base} \geq 0.$$

By the nature of the experiment we would not, however, expect and could not guarantee that $\mu_{t.eos} = \mu_{t.base}$ and so it would not be advisable to assume $\left(\frac{\mu_{t.eos}}{\mu_{t.base}}\right)^2 = 1$. As a well as this, the approximation (2.24) is not considering random variables being dealt with on the natural logarithmic scale.

2.5.1.3 Percentage estimate derivation 3: pre- and post-treatment counts (negative control and treated groups of cattle present)

To derive the corresponding confidence interval for percentage estimate (2.4); consider the population means $\mu_{t.base}$, $\mu_{t.eos}$, $\mu_{c.base}$, $\mu_{c.eos}$ for the baseline and end of study FEC counts for the positive treatment and negative control groups, respectively. Our random variables are defined as the FECs at baseline and end of study for the positive treatment group and negative control groups, i.e. $X_{t.base}$, $X_{t.eos}$, $X_{c.base}$ and $X_{c.eos}$, respectively, and let $Y = ln(X) = ln\left(\frac{X_{c.base}X_{t.eos}}{X_{c.eos}X_{t.base}}\right)$.

From the approximations (2.11) and (2.12), we derive the following:

$$\begin{aligned}
E[Y] &= E\left[ln\left(\frac{X_{c.base}X_{t.eos}}{X_{c.eos}X_{t.base}}\right)\right] \\
&= E[ln\left(X_{c.base}X_{t.eos}\right)] - E[ln\left(X_{c.eos}X_{t.base}\right)] \\
&= E[ln(X_{c.base})] + E[ln(X_{t.eos})] - E[ln(X_{c.eos})] - E[ln(X_{t.base})] \\
&\approx [ln(\mu_{t.eos}) - ln(\mu_{t.base})] + [ln(\mu_{c.base}) - ln(\mu_{c.eos})] \\
&\approx [ln(T_i) - ln(T_0)] + [ln(C_0) - ln(C_i)] \\
&= ln\left(\frac{C_0T_i}{C_iT_0}\right)
\end{aligned}$$

where C_0 is the sample mean of the FECs from the control group at baseline of study, i.e. Day 0.

Additionally,

$$\begin{aligned}
Var[Y] &= Var\left[ln\left(\frac{X_{c.base}X_{t.eos}}{X_{c.eos}X_{t.base}}\right)\right] \\
&= Var[ln(X_{t.eos}) - ln(X_{t.base}) + ln(X_{c.base}) - ln(X_{c.eos})] \\
&= Var[ln(X_{t.eos}) - ln(X_{t.base})] + Var[ln(X_{c.base}) - ln(X_{c.eos})] \\
&\quad - 2Cov[ln(X_{t.eos}) - ln(X_{t.base}), ln(X_{c.base}) - ln(X_{c.eos})].
\end{aligned}$$

Again, as in Section 2.5.1.1, if we assume independence between the treatment and control group then we obtain

$$\begin{aligned}
Var[Y] &= Var[ln(X_{t.eos}) - ln(X_{t.base})] + Var[ln(X_{c.base}) - ln(X_{c.eos})] \\
&= Var[ln(X_{t.eos})] + Var[ln(X_{t.base})] + Var[ln(X_{c.eos})] + Var[ln(X_{c.base})] \\
&\quad - 2Cov(ln(X_{t.eos}), ln(X_{t.base})) - 2Cov(ln(X_{c.eos}), ln(X_{c.base})).
\end{aligned}$$

Similarly as in Section 2.5.1.2, it would not be reasonable to assume that there is independence between the counts obtained from baseline and end of study, with respect to either the control or treatment groups since faecal samples obtained on these days would be from the same experimental units. As a result, $Cov[ln(X_{t.eos}), ln(X_{t.base})] \neq 0$ and $Cov[ln(X_{c.eos}), ln(X_{c.base})] \neq 0$, and so we obtain the following approximation for $Var[Y]$ in a similar fashion as to Section 2.5.1.2:

$$Var[Y] \approx \underbrace{\frac{s^2_{t.eos}}{T_i^2} + \frac{s^2_{t.base}}{T_0^2} - 2Cov[ln(X_{t.eos}), ln(X_{t.base})]}_{Variance1} + \underbrace{\frac{s^2_{c.eos}}{C_i^2} + \frac{s^2_{c.base}}{C_0^2} - 2Cov[ln(X_{c.eos}), ln(X_{c.base})]}_{Variance2}.$$

where $s^2_{t.eos}$, , $s^2_{t.base}$ represent the sample variances for the treated group at end of study and baseline, respectively. Also, $s^2_{c.eos}$, , $s^2_{c.base}$ represent the sample variances for the control group at end of study and baseline, respectively.

Therefore, the $100(1-\alpha)\%$ confidence interval for the percentage estimate (2.4) can be derived in a similar fashion to that of (2.22) in Section 2.5.1.1 and is defined as:

$$100\left(1 - \left(\frac{C_0 T_i}{T_0 C_i} \exp\left(\pm Z_{(\frac{\alpha}{2})} \sqrt{\frac{Variance1}{n_{treat}} + \frac{Variance2}{n_{control}}}\right)\right)\right)\%$$

where n_{treat} and $n_{control}$ are the number of animals in the treatment and control groups respectively (the number of animals at baseline and end of study should be the same with respect to each group).

For finite n_{treat} and $n_{control}$ we have the following $100(1-\alpha)\%$ confidence interval

$$100\left(1-\left(\frac{C_0T_i}{T_0C_i}\exp\left(\pm t_{(n_{treat}+n_{control}-2)}\sqrt{\frac{Variance1}{n_{treat}}+\frac{Variance2}{n_{control}}}\right)\right)\right)\%. \tag{2.25}$$

Again, it is important to note that if a percentage estimate of 100% was to be observed, then confidence interval (2.25) would not be defined since $T_i = 0$.

2.5.2 Involving the Bootstrap

When considering Bootstrapping, there are many ways of constructing confidence intervals for estimates. With reference to Efron and Tibshirani (1993) and Carpenter and Bithell (2000), in this Section we discuss some of the more well-known methods of forming these intervals - all of which can be obtained through R software through the *bootstrap* and *boot* packages (Kostyshak 2015; Canty and Ripley 2015; CRAN 2018).

2.5.2.1 Basic confidence intervals

The most basic $100(1-\alpha)\%$ confidence intervals that can be formed for an estimate $\hat{\theta}$ are those based on assuming a normal distribution is present, that is

$$\hat{\theta} \pm Z_{(\frac{\alpha}{2})}se[\hat{\theta}] \tag{2.26}$$

and

$$\hat{\theta} \pm t_{(n-1)}se[\hat{\theta}] \tag{2.27}$$

where $t_{(n-1)}$ is the $\frac{\alpha}{2}$ upper-tail probability for a Student's t-distribution with $(n-1)$ degrees of freedom. Note: we are basing the degrees of freedom on the

value of n since the original data set described earlier in Section 2.4.2, $\underline{X}$, has a sample size n.

The confidence interval (2.26) is formed based on the following asymptotic result involving the standard normal distribution with mean zero and variance one:

$$\frac{\hat{\theta} - \theta}{se[\hat{\theta}]} \sim \mathcal{N}(0, 1)$$

and is valid as $n \to \infty$ but is an approximation for finite samples, which is mainly what we would be dealing with in the Bootstrap framework. Hence, the confidence interval (2.27) is usually used in practice.

These confidence intervals do not account for any skewness in the data, or any other errors that can result when $\hat{\theta}$ is not the sample mean. As a result we consider the *Bootstrap-t* interval, which can adjust for these types of errors. The procedure in forming this type of interval involves estimating the distribution of Z directly from the data. When we consider the Bootstrap techniques described in Section 2.4.2, once we have generated B Bootstrap data sets $\{x_1^*, \ldots, x_B^*\}$ we then compute

$$Z^*(i) = \frac{\hat{\theta}_i^* - \hat{\theta}}{se[x_i^*]},$$

where $\hat{\theta}_i^*$ is the estimate for θ and $se[x_i^*]$ is the standard error corresponding to the *i-th* Bootstrap sample x_i^* (from $1 \leq i \leq B$).

Then the $\frac{\alpha}{2}$th percentile of $Z^*(i)$ is estimated by the value $t^*(\frac{\alpha}{2})$ such that:

$$\frac{\{Z_i^* \leq t_{\frac{\alpha}{2}}^*\}}{B} \leq \frac{\alpha}{2}$$

This value is also known as an *approximate pivot*, where the distribution obtained

is approximately the same for each value of θ and so we can obtain the following confidence interval:

$$(\hat{\theta} - t^*_{1-\frac{\alpha}{2}} se[\hat{\theta}], \hat{\theta} - t^*_{\frac{\alpha}{2}} se[\hat{\theta}]). \tag{2.28}$$

This interval is known as the *Studentised* interval in the *boot* and *bootstrap* packages in *R*/*RStudio*. It is also worth mentioning that in order to obtain an estimate for $se[\hat{\theta}]$; a *Double Bootstrap* must occur, i.e. one generated set of Bootstrapped data sets are obtained to estimate $\hat{\theta}$ and another set is obtained to estimate $se[\hat{\theta}]$, in order to eventually obtain the confidence interval (2.28). However if we were Bootstrapping for B Bootstrap replications say, then to obtain the Studentised interval we would require B^2 Bootstrap replications due to the double Bootstrap being present. This would be time consuming to compute.

In large samples, the coverage probability of the Bootstrap-t intervals tends to be closer to the desired level, in comparison to the coverage that the confidence intervals (2.26) and (2.27) can provide. However, this gain in accuracy is at a cost of generality according to Efron and Tibshirani (1993); confidence interval (2.26) can be applied to data sets of all sizes, confidence interval (2.27) can be applied to data sets of a finite sample size and the confidence interval (2.28) can be applied to the given data set of interest. Yet, the Bootstrap-t interval can be asymmetric about the value of zero and this is due to the improvement in the coverage offered by this interval.

It is also worth noting that the Bootstrap-t interval (2.28) is a generalisation of interval (2.27). The Bootstrap-t interval is not however *transformation-respecting*, i.e. the scale used to construct the interval does affect the interval itself and different scales work better than others. We can however use the Bootstrapping approaches, on the provided data, to estimate an appropriate transformation for the construction of the Bootstrap-t interval. In fact, these transformations approximately normalise and stabilise the variance of the estimate $\hat{\theta}$, the latter being regarded as more important (refer to Efron and Tibshirani (1993) for how to construct these *variance-stabilised Bootstrap-t* intervals).

2.5.2.2 Percentile methods

Percentile methods have proved popular in the veterinary and parasitological communities when constructing confidence intervals (Cabaret and Berrag 2004; Vidyashankar *et al.* 2007; Traversa *et al.* 2009; Lester and Matthews 2013; Lester *et al.* 2013). These intervals are mainly data-driven, empirical and depend much on the sampling distribution of the estimate $\hat{\theta}$, despite whether one is using a parametric and non-parametric Bootstrapping approach.

When we consider the Bootstrapping approaches described in Section 2.4.2, once we have obtained the sampling distribution $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$ for our estimate $\hat{\theta}$ an obvious choice of construction for a $100(1-\alpha)\%$ confidence interval is to read off the $\frac{\alpha}{2}$ and $(1-\frac{\alpha}{2})$ percentiles of the (ordered) sampling distribution. This is known as the *percentile* interval for the estimate $\hat{\theta}$ and is mathematically written as follows:

$$(\hat{\theta}^*_{(\frac{\alpha}{2})}, \hat{\theta}^*_{(1-\frac{\alpha}{2})}). \tag{2.29}$$

This interval (2.29) is simple to calculate (Carpenter and Bithell 2000) in comparison to other intervals discussed in this Section and is ideal when the number of Bootstrap replications is infinite; but in practice we can only work with a finite number of these replications and this is accounted for when constructing this interval based on the sampling distribution. If the Bootstrap sampling distribution is approximately normal, then the confidence intervals (2.26), (2.27), (2.28) and (2.29) will nearly agree; otherwise they will differ.

In fact based on Efron and Tibshirani (1993), the percentile interval described above is also transformation-respecting and also *range-preserving*; meaning if there are restrictions on the values that parameters can take, then the interval is able to respect and keep within or equate to these restrictive values. This is due to the fact that $\hat{\theta}$ and the values of the bootstrap estimates $\hat{\theta}_i^*$ will keep within the same restrictions as the parameter value θ.

However there are ways in which the confidence intervals discussed so far can

fail, such as the case of a bias estimator being present. We now go on to discuss if there are any modifications that can be made to the intervals that have been discussed so far that could possibly correct for these types of estimators being used in calculations.

2.5.2.3 Bias-Correction intervals

In this Section we consider a modification of the percentile interval namely the *Bias-corrected and accelerated* (BCA) interval, which can be calculated in the *boot* package in *R/RStudio*.

Again we refer to the sampling distribution of the estimate $\hat{\theta}$: $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$. The $100(1-\frac{\alpha}{2})\%$ BCA interval is defined as follows:

$$(\hat{\theta}^*_{\alpha 1}, \hat{\theta}^*_{\alpha 2}),$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{(\frac{\alpha}{2})}}{1 - \hat{a}(\hat{z}_0 + z_{(\frac{\alpha}{2})})}\right)$$

and

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{(1-\frac{\alpha}{2})}}{1 - \hat{a}(\hat{z}_0 + z_{(1-\frac{\alpha}{2})})}\right).$$

Here, $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z_{(\frac{\alpha}{2})}$ is the $100(1-\alpha)th$ percentile of a standard normal distribution.

$\hat{z}_0$ represents the bias-correction (this concept was explained earlier in Section 2.4.2), which is calculated from the proportion of the Bootstrap replications which

are less than the original estimate $\hat{\theta}$ and $\hat{a}$ represents the acceleration that can be described as the rate of change of the standard error associated with $\hat{\theta}$ with respect to the true parameter θ (measured on a normalised scale). In other words, $\hat{a}$ is proportional to the skewness of the sampling distribution $dist\{\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*\}$, estimated via a jackknife approach Efron and Tibshirani 1993. However, Carpenter and Bithell (2000) tell us that the calculation of $\hat{a}$ is tortuous, and therefore time consuming, particularly for complex parametric problems.

The BCA interval is also transformation-respecting and also more accurate in comparison to the other intervals (in fact, the BCA interval can be shown to be *second-order accurate*, but we do not show this here and refer the reader to Efron and Tibshirani (1993) for more details).

This indeed looks like a complicated interval in comparison to the other intervals discussed so far. But there are other confidence intervals such as the *Approximate Bootstrap Confidence* (ABC) intervals that can reduce the computation needed to calculate BCA intervals. BCA intervals also require a large number of Bootstrap replications. Efron and Tibshirani (1993) tell us that at least 1000 are needed in order to reduce sampling error.

It is also worth noting that even though the intervals described in this Section all have theoretical advantages and disadvantages one must also consider the nature of their data when it comes to choosing the most appropriate confidence interval to make inferences about the parameters of interest.

2.5.3 Credible intervals

As mentioned in Section 2.4.3, a variety of point estimates are often used to describe different characteristics of the posterior distribution such as central tendency, spread and correlation. However, these point estimates do not provide information on the skewness of the distribution or bi-modality for instance. To aid in providing this type of additional information, interval estimates can be obtained. A Bayesian $100(1-\alpha)\%$ credible interval is defined to be an interval that contains $100(1-\alpha)\%$ of the posterior distribution of the parameter of interest.

There are two types of credible intervals that are unique. There is the well-

known symmetric credible interval (symmetric with respect to probability), that is obtained by quoting the lower $\frac{\alpha}{2}$ quantile and the upper $(1-\frac{\alpha}{2})$ quantile of the posterior distribution, which is a similar approach to forming the Bootstrapped percentile interval (2.29) discussed in Section 2.5.2.2. There is also the *Highest Posterior Density Interval* which is the *shortest* $100(1-\alpha)\%$ credible interval and it obtains the region of the highest posterior density. In the case of unimodal and symmetric data this interval and the symmetric credible interval are identical; though they can dramatically differ when data are skewed. In general, it is preferred to use the symmetric credible interval since for this interval an easier interpretation can be concluded, is transformation-respecting and is easy to compute. Though in the presence of skewed data, one would be most likely to adopt the use of the Highest Posterior Density interval.

2.5.4 Frequentism vs. Bayesianism: differences in approaches and interpretations

When considering these approaches to statistics, the differences often stem from considering the context of probability. In classical statistics, probabilities are related to frequencies of events and so the interpretation is that parameters are fixed (but unknown) and data are random; whereas in the Bayesian paradigm probabilities are extended to cover the degree of certainty about statements and are related to a degree of subjectiveness and so the interpretation is that parameters are random and data are fixed.

Over and above this, there are subtle differences in the interpretations of intervals for parameters of interest. The $100(1-\alpha)\%$ credible interval is the Bayesian analogue of the confidence interval. According to Rice (2007) and VanderPlas (2014), a credible interval tells us that given the observed data, there would be a $100(1-\alpha)\%$ probability that the true value of the parameter, θ say, lies within the credible region. For example, a 95% credible interval for a parameter θ, tells us that we can be 95% sure that the parameter lies in the credible region obtained.

For a typical $100(1-\alpha)\%$ confidence interval, we would say that if an experiment was to be repeated many times; in $100(1-\alpha)\%$ of these cases the computed confi-

dence interval will contain the true parameter value θ. Here we make a statement of probability about the confidence interval itself given a fixed parameter value. For instance, a 95% confidence interval tells us the range in which the parameter, θ say, would occur 95% of the time with repeated sampling of the population.

2.6 Determining resistance with the FECRT

According to WAAVP, anthelmintic resistance is confirmed if the percentage estimate (1.2) is less than 95% and the associated lower confidence limit, of the confidence interval (1.3) is less than 90%. If only one of these criteria are met; anthelmintic resistance is suspected.

Vidyashankar *et al.* (2007) describes these thresholds as arbitrary for livestock species, though these thresholds have been widely accepted and adopted in relation to the other percentage estimates described in Section 2.3 and their respective confidence intervals reviewed in Section 2.5.1 (Torgerson *et al.* 2005; Dobson *et al.* 2012; Falzon *et al.* 2013; Lyndal-Murphy *et al.* 2014). For sheep and cattle FEC data, Lyndal-Murphy *et al.* (2014) suggests that as well as the thresholds recommended by Coles *et al.* (1992), a threshold for the upper confidence limit of the respective confidence interval, that is less than 95% should also be considered - especially if the percentage estimate is in the range of 90-95% for which resistance could be suspected.

Yet, for some livestock species, studies have used different thresholds for determining resistance. For example, Kaplan (2002) tells us that for horses, resistance to the *Benzimidazole* class of anthelmintics is declared if a percentage estimate is less than 90%. In addition, some equine studies have classed an anthelmintic to be effective, suspected resistance and ineffective if the FEC reduction is greater than 90%, between 80-90% and less than 80%, respectively. These are however thresholds for percentage estimates; there is no mention of thresholds for confidence limits for equine studies and so the WAAVP recommendations are adopted.

EMA regards the FECRT as an estimation of efficacy, and not confirmation of resistance (EMA 2014), since egg counts give an indirect measure of effi-

cacy/resistance. True resistance must be confirmed through laboratory slaughter studies, potentially supported by molecular level studies, or methods such as egg hatch tests (Vidyashankar *et al.* 2012) and so any interpretation from calculations as part of a FECRT would be considered with respect to *apparent* efficacy and resistance of livestock.

2.7 Discussion

As part of this review, experimental designs and considerations and the various statistical frameworks for which point and interval estimation can possibly be carried out for the conduct of the FECRT, have been examined. This examination has highlighted the need to investigate which experimental designs are appropriate to determine and classify apparent anthelmintic efficacy in a robust manner, with respect to statistical calculations and frameworks that can be adopted for the FECRT. With respect to statistical frameworks, derivations of confidence intervals using asymptotic approximations that assume normality of data, for the most commonly used percentage estimates, were presented. Bootstrap and Bayesian frameworks to carry out statistical calculations, as part of the FECRT, were also discussed.

This review highlighted the most commonly used percentage estimates that have been observed in the literature, as well as alternative percentage estimates and approaches, such as the Symmetrised Percentage Change and averaging over individual-based FEC reductions/changes, that could possibly be of use when conducting the FECRT. It is also worth mentioning that the percentage estimates currently being recommended and popularly utilised in the literature involve arithmetic means of treatment group FECs, which are the central tendency/maximum likelihood estimates for data assumed to be Poisson or follow a Negative Binomial distribution. Given that, in general, research and obtaining cattle FEC data has been limited in the past, it could be meaningful to consider if other appropriate discrete distributions, and their associated central tendencies, such as maximum likelihood estimators, would be better representative of the real cattle FEC data that this project provides.

The statistical frameworks for constructing corresponding intervals for these percentage estimates were also examined. When considering confidence intervals derived from asymptotic approximations, it would be meaningful to verify statistical assumptions that arise in the derivations presented for cattle FEC data (even upon transformation), i.e. normality of data. This is because communications have traditionally worked with these types of confidence intervals to conclude on the apparent anthelmintic efficacy/resistance status of livestock herds. With respect to Bootstrapping and Bayesian paradigms, the methodologies of obtaining point estimates and Bootstrapped confidence intervals and credible intervals were presented and have fairly recently been utilised as part of research conducted in anthelmintic studies. It could be meaningful to develop and compare the performance of methodologies within each framework to conclude on the percentage estimates examined as part of this review, and by extension the experimental designs, that provide robust statistical calculations for the conduct of the FECRT.

Chapter 3

Probabilistic distributions to represent cattle FECs

3.1 Introduction

In Chapter 2, it was highlighted that percentage estimates currently being recommended and popularly utilised in the literature, i.e. percentage estimates (2.2), (2.3) and (2.4) (with $i = 14$), involve arithmetic means of treatment group FECs, which are the central tendency/maximum likelihood estimates for data assumed to be Poisson or follow a Negative Binomial distribution (for which the Poisson distribution is a specific case).

When considering the probabilistic distribution of FECs, we would expect this to be Poisson with location parameter μ, which would be estimated by the arithmetic mean $\overline{x}$ (by maximum likelihood estimation). If we let the number of eggs per gram of faeces be a random variable Y, then the probability mass function (pmf) of the Poisson distribution, denoted as $PO(\mu)$, for the random variable Y is given by (3.1):

$$P(Y = y|\mu) = \frac{e^{-\mu}\mu^{y}}{y!}, \tag{3.1}$$

where $y = 0, 1, 2, \ldots$ and $\mu > 0$.

This assumption arises because we are dealing with discrete, count ecological data (Linden and Mantyniemi 2011) and would naturally assume that the eggs are randomly dispersed in the faeces. However, most hosts are uninfected with parasites (or contain very few parasites) and very few hosts are heavily infected; hence parasite populations are known to be statistically aggregated (Shaw and Dobson 1995; Levecke *et al.* 2012; Wang *et al.* 2017), as mentioned in Chapter 2. Denwood *et al.* (2008) tell us that due to this aggregation, FEC data and the associated number of parasites, would follow an *over-dispersed* Poisson distribution. As a result, for parasite populations in general, different parameterisations of the Negative Binomial distribution (an extension of the Poisson but still having a location parameter μ) are frequently used to represent the number of parasites in a given host in order to account for this statistical aggregation and some communications have used this distribution to represent FEC data (Wilson *et al.* 1996; Morgan *et al.* 2005; Levecke *et al.* 2012). However, how valid is it to assume that cattle FEC data follow a Negative Binomial distribution, and hence use the location parameter μ in FECRT calculations? Also, if there are any other distributions that better represent cattle FEC data, then how do FECRT calculations compare when these distributions' location parameters are utilised in comparison to using the location parameter μ, which is currently recommended to be used in FECRT calculations?

As well as this, the associated confidence intervals (2.22), (2.23) and (2.25) for their respective percentage estimates (2.2), (2.3) and (2.4) (with $i = 14$) described in Section 2.5.1, were derived in Chapter 2 using an asymptotic approximation which assumes the natural logarithm of FEC data and the original FEC data being normally distributed. But again, are these assumptions valid?

In this Chapter, the above shall be investigated using the data described in Section 1.6.1.

3.2 Validity of normality assumption for confidence intervals

3.2.1 Data transformations

A random variable X, such as FECs, may be transformed by some function $g(\cdot)$ to define a new random variable Y. In this case, $g(\cdot)$ is referred to as a *transformation* of the random variable X (Mood *et al.* 1913; Upton and Cook 2011). There are many transformations that exist for both continuous and discrete random variables, however, most are employed mainly to correct for skewness in an attempt to make the distribution of the data more symmetric, i.e. (at least) approximately normal. As a result, one is able to quote the arithmetic mean as the central tendency of the distribution, rather than other location parameters such as the median or the mode.

According to Newton and Rudestam (2012), one type of transformation that can be used to reduce skewness, condense outliers and condition the distribution of data to approximate the normal curve are *power transformations*, i.e. $Y = X^p$ for some power p. In fact in the biological sciences, it is most common to use transformations for which zero values (if any) are easily identifiable on the transformed scales (Zar 1996) such as the square-root transformation. However, one of the biggest limitations of transforming data is the possible challenge of interpreting them (Newton and Rudestam 2012).

Transformations of FEC data have been considered by several authors, for example $log_{10}(\cdot)$, the natural logarithm $ln(\cdot)$, square-root and arcsine transformations (Fulford 1994; Pook *et al.* 2002; Mejia *et al.* 2003; Torgerson *et al.* 2005; Vidyashankar *et al.* 2007; Dobson *et al.* 2009). Transformations such as $ln(x+a)$, where $a > 0$, have also been considered in order to take account of FEC data with values of zero (Torgerson *et al.* 2005).

It is thus of interest, not only to assess the normality of counts that have been transformed using some variation of the natural logarithmic function (in order to guarantee that the sampling distribution of the ratio of means on the appropriate scale is asymptotically normal), but also to observe whether or not other trans-

formed versions of FECs could prove useful to obtain data of a normal nature. If this were the case, it could be possible to utilise the Delta Method and provide asymptotic results of the sampling distribution of the ratio of means, as in Section 2.4.1.2, with respect to the appropriate transformation scale.

3.2.2 Null hypothesis significance testing

A statistical hypothesis is a statement about a statistical parameter, θ say, or about the nature of the probability distribution of a random variable. According to Upton and Cook (2011), hypothesis testing is a procedure for deciding between two hypotheses on the basis of the value of a *test statistic*, which is a function of the observations in a random sample, based on the Neyman-Pearson lemma (Rice 2007).

A hypothesis is considered *simple* if the hypothesis considers one parameter value θ or a single (fully specified) probability distribution, otherwise it is a *composite* hypothesis. In a test considering the value of an unknown parameter, θ, the *null hypothesis*, usually denoted as H_0 specifies a particular value for the parameter θ, whereas the *alternative hypothesis*, denoted as H_1, specifies either an alternative value or a range of alternative values. For example, a typical null hypothesis might state that the population mean $\mu = 50$. The alternative hypothesis may be that $\mu < 50$ and this is classified as a *one-tailed* hypothesis test. Conversely, if an alternative hypothesis such as $\mu \neq 50$ was stated then the hypothesis test is considered to be *two-tailed*.

Upton and Cook (2011) tell us that the probability of obtaining a value for the test statistic that is as extreme, or more extreme, than when H_0 is true is called a *p-value*, which can be interpreted as the probability that one's results have occurred through chance. If the actual value of the statistic is too far from its expected value, i.e. the probability that the results obtained have occurred by chance is extremely low, then the test is deemed to be *significant* and the decision is to reject H_0. However, if the value of the statistic is close to its expected value the test is considered to be *non-significant* and we fail to reject H_0. The set of values of the statistic that lead to the rejection of H_0 is called the *critical/rejection*

region.

When considering the outcomes of null hypothesis significance testing procedures, there are two cases when the test leads to the correct result, i.e. when H_0 is in fact true and the test leads to the failure of its rejection or when H_1 is indeed true and the test procedure leads to the rejection of H_0 (Rice 2007). Similarly, there are two cases when the testing procedure leads to an incorrect result, namely H_0 is true but the test leads to the rejection of H_0, i.e. a *Type I* error, or when H_1 is true and the test leads to a failure of the rejection of H_0, i.e. a *Type II* error.

Upton and Cook (2011) also tell us that the size of the critical region is determined by the desired *significance level*, denoted as α, which is in fact the probability of making a Type I error - it is this which is pre-determined before carrying out the hypothesis testing procedure by the researcher and is usually set to 5% (Cowles and David 1982). The smaller the value of α, the smaller the size of the critical region. Furthermore, the probability of a Type II error is often denoted as β and it is this value which contributes to the *power* of the test, which is defined to be $1 - \beta$ i.e. the probability that one favours the alternative hypothesis when it is in fact true, and is often desired to be of at least 80% (Gardiner and Gettinby 1998). If there is a choice of test with pre-determined α, it is usual to choose the test (if one exists) which maximises the power of the test (Rice 2007; Newton and Rudestam 2012).

3.2.3 Assessing normality of original and transformed FECs

In our study, using the data described in Section 1.6.1, an assessment of normality was conducted by the use of a *Shapiro-Wilk* normality test at the 5% significance level, in which the null hypothesis is that the data follow a normal distribution (Royston 1982a; Royston 1982b; Royston 1995; Razali and Wah 2011). Each original data set was assessed for normality and the $ln(x+1)$ (where x is defined as a FEC), the square-root and $x^{\frac{2}{3}}$ power transformations were also applied and assessed for normality. These transformations were chosen in order to account for any zero-valued counts present and could therefore be easily identifiable on the transformed scales.

	Original Data		ln(x+1) Data		Square-Root Transformed Data		$\frac{z}{x^3}$ Transformed Data	
	Day 0	Day 14	Day 0	Day 14	Day 0	Day 14	Day 0	Day 14
Data sets that were considered normal	38 (12.50%)	12 (3.90%)	78 (25.70%)	24 (7.90%)	154 (50.70%)	41 (13.50%)	104 (34.20%)	29 (9.50%)
Data sets that were considered non-normal	266 (87.50%)	285 (93.80%)	226 (74.30%)	273 (89.80%)	150 (49.30%)	256 (84.20%)	200 (65.80%)	268 (88.20%)
Data sets that were inconclusive	0 (0%)	7 (2.30%)	0 (0%)	7 (2.30%)	0 (0%)	7 (2.30%)	0 (0%)	7 (2.30%)

Table 3.1: *Shapiro-Wilk* Normality test results for Day 0 and Day 14 data and the various transformations applied to these data

In the case of an inconclusive result being obtained from the normality test, this would be due to either the relevant data set being of too small a sample size for the test to be conducted or the counts present in the data set being all of the same value.

3.2.4 Results

Table 3.1 highlights the normality results of the original 304 Day 0 and Day 14 FEC data and their transformed versions. Of the original Day 0 and Day 14 FEC data, 87.5% and 93.8% of these data sets, respectively, were classed as non-normal. With respect to the transformed versions of these data, apart from the square-root transformed Day 0 data, the majority of the transformed versions of the Day 0 and Day 14 FEC data were also considered to be non-normal. As a result, it would not be valid to assume that FEC data, even upon transformation, follow a normal distribution.

3.3 Investigating distributions to be used for representing cattle FEC data

3.3.1 Issues with discrete count data

3.3.1.1 Over-dispersion of count data

Over-dispersion is the term given to discrete count data, such as FECs, where the variability present between counts is far larger than would be expected. In the case of the Poisson distribution, we expect the variability in count data to be the equivalent of the location parameter μ (Zuur *et al.* 2009), that is $E[Y] = Var[Y] = \mu$. According to (Rigby *et al.* 2014) however, very often when dealing with count data $E[Y] = \phi Var[Y]$, where $\phi > 0$ and is referred to as the *dispersion parameter* (Zuur *et al.* 2009). If $\phi > 1$ then we say the count data is over-dispersed and it is suggested that a distribution other than the Poisson should be used to represent the data. However, if ϕ is close or equal to the value of one, resulting in $E[Y] \approx Var[Y]$ then this would suggest that the Poisson distribution is an appropriate representation for the data. For $\phi < 0$, then we would say the data is under-dispersed.

Various solutions to the problem of over-dispersion in count data have been suggested (Consul and Famoye 1989; Dossou-Gbété and Mizère 2006), one of which is to assume a random effect at the observation level (Rigby *et al.* 2014), i.e. the use of *compound distributions* and this will be further discussed in Section 3.3.2 when considering FEC data.

3.3.1.2 Excess zeros in count data

Another problem that can appear when dealing with count data is the excess or shortage of zero counts than expected from a Poisson distribution (more often the former case). Zuur *et al.* (2009) tell us that in the case of an excess of zeros in count data, then this data is referred to as *zero inflated*.

In the presence of zero inflation, zeros are considered to comprise of two groups:

the first containing only zeros that are the false/excess zeros which have occurred due to experimental design, survey or observer errors. This group of zeros is called the observations with zero mass or *structural* zeros. The second group of zeros comes from the appropriate discrete count distribution which is assumed (and this distribution is dependent on the dispersion of the count data) and these zeros can be classed as *sampling* zeros. It is worth noting here that we do not know which of these two groups the zeros belong too, we just assume that these two groups co-exist.

In the case of dealing with FEC data, it is likely that the choice of counting technique would influence the presence of excess zeros, since for diagnostic sensitivities with values greater than 1 epg, zero eggs being found may not necessarily correspond to no eggs being present; it is more likely to mean that the counting technique is not precise enough to be able to detect any eggs that could in fact be present. Although, excess zeros could also be the result of hosts being uninfected with any parasites, resulting in no eggs being obtained as part of sampled faeces.

3.3.2 Compound distributions

Compound distributions are distributions that result from allowing their associated parameters to vary, i.e. follow other distributions (Upton and Cook 2011). According to (Rigby *et al.* 2014), these types of distributions can account for over-dispersion present in count data.

Assume that the conditional distribution of the response variable Y, such as FECs, is a discrete probability function $P(Y = y|\Gamma = \gamma)$, given $\Gamma = \gamma$, a value of a continuous random effect variable Γ which has probability density function $f(\gamma)$. The marginal probability function of Y, which results in a *continuous mixture of discrete distributions*, is thus given by (3.2)

$$P(Y = y) = \int P(Y = y|\Gamma = \gamma) f(\gamma) d\gamma. \tag{3.2}$$

If however the random effect variable Γ has a discrete probability function $P(\Gamma =$

γ), then the resulting probability distribution of Y is called a *discrete mixture of discrete distributions*, and is given by (3.3):

$$P(Y = y) = \sum P(Y = y|\Gamma = \gamma)P(\Gamma = \gamma). \tag{3.3}$$

For example, $Y|\gamma \sim PO(\mu\gamma)$ where PO denotes the Poisson distribution and $\gamma \sim GA(1, \sigma^{\frac{1}{2}})$, for which GA denotes the Gamma distribution and $\sigma > 0$, gives rise to the marginal distribution that is the Negative Binomial distribution that has parameters mean μ and scale parameter σ – this distribution is also referred to as the Gamma-Poisson distribution (Denwood 2010).

Additionally, these types of distributions can also solve the problem of having excess zeros in count data, i.e. by considering special cases of compound distributions known as zero inflated and zero-adjusted distributions. Subsections 3.3.2.1 and 3.3.2.2 give brief definitions of these distributions.

3.3.2.1 Zero inflated distributions

In general, for a zero inflated distribution (ZID) we consider a discrete response variable Y that can exhibit a greater probability of value zero than that of a certain discrete distribution, Y_1 say. That is $P(Y = 0) > P(Y_1 = 0)$ and thus we have that $Y \sim ZID(\cdot)$, which is a discrete mixture of two components, namely (3.4):

$$\begin{aligned} P(Y = 0) &= \nu + (1 - \nu)P(Y_1 = 0) \\ P(Y = y) &= (1 - \nu)P(Y_1 = y), \end{aligned} \tag{3.4}$$

where $0 < \nu < 1$ and $y = 1, 2, 3, ...$ (Zuur *et al.* 2009; Rigby *et al.* 2014).

The parameter ν is defined as the proportion of structural zeros present, and is subsequently estimated as the proportion of zeros in count data. This distribution is a discrete mixture of two components: value zero with probability ν and the discrete counting distribution Y_1 involved with probability $(1 - \nu)$.

In fact, Bohning *et al.* (1999) tell us that the central tendency for a ZID, is the maximum likelihood estimator μ_1, say, and is derived as

$$\mu_1 = \frac{\mu}{(1-\nu)}. \tag{3.5}$$

The parameter μ_1 given in equation (3.5) of the zero inflated distribution can be interpreted as the ratio of two variables; the parameter μ, estimated as the arithmetic mean of the data, and the proportion of non-structural zero counts present in the data. Thus, μ_1 can be thought of as a re-scaling of the arithmetic mean parameter μ with respect to the proportion of non-structural zero counts present to a value representative of the discrete counting distribution. This re-scaled value will be greater than that of the value of the arithmetic mean μ, since:

$$\begin{aligned} & \nu > 0 \\ \Leftrightarrow\ & 1-\nu < 1 \\ \Leftrightarrow\ & \frac{1}{1-\nu} > 1 \\ \Leftrightarrow\ & \frac{\mu}{1-\nu} > \mu \\ \Leftrightarrow\ & \mu_1 > \mu. \end{aligned}$$

Hence, the parameter μ_1 can take account of non-structural zero counts and higher-valued data points present in zero inflated data, which may be less accounted for when locations such as the arithmetic mean are used. This is due to zero inflation of potential excess zeros decreasing these values to that of a value closer to zero. Therefore, we are essentially trying to correct/compensate for the presence of zero inflation by using μ_1 as an estimate of central tendency/location of the discrete count distribution involved.

When we consider μ_1 as a ratio, there are two scenarios which may result:

• μ_1 is large, due to either μ being large (due to outliers present in the data) and/or ν being large (due to the proportion of zeros being high, as expected),
• or the less likely scenario in the context of zero inflation, that μ_1 is small, which involves either μ being small or ν being small (in this case we would likely be considering a data set with low-valued counts).

3.3.2.2 Zero adjusted distributions

Zero adjusted distributions ($ZADs$) are used when a discrete response variable Y can exhibit either a greater or less probability of value zero than that of a particular discrete distribution, Y_2 say. We say that $Y \sim ZAD(\cdot)$ which is a discrete mixture of two components (Rigby *et al.* 2014),

$$P(Y = 0) = \nu$$
$$P(Y = y) = \frac{(1 - \nu)P(Y_2 = y)}{1 - P(Y_2 = 0)},$$

where y and ν are defined as before in Subsection 3.3.2.1.

Rigby *et al.* (2014) tell us that $P(Y = 0)$ can be greater or less than $P(Y_2 = 0)$ and hence the distribution is called zero-adjusted. If $\nu = P(Y_2 = 0)$, then $Y = Y_2$ and the ZAD becomes the discrete distribution Y_2. If $\nu > P(Y_2 = 0)$, then the zero adjusted distribution is just a re-parameterisation of the zero inflated distribution defined in (3.4). Hence the zero inflated distribution is a re-parameterised sub-model of the zero adjusted distribution. If $\nu < P(Y_2 = 0)$, however, then the zero adjusted distribution becomes the discrete count distribution Y_2 but with ν less than $P(Y_2 = 0)$, i.e. the proportion of zeros we would expect under the given discrete counting distribution Y_2.

3.3.3 Fitting distributions to count data

3.3.3.1 General Additive Models for Location, Scale and Shape (GAMLSS) package

The General Additive Models for Location, Scale and Shape (GAMLSS) package available in *R/RStudio* (Rigby and Stasinopoulos 2005; Stasinopoulos and Rigby 2007) is flexible in the fitting of distributions in the sense that it can consider distributions that may not necessarily belong to the exponential family, e.g. compound distributions. This package is also able to estimate relevant parameters other than the location of a given distribution. The $gamlssML(\cdot)$ function utilises maximum likelihood estimation, for parameters such as location, through a non-linear maximisation algorithm, i.e. the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) algorithm (Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970). This algorithm is the same as the one used in the $mle(\cdot)$ function, which is available in the *stats4* package in *R/RStudio*'s system library. The BFGS algorithm is the most popular *Quasi-Newton* method for carrying out optimisation for the log-likelihood function, mainly due to it being a gradient descent method which does not involve the necessity of evaluating second derivatives that feature as part of the Hessian matrix for the maximum likelihood estimation. As a result, this package and the $gamlssML(\cdot)$ function were utilised for fitting discrete count distributions to the cattle FEC data, as described in Section 1.6.1 using *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.).

The $gamlssML(\cdot)$ function also gives us the option of specifying *sensible* initial values for parameters before carrying out the maximum likelihood process, otherwise the algorithm picks a random number as its initial value. In using this function, intuitive starting values for the location parameter μ were used as the arithmetic mean of the data, except for in the instance of zero altered or inflated distributions. In these cases initial starting values were given as the estimated location parameter μ_1 in equation (3.5). For the scale parameter σ, an initial value was specified as the standard deviation of the data and the parameter ν for the zero altered and inflated distributions was assigned an initial value estimated by the proportion of zeros in the data. Some distributions fitted also involved skew-

ness parameters, for which an initial value of $(\frac{\mu}{\sigma})^3$ was estimated from the data. However, if no distributions could be fitted to a particular data set due to the counts being all of the same value, namely the value of zero, the fit was recorded and classed as *inconclusive*, and this was also summarised, if appropriate. The distributions available (Wimmer and Altmann 1999; Johnson *et al.* 2005; Rigby *et al.* 2014) to be fitted to discrete count data are explained in more detail as follows.

3.3.3.2 Distributions available to be fitted using GAMLSS

For count data, this package offers seventeen potential discrete distributions. These include: the Poisson distribution given by (3.1), denoted as $PO(\mu)$ and two parameterisations of the Negative Binomial distribution type I and type II, denoted as $NBI(\mu, \sigma)$ and $NBII(\mu, \sigma)$ respectively. The main difference between these two distributions is that the former is derived using a Gamma distribution $GA(1, \sigma^{\frac{1}{2}})$ and the latter uses a Gamma distribution $GA\left(1, \left(\frac{\sigma}{\mu}\right)^{\frac{1}{2}}\right)$ but both have a maximum likelihood estimator (mle), and hence central tendency, μ. In addition, $NBI(\mu, 1)$ is the equivalent to the Geometric distribution, denoted as $GEOM(\mu)$, and this distribution can also be fitted.

The Poisson inverse-Gaussian, Sichel and Delaporte distributions can also be fitted - denoted as $PIG(\mu, \sigma)$, $SICHEL(\mu, \sigma, \nu)$ and $DEL(\mu, \sigma, \nu)$ respectively - and these are further examples of continuous mixtures of discrete distributions described by equation (3.2), where the discrete distribution involved is the Poisson distribution and $f(\gamma)$ takes the form of an inverse-Gaussian ($IG(1, \sigma^{\frac{1}{2}})$), generalized inverse-Gaussian ($GIG(1, \sigma^{\frac{1}{2}}, \nu)$) and a shifted Gamma mixing distribution ($SG(1, \sigma^{\frac{1}{2}}, \nu)$) respectively. These distributions have a mle, and hence central tendency, μ.

With respect to zero inflated distributions, based on the definition given by equation (3.4), $Y_1 \sim PO(\mu)$ and the $gamlssML(\cdot)$ function can fit two different types of zero inflated Poisson distributions: one denoted as $ZIPI(\mu_1, \nu)$, which has mle μ_1, and $ZIPII(\mu, \nu)$, which is re-parameterised to have mle μ (described by Rigby *et al.* (2014)). In addition it is also possible to fit a zero inflated Poisson

inverse-Gaussian distribution by letting $Y_1 \sim PIG(\mu, \sigma)$ and this is denoted as $ZIPIG(\mu_1, \sigma, \nu)$ and a zero inflated Negative Binomial (type I) distribution can also be fitted, where $Y_1 \sim NBI(\mu, \sigma)$ and this is denoted as $ZINBI(\mu_1, \sigma, \nu)$ both having mles μ_1 that can be estimated.

For zero adjusted distributions, we are able to fit a zero adjusted Poisson and zero adjusted Negative Binomial distributions and these distributions are denoted as $ZAP(\mu_1, \nu)$ and $ZANBI(\mu_1, \sigma, \nu)$, respectively, both having mles μ_1.

There are another four distributions that the $gamlssML(\cdot)$ function is able to fit to count data, namely the logarithmic, zero adjusted logarithmic, Yule and Waring distributions. However, the latter two distributions consist of shape parameters being estimated only, not parameters of location, and the Logarithmic distribution is the limiting distribution of the ZANBI distribution, which we consider already. It is for these reasons we do not consider these distributions here but refer the reader to Wimmer and Altmann (1999), Winkelmann (2000) and Johnson *et al.* (2005) for further information on these distributions. Table 3.2, gives a summary of the distributions fitted and their associated parameters.

3.3.4 Distribution selection methods

3.3.4.1 Hypothesis test based selection methods

For ecological and biological data, there exist various methods for selecting distributions of best representation and fit. In general, selection methods strive to find a balance between complexity/flexibility and parsimony, i.e. wanting to consider more complex distributions but only if it is worthwhile in doing so. It is worth noting that the terms *distribution* and *model* are often used in communications interchangeably, and throughout this Section the former term is utilised.

One way to determine whether or not a distribution is of adequate fit is by using the χ^2 goodness-of-fit test. This is a one-tailed hypothesis test procedure where the null hypothesis, H_0 say, assumes that the data do follow the distribution under consideration (Weiss and Hasset 1991; Bolker 2008; Upton and Cook 2011). According to Weiss and Hasset (1991), if we assume the sample size involved is

Distribution (denoted as)	Parameters
Poisson (PO)	μ
Geometric (GEOM)	μ
Negative Binomial Type I (NBI)	μ, σ
Negative Binomial Type II (NBII)	μ, σ
Poisson inverse-Gaussian (PIG)	μ, σ
Sichel (SICHEL)	μ, σ, ν
Delaporte (DEL)	μ, σ, ν
Zero Inflated Poisson Type I (ZIPI)	μ_1, ν
Zero Inflated Poisson Type II (ZIPII)	μ, ν
Zero Inflated Poisson inverse-Gaussian (ZIPIG)	μ_1, σ, ν
Zero Inflated Negative Binomial Type I (ZINBI)	μ_1, σ, ν
Zero Adjusted Poisson (ZAP)	μ_1, ν
Zero Adjusted Negative Binomial Type I (ZANBI)	μ_1, σ, ν

Table 3.2: Distributions fitted and their parameters

large, and consider a significance level α, then if H_0 is true, the random variable is given by

$$\chi^2 = \frac{\sum(O-E)^2}{E} \sim \chi^2_{k-1}$$

where O and E represent the observed and expected frequencies of the data, respectively and the term $k-1$ represents the degrees of freedom which is interpreted as one less than the number of categories in the distribution.

In general, each expected frequency is computed as $E = np$ where n is the sample size and p is the relative frequency of the categories of the data involved. The assumptions of this hypothesis test procedure include all expected frequencies

being of a greater value than one and at most 20% of the expected frequencies being of a value less than five. However, it is not always possible to satisfy these assumptions in practice and it is worth noting that this hypothesis test procedure is essentially an asymptotic result, since a large sample is assumed to be worked with and the random variable χ^2 is said to *approximately* follow a χ^2 distribution. In fact, Bolker (2008) tell us that this test works only for *simple* count data and if data are continuous or over-dispersed then the χ^2 goodness-of-fit test is no longer useful as another parameter describing the variance is present. Upton and Cook (2011) also refers us to using the non-parametric test that is the *Kolmogorov-Smirnov* test, which assesses the null hypothesis that a random sample has been drawn from a specified distribution, however, this test is applicable mainly for considering the fit of continuous distributions (Birnbaum and Tingey 1951; Conover 1971; Durbin 1973; Sprent 1989; Marsaglia *et al.* 2003).

Another method of selecting distributions is by means of the likelihood ratio test (LRT) which was originally developed to compare nested distributions (Newyman and Pearson 1928) but can also be extended to assess non-nested distributions under the asymptotic assumption that the sample size tends to infinity (Cox and Hinkley 1974). For example, Zuur *et al.* (2009) tell us this test can be used to assess whether a Poisson or Negative Binomial distribution are more appropriate to describe a discrete, count response since the Poisson distribution is a nested distribution of the Negative Binomial and this test can also be utilised when comparing the zero inflated versions of these distributions.

Within the GAMLSS framework, the LRT is carried out using a fitted distributions *global deviance*, denoted as GD, which is essentially $-2l(\hat{\theta})$, where $l(\cdot)$ is the log-likelihood function and $\hat{\theta}$ represents the estimated parameters associated with the relevant distribution. According to Stasinopoulos *et al.* (2015), the GD is different from the deviance that is considered with generalised linear and generalised additive modelling. The global deviance is *exactly* minus twice the log-likelihood function, including all constant terms in the log-likelihood, whereas in the other modelling frameworks the deviance is calculated as a deviation from the simpler model and it does not include *constant* terms in the fitted log-likelihood.

As a result let D_0 and D_1 be two different distributions with fitted global deviances GD_0 and GD_1, respectively such that D_0 is nested within D_1, with respective degrees of freedom df_0 and df_1. Then we consider the statistic

$$\Lambda = GD_0 - GD_1$$

which has an asymptotic χ^2 distribution, under the null hypothesis that the correct distribution for representation is D_0 with degrees of freedom $df_0 - df_1$. However, as with the χ^2 goodness-of-fit test, the LRT is essentially an asymptotic result, meaning the random variable Λ will approximately follow a χ^2 distribution for a finite sample.

One final hypothesis test based selection method worth mentioning is the *vuong*'s test, which is essentially a likelihood ratio test for non-nested distributions (Vuong 1989), though it has recently been commented on it's misuse for the assessment of zero inflation for discrete count data (Wilson 2015). Again, however, this test is essentially an asymptotic result.

The hypothesis test procedures described here are frequently used with the intention of selecting a probability distribution that can best represent one's data. However, it has been repeatedly mentioned that these procedures are carried out at an asymptotic level, meaning large samples are assumed to be worked with - but this can not always be guaranteed when working with real data. Assuming this is achieved, the random variables involved with these procedures are also said to be approximately distributed with various forms of the χ^2 distribution. As a result, it would be a natural response to be concerned about how adequate these approximations are, on the grounds that this could potentially result in various type I or type II errors occurring when carrying out these hypothesis test procedures. It is for these reasons that no hypothesis test procedures were used in the process of selecting the best-fitting distributions in this part of the project, and as a result information criteria methods were considered and are explained in more detail as follows.

3.3.4.2 Information criteria based selection methods

An alternative distribution selection strategy is through the use of information criteria, where we are able to evaluate these criterion for every distribution fitted and are able to compare these evaluations, despite whether or not the distributions are nested, to conclude on distributions of best fit. For example, the *Akaike's Information Criterion* (AIC) is frequently referred to when dealing with selections (Akaike 1973; Bolker 2008; Zuur *et al.* 2009; Upton and Cook 2011; Zuur *et al.* 2012). This is defined as

$$AIC = -2(l) + 2(df) \tag{3.6}$$

where l represents the fitted log-likelihood function and df represents the associated degrees of freedom, which is the number of associated parameters of the distribution being fitted.

By observing the AIC, the distribution fit is being accounted for through the $2l$ term, whilst at the same time being penalized for complexity through addition of the $2df$ term. The AIC value (3.6) strives for balance between flexibility/complexity (rewarding the distribution being fitted, at least indirectly, as more complex distributions make it easier to explain the data involved and hence lead to a decrease in the value of $-2l$) and parsimony, i.e. consider more complex distributions only if it is worthwhile doing so, by penalizing the fit through the penalty term $2df$. Distributions which exhibit the lowest AIC values are considered to be of the best fit.

Another criterion to consider is the *Schwartz Bayesian Criterion* (Schwarz 1978), which is also referred to as the *Bayesian Information Criterion* (BIC), which is defined as

$$BIC = -2(l) + ln(n)(df) \tag{3.7}$$

where l is defined as before, $ln(\cdot)$ is the natural logarithmic function and n is the sample size of the data involved. In a similar way to the AIC value (3.6), minimum values of the BIC value (3.7) indicate distributions of the best fit, however the penalty term is harsher for the inclusion of the $ln(n)$ component. As a result, the use of the BIC will usually result in the simpler distribution being favoured, whereas the AIC will favour the more complex.

These information criteria were derived under different assumptions and are useful in different settings. The AIC was derived under the assumption that a true model requires an infinite number of parameters (Akaike 1973) and attempts to minimize the information lost by using a given finite dimensional model to approximate this. The BIC however, was derived as a large sample approximation to Bayesian selection among a fixed set of finite dimensional models (Schwarz 1978), resulting in it being consistent for selection among a fixed family of models (where one is assumed to be the true model).

One final criterion worth mentioning is the *Deviance Information Criterion* (DIC), which (Denwood 2010) tell us is used for evaluating the fit of distributions/models within the Bayesian framework (Spiegelhalter *et al.* 2002).

All the criteria discussed are valid methods for selecting distributions of best fit. Within the GAMLSS package, the function $gamlssML(\cdot)$ evaluates both the AIC and BIC, since maximum likelihood estimation is used as part of the fitting process. Additionally, communications referring to the use of the GAMLSS package, base their decisions on selecting distributions/models using the AIC value. It is also worth noting however when comparing these two criteria, the BIC was derived under conditions for which one candidate model is assumed to be the true model - but given the vast number of distributions that can be obtained when considering compound distributions and the finite, available selection of distributions for which the GAMLSS has to offer; it would be unfeasible to assume in every case that one distribution is the true distribution of the observed data - unless the number of distributions to be considered is indeed the entire number of distributions that can be evaluated, but this is would be unachievable in reality.

Hence, in the interest of investigating if more complex distributions give a better representation of FEC data, the AIC value was chosen to be evaluated for each

distribution fitted and the distribution with the minimum AIC value for each data set was recorded.

3.3.5 Results

A total of 304 sets of FEC data were obtained from Day 0 and a summary of the frequencies of the best-fitting distributions for each of the diagnostic sensitivity groupings are shown in Table 3.3. With respect to the Day 14 data, a total of 304 sets of FEC data were also considered and a similar summary of the frequencies of best-fitting distributions for each diagnostic sensitivity grouping is displayed in Table 3.4. From both tables we observe a high occurrence of ZIDs with central tendency (3.5), being reported as the best-fitting types of distributions in the majority of the diagnostic sensitivity groupings, except for the 15EPG_McM_SCFT data; in this case the most common best-fitted distributions were those associated with the Negative Binomial distribution, with central tendency μ.

As a means of demonstrating the goodness of fit for the selection of distributions that could be fitted, the Fenbendazole Day 0 and Day 14 treatment group FEC data from an example farm, are displayed in Figures 3.1 and 3.2. Examples of fitted distributions based on 1000 simulated random samples are also included in these figures, where these samples were simulated from distributions using parameters estimated from the $gamlssML(\cdot)$ function. Their associated AIC values are displayed - where the lowest valued AIC displayed in these figures indicates the distribution of best fit for that particular set of data. From these figures, we can observe that zero inflated distributions represent the data obtained using 30 or 15 epg sensitivities very well, particularly for Day 14 FEC data, and how well distributions associated with the Negative Binomial fit the data in comparison to one another for the 15EPG_McM_SCFT data.

Best Fitting Distributions	30EPG_McM1 Data (%)	30EPG_McM2 Data (%)	15EPG_McM Data (%)	15EPG_McM_SCFT Data (%)
DEL	0.00	0.00	0.00	7.89
GEOM	2.63	5.26	10.53	27.63
NBII	5.26	3.95	5.26	27.63
PIG	5.26	6.58	14.47	21.05
SICHEL	0.00	0.00	0.00	6.58
ZINBI	26.32	14.47	25.00	6.58
ZIPIG	60.53	69.74	44.74	2.63

Table 3.3: Percentages of the best-fitting distributions for Day 0 data sets, categorised by the four diagnostic sensitivity groups (76 data sets in each grouping)

Best Fitting Distributions	30EPG_McM1 Data (%)	30EPG_McM2 Data (%)	15EPG_McM Data (%)	15EPG_McM_SCFT Data (%)
DEL	0.00	0.00	0.00	10.53
GEOM	0.00	0.00	0.00	15.79
INCONCLUSIVE	3.95	2.63	2.63	0.00
NBII	1.32	2.63	2.63	30.26
PIG	0.00	0.00	0.00	22.37
PO	0.00	0.00	0.00	1.32
SICHEL	0.00	0.00	0.00	1.32
ZINBI	7.89	6.58	10.53	5.26
ZIPI	18.42	26.32	19.74	5.26
ZIPIG	68.42	61.84	64.47	7.89

Table 3.4: Percentages of the best-fitting distributions for Day 14 data sets, categorised by the four diagnostic sensitivity groups (76 data sets in each grouping)

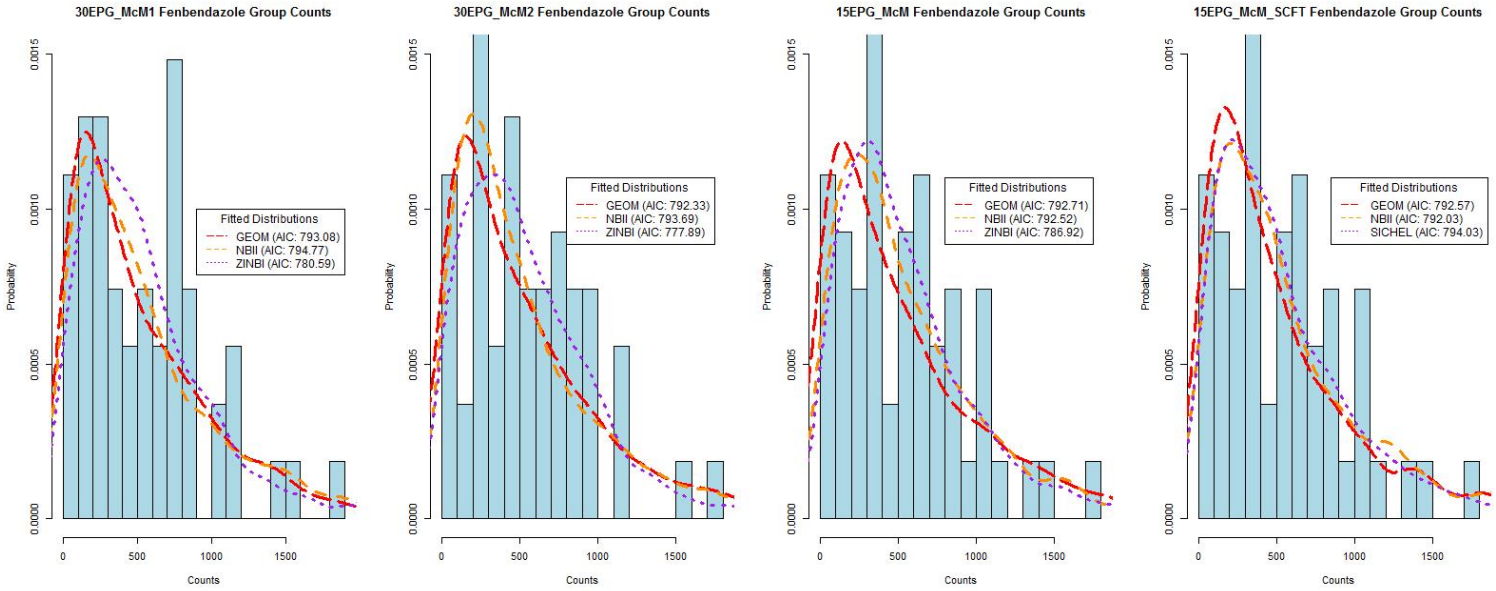

Figure 3.1: Farm E32 Fenbendazole Day 0 FEC data with example fitted distributions and their associated AIC values.

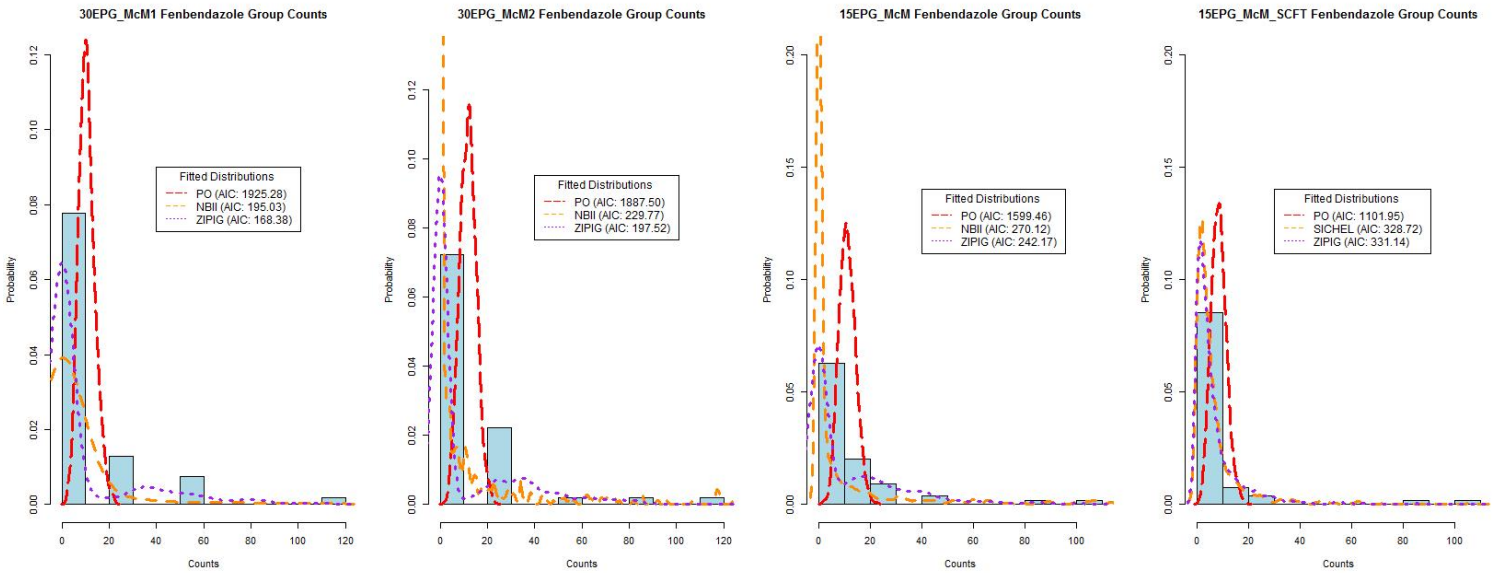

Figure 3.2: Farm E32 Fenbendazole Day 14 FEC data with example fitted distributions and their associated AIC values.

3.4 Comparing FECRT calculations: location parameters from best fitting distributions vs. arithmetic means

Given the results in Sections 3.2.4 and 3.3.5, it was of interest to compare the FECRT calculations of the percentage estimates (2.2), (2.3) and (2.4) (with $i = 14$) and their associated 95% upper confidence limits (UCLs) and lower confidence limits (LCLs) using estimated arithmetic means $\bar{x}$ (the estimator for the location parameter μ) and the central tendency estimates from the appropriate best-fitted distributions across all four diagnostic sensitivity grouping of cattle FEC data. From the available data for the project described in Section 1.6.1, twenty $100(1-\frac{T_{14}}{C_{14}})\%$ and $100(1-\frac{C_0T_{14}}{T_0C_{14}})\%$ percentage estimates and associated 95% confidence limits and sixty-one $100(1-\frac{T_{14}}{T_0})\%$ percentage estimates and 95% confidence limits (essentially all of the available sixty-one treatment groups) were able to be evaluated. Comparisons of each percentage estimate and associated 95% UCL and LCL, for each of the diagnostic sensitivity groupings, were made visually via scatterplots, where estimates based on utilising the location parameters of the best fitting distributions and based on utilising arithmetic means are viewed on the x and y axes, respectively. Straight lines were superimposed on these plots to represent the scenario of any percentage estimate, 95% UCL and/or 95% LCL being equal, irrespective of using arithmetic means and the location parameters of the best fitting distributions.

Since results highlighted in Table 3.1 indicated that the majority of cattle FEC data are of a non-normal nature, even upon transformation, then 95% Bootstrapped percentile intervals were estimated from 5000 iterations for each of the Day 0 and Day 14 negative control and positive treatment group data. Every combination of the 5000 estimates obtained for each set of data was considered, resulting in a sampling distribution of $2.5\text{x}10^7$ percentage estimates, from which the percentile intervals were derived. When considering distributions that were classed as inconclusive, the central tendency μ was used to represent the counts involved that all had the same value of zero, as this is the only appropriate central tendency that can be used to represent these data.

3.5 Results

Figures 3.3 and 3.5 give visual representations of the comparison of the twenty $100(1-\frac{T_{14}}{C_{14}})\%$ and $100(1-\frac{C_0T_{14}}{T_0C_{14}})\%$ percentage estimates and associated confidence limits, respectively, using the data available from the four diagnostic sensitivity groupings. Figure 3.4 displays similar information for the sixty-one $100(1-\frac{T_{14}}{T_0})\%$ percentage estimates and 95% confidence limits that were able to be evaluated. The straight lines featuring in these plots represent the scenario of either the percentage estimates (figures labelled (b), (e), (h) and (k)) UCLs (figures labelled (a), (d), (g) and (j)) or LCLs (figures labelled (c), (f), (i) and (l)) being equal when evaluated using both sets of estimates.

Overall, for each type of FECRT calculation method, the percentage estimate and the associated confidence limits estimated for FEC data obtained using 30 or 15 epg sensitivities using arithmetic means resulted in higher valued percentage estimates and interval estimates being obtained (i.e. values that lie above the straight lines), in comparison to those estimated using the central tendencies of the best-fitting distributions, i.e. zero inflated distributions. This was also the case for the comparisons that could be considered as outliers in Figures 3.3, 3.4 and 3.5, since these points lie above the straight lines in these figures. However, there was good agreement between the percentage estimates and confidence limits estimated using arithmetic means and central tendencies of the best-fitting distributions when considering the 15EPG_McM_SCFT data; since the majority of comparisons lie on the straight lines in the associated figures, which is due to the majority of 15EPG_McM_SCFT data being best fitted by distributions that are associated with the Negative Binomial distribution (Tables 3.3 and 3.4) and hence have the location parameter μ.

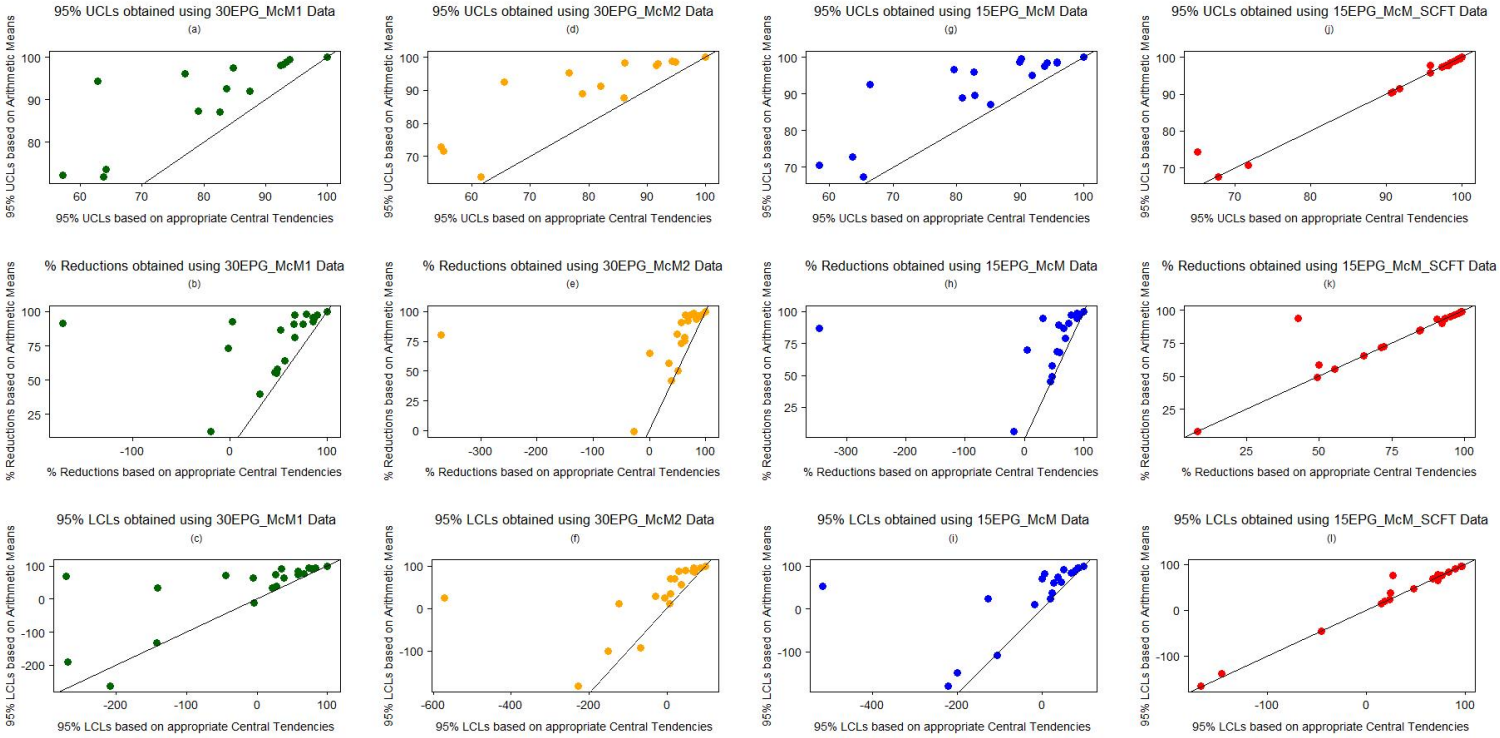

Figure 3.3: Comparison of $100(1-\frac{T_{14}}{C_{14}})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data.

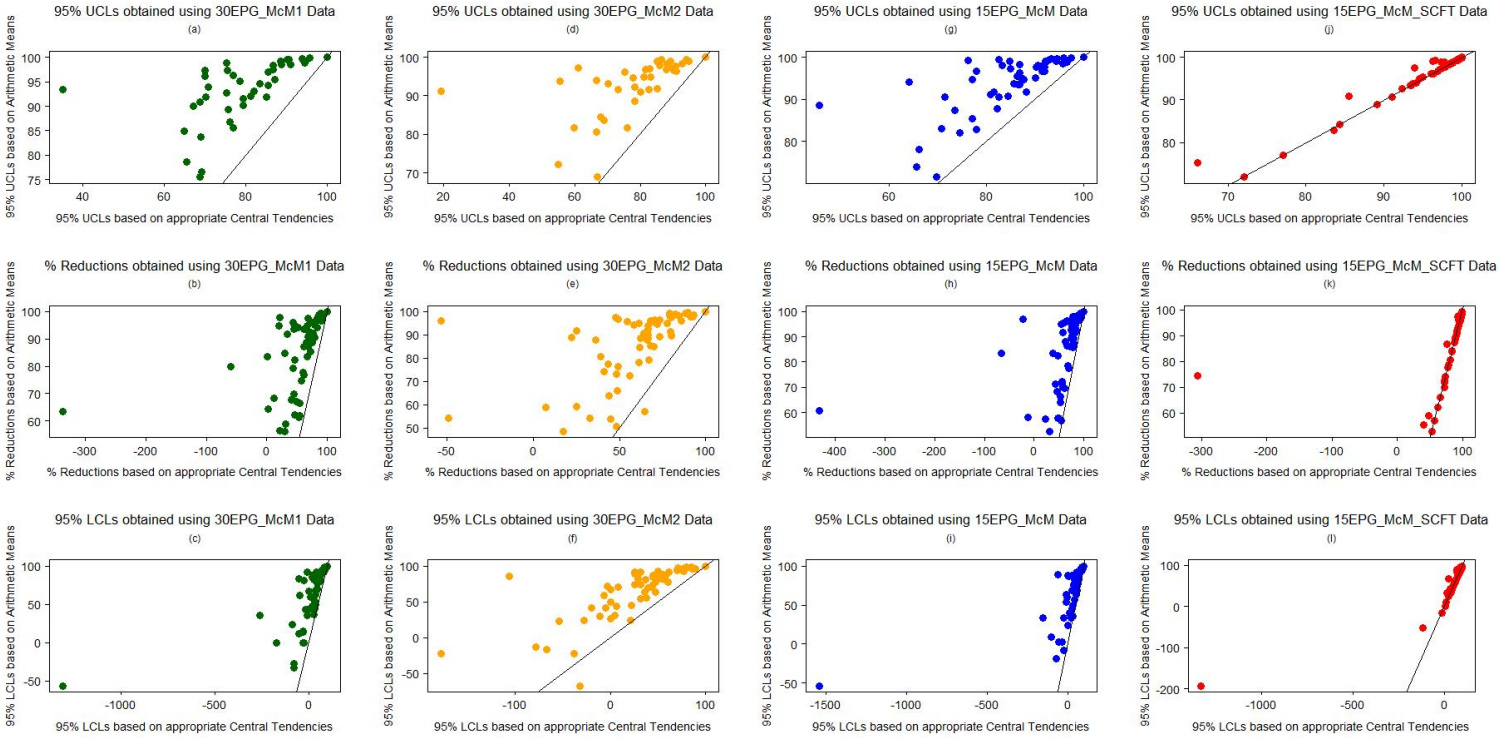

Figure 3.4: Comparison of $100(1-\frac{T_{14}}{T_0})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data.

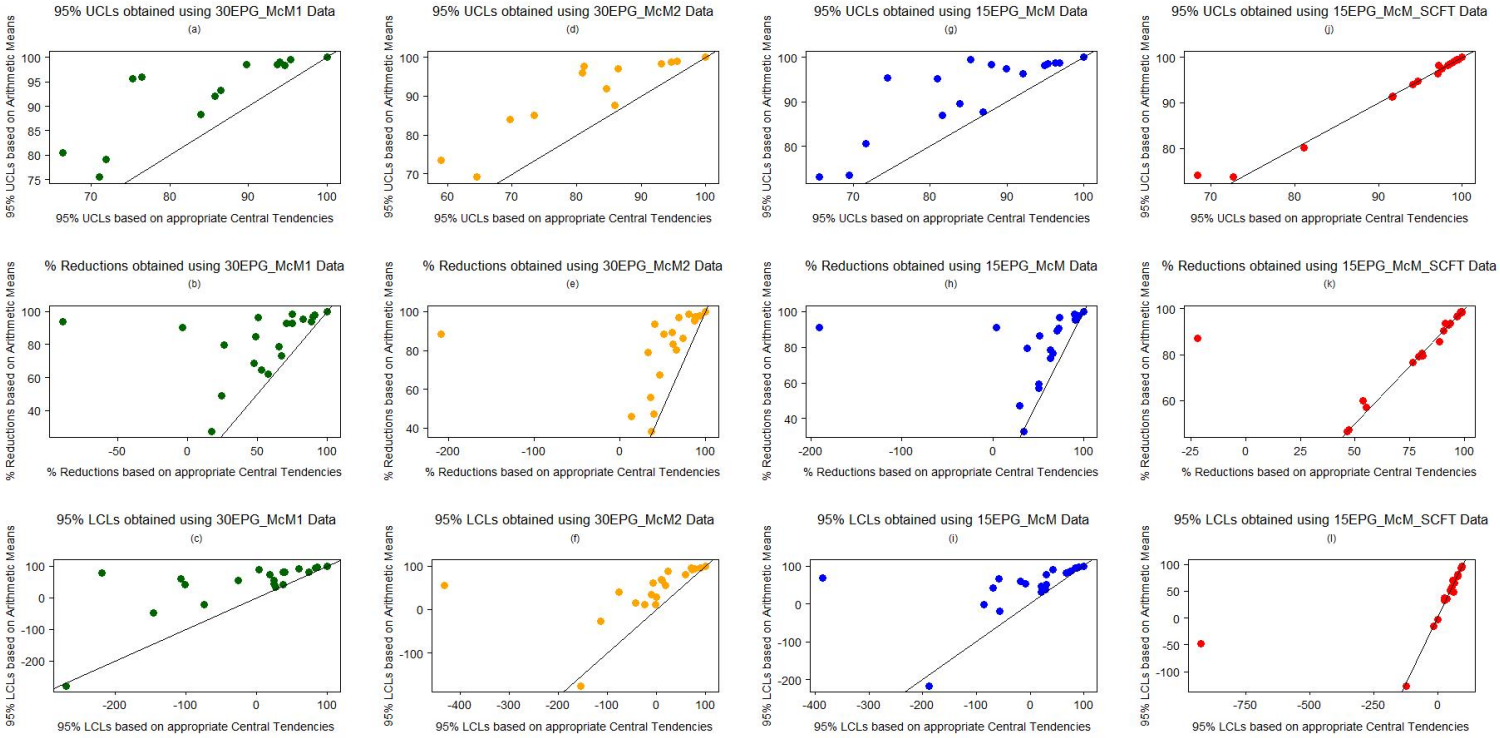

Figure 3.5: Comparison of $100(1 - \frac{C_0T_{14}}{T_0C_{14}})\%$ estimates and corresponding 95% UCLs and LCLs obtained using FEC data (central tendency estimates from best fitted distributions used vs. arithmetic group means used). Figures (a)–(c) based on 30EPG_McM1 data, (d)-(f) based on 30EPG_McM2 data, (g)-(i) based on 15EPG_McM data and (j)-(l) based on 15EPG_McM_SCFT data.

3.6 Discussion

The FECRT remains a widely used field test for anthelminthic efficacy, despite it never having been validated against slaughter studies. This Chapter provides insight into the potential statistical distributions that could be applied and represent cattle FEC data obtained by counting techniques of varying sensitivity, and the associated central tendencies that can be therefore utilised as part of evaluating percentage estimates in FECRT calculations. The normality of cattle FEC data, even upon transformation, has also been considered, which is an assumption required to ensure that the use of associated 95% confidence intervals, derived from asymptotic approximations, is valid.

The original 304 Day 0 and Day 14 FEC data sets and transformed versions of these were assessed for normality, since confidence intervals currently recommended to be used in a FECRT are derived assuming relevant data to be normal to obtain approximate estimates for the ln-transformed ratio of means of FEC data and its associated variance. For smaller sample sizes (<30), the Student's t-distribution is generally utilised to generate confidence intervals since it provides a more conservative estimate in comparison with the standard normal distribution. Furthermore, the transformations were used in an attempt to correct the usual skewness present in FEC data, to obtain data that would be considered as symmetric (Zar 1996; Torgerson *et al.* 2005; Vidyashankar *et al.* 2007).

The majority of the original and transformed data sets, both on Day 0 and Day 14, were found to be non-normal via the Shapiro-Wilk normality test. As a result, it would not be recommended to use confidence intervals that are based on large sample approximations that assume normality. Moreover, some of the confidence intervals derived by the Delta Method rely on correlations of natural logarithmic-transformed FEC data being evaluated; but this would not be possible if zero-valued FECs were obtained. In fact, given the nature of these data; confidence or credible intervals would be more suitably estimated using alternative methods such as Bootstrapping or a Bayesian approach, since generating these types of intervals from such frameworks is not necessarily dependent on assumptions of normality. Bootstrapping is a computer intensive and data driven technique that involves re-sampling observed data (Efron and Tibshirani 1993) and is generally

regarded by the veterinary and parasitological communities to potentially offer a simple, accessible and robust method to generate and infer confidence intervals for percentage estimates, even in the presence of small sample sizes (Cabaret and Berrag 2004; Lester and Matthews 2013; Lester *et al.* 2013). Bayesian Statistics lead us to work with a distribution for the parameters of interest (as opposed to fixing parameters to be estimated from data) for which credible intervals can be generated, and is the basis for subsequent inference within the Bayesian paradigm (Rice 2007). A Bayesian approach to analysing data offers benefits such as the usual normality assumption within statistical models being removed and unrealistic assumptions and simplifications being avoided when considering data, but Matthews (2014) highlights that a limitation to adopting Bayesian methods in analysing FEC data is the ability to use advanced statistical programmes, which the layperson may not be familiar with. The work presented here makes use of maximum likelihood estimation through the GAMLSS package, which is able to estimate distributional parameters, other than central tendencies such as variability, proportion of zeros etc., without the extra computational intensity that Bayesian inference can involve.

The results from this Chapter suggest that for cattle FECs obtained by sensitive counting techniques (such as the SCFT with a diagnostic sensitivity of 1 epg), distributions associated with and including the Negative Binomial distribution could be recommended to represent these data. Hence, percentage estimates and confidence limits can be estimated using arithmetic group means (i.e. the central tendency estimates associated with these distributions) in order to evaluate apparent anthelmintic efficacy. If cattle FEC data are obtained with less sensitive counting techniques (such as the McMaster technique with diagnostic sensitivities of 30 epg or 15 epg), ZIDs are recommended to represent these data, with central tendency μ_1 being used when calculating percentage estimates and confidence limits, due to excess zeros being produced by the counting techniques employed. As a result, this study demonstrates that the diagnostic sensitivities used in egg counting techniques influence the distribution of best representation for FEC data. For cattle, this is a consistent result with the study of El-Abdellati *et al.* (2010), who also reported that detection limits of counting techniques used in experimental studies are confounding factors of major importance when inves-

tigating anthelmintic resistance.

With respect to ZIDs, these distributions (more specifically zero inflated Negative Binomial distributions) have been used in worm egg count simulation studies involving sheep (Denwood *et al.* 2008) and horses (Denwood 2010; Denwood *et al.* 2010). The work in this Chapter compliments these earlier studies and advocates the use of ZIDs for representing real cattle FECs obtained using less sensitive counting techniques, but with alternative estimators being utilised. One could argue though, that in the present study there were animals involved whose FECs were less than the 100 epg threshold recommended by Coles *et al.* (2006) and so this introduces the need for ZIDs. However, the level of egg excretion is generally low and highly aggregated in cattle, i.e. the majority of cattle will be shedding low numbers of eggs in their faeces as well as few animals shedding a higher number of eggs (Demeler *et al.* 2009; El-Abdellati *et al.* 2010; Levecke *et al.* 2012) and is the reason as to why cattle with FECs less than 100 epg were included.

The choice of which central tendencies be used to represent FEC data, and therefore be used as part of a FECRT, has been long debated in veterinary parasitology research, despite the fact that the choice of central tendency depends on the distributions of best fit. For instance, the use of geometric means has been suggested previously (Presidente 1985; Mejia *et al.* 2003), but the use of arithmetic means has been more widely adopted (Dash *et al.* 1988; Dobson *et al.* 2009). In fact, Geurden *et al.* (2015) investigated anthelmintic efficacy in cattle in Europe, where egg counts were obtained using a diagnostic sensitivity of 12.5 epg (and 15 epg in one country) and arithmetic means were used to calculate percentage estimates. Our study, however, recommends the use of the central tendency μ_1 for FECRT calculations as opposed to the use of arithmetic means when dealing with ZIDs, on the basis that this is the maximum likelihood estimator for these types of distributions.

With this recommendation in mind, it naturally leads us to ask for what diagnostic sensitivities between 15 epg and 1 epg do we start accepting distributions, such as the Negative Binomial, being the better representation in comparison to ZIDs? As part of the current study we are unable to answer this question, but

this could be investigated as part of future studies.

Percentage estimates and their associated 95% UCLs and LCLs were evaluated using arithmetic group means and using the central tendency estimates of the best-fitted distributions for each of the different types of FEC data. With regards to FEC data obtained using 30 or 15 epg sensitivities, using central tendency estimates of the best-fitted distributions resulted in lower percentage estimates and confidence limits being obtained, in comparison to using arithmetic means in calculations. This is because the central tendency featuring in calculations was μ_1 in most cases, since the majority of these distributions were zero inflated distributions. As a result, for FEC data obtained by less sensitive counting techniques, an anthelmintic could be interpreted as over-performing (bearing in mind that a FECRT gives an indirect indication of efficacy) when arithmetic group mean estimates are used in the presence of zero inflated data.

Based on the hybrid sets of data with egg counts obtained with a 1 epg sensitivity, there was good agreement between the percentage estimates and confidence limits estimated using both arithmetic means and central tendencies of the best-fitting distributions. This was because the majority of these data were best represented by distributions whose central tendency was the arithmetic mean, i.e. those associated with the Negative Binomial distribution.

The work presented in this Chapter has been concerned with the analysis of cattle FEC data, and so it would be of interest to observe, using the above results and recommendations, what form of percentage estimates, and by extension what experimental designs, would be more appropriate to use for the FECRT in producing *robust* statistical analyses in a Bootstrapping framework. This will be considered in Chapter 4.

Chapter 4

Identifying a robust design of experiment via a simulation study involving Bootstrap methodology

4.1 Introduction

In chapter 2, it was highlighted that, as agreed by many researchers, there is a strong need to investigate which experimental designs are appropriate to determine and classify apparent anthelmintic efficacy in a robust manner, with respect to statistical calculations being carried out as part of the FECRT (Vidyashankar *et al.* 2007; Vidyashankar *et al.* 2012; Lyndal-Murphy *et al.* 2014). In Chapter 3, the probability theory associated with cattle FEC data has been underpinned and we shall utilise the findings from this Chapter to investigate what form of percentage estimates, and by extension what experimental designs, would be more appropriate to use for the FECRT in producing *robust* statistical analyses in a Bootstrapping framework.

In this section, we will primarily investigate the performance of confidence intervals using Bootstrap methodology, as a means of investigating the robustness of percentage estimates and, by extension, associated experimental designs. The performance of these intervals will be assessed via a *simulation study* using *RStu-*

dio software (version 0.98.994 along with *R* software version 3.1.1.). In the following section, we explore different aspects of confidence intervals and what is currently used as criteria for utilising appropriate percentage estimates in anthelmintic studies.

4.2 Criteria currently used to utilise appropriate percentage estimates in anthelmintic studies: confidence intervals

In Section 3.2.2, concepts of classical, null hypothesis significance testing (NHST) related to statistical parameters were explored. One disadvantage of NHST is that related inference procedures provide no information on the magnitude of an experimental effect (Gardiner and Gettinby 1998). Confidence intervals provide means of estimating this type of magnitude and this type of interval is defined in the general form as follows :

$$estimate \pm (critical\ value)(measure\ of\ variability). \tag{4.1}$$

The interval (4.1) tells us the range in which the true value of a parameter would occur with repeated sampling of the population with a pre-determined confidence level $100(1-\alpha)\%$, where α is defined as the significance level for the corresponding hypothesis test for which the confidence interval is providing the magnitude of effect for. It is worth noting that confidence intervals can take account of one-tailed and two-tailed hypotheses by way of the *critical value* in interval (4.1).

With respect to the confidence intervals that can be produced for a FECRT, Cabaret and Berrag (2004) tell us that one criterion for a good FECRT is one which has a small confidence interval, i.e. one with a small width. Factors which will influence the widths of a confidence interval are variability, which is influenced by the choice of study design (Borenstein *et al.* 2009), and sample size. The more variation present in a given sample; the wider the confidence intervals

that will be formed. On the other hand, if one were to increase the sample size then this would reduce the widths of confidence intervals because more of the study population of interest is being sampled from. For example, if we were to consider the confidence interval (2.22):

$$100\left(1-\frac{T_i}{C_i}\exp\left(\pm t_{(n_{treat}+n_{control}-2)}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T_i^2}+\frac{s^2_{c.eos}}{n_{control}C_i^2}}\right)\right)\%,$$

we can see that by increasing the values $s^2_{t.eos}$ and/or $s^2_{c.eos}$ or by decreasing the sample sizes n_{treat} and/or $n_{control}$ we would be able to produce a confidence interval with a wider width.

The widths of confidence intervals are an indication of whether or not there is sufficient data and whether or not the data are too variable to obtain a precise estimate to infer about the population parameter of interest, bearing in mind that *precision* can be thought of as the associated uncertainty of a given estimate. Hence, confidence intervals with large widths are an indication of insufficient data or large uncertainty.

One other thing that can impact on the width of a confidence interval is the *advertised coverage* of the interval. Gardiner and Gettinby (1998) tell us that in practice, the coverage offered by intervals is usually set to 95%, but can also be set to 99% and 90% in some circumstances. It is also worth noting that the higher the value of the advertised coverage as set by the researcher, the wider the resulting confidence interval will be.

Related to this, the *coverage probability* is the proportion of time that the obtained confidence interval actually contains the true specified parameter value upon repeated sampling of the population and this should be approximately equal to the coverage that is advertised. For example, if we were to sample from the same user population 1000 times, we would expect the average to fall within the interval 950, 990 and 900 times out of 1000 for the respective values of coverage mentioned above. In fact, it is this performance measure which will be primarily examined in our simulation study and will indicate the robustness of Bootsrapped percentile

intervals for the different percentage estimates - since the probabilities obtained will indicate whether or not the confidence intervals contain the true percentage estimates as advertised.

Before this however, the variability associated with the different experimental designs, from which relative percentage estimates are considered, must be investigated. Due to the nature of the positive treatment group paired study design (or parallel group design where a negative control is present) our simulation study will have to involve simulating paired data and we must ensure that the variability associated with these types of data/designs is reflected in our simulations. In doing this, we will be able to simulate representative FEC data, for which the coverage probability of confidence intervals for different percentage estimates can then be examined. As a result, the following section explores the variability associated with paired study designs.

4.3 Variability of the measured responses in anthelmintic studies

4.3.1 Variability of paired data

To investigate the variability associated with the paired study design, if we let, say, random variables X and Y be the FECs collected on Days 0 and 14, respectively, then it is the *differences* between the measured responses that are to be examined, i.e. $X - Y$ (Gardiner and Gettinby 1998). To make inference with paired study designs, the variance of the differences, i.e. $Var[X - Y]$, is required and usually estimated from the pairs of responses. Although traditionally the logarithm of the ratio $\frac{Y}{X}$ is considered, this equates to $ln(Y) - ln(X)$, and so a difference between measured responses can still considered.

Due to the fact that animals are considered as their own controls and FECs are obtained over time, then it is unlikely that the measured responses taken from Days 0 and 14 are independent, i.e. $Var[X - Y] = Var[X] + Var[Y]$. In fact, we would expect some form of correlation to be present between these

measurements, i.e. some form of dependence between FECs, due to the fact that measured responses are being obtained from the same individual animals throughout the study.

In fact, Rice (2007) tell us that the appropriate form for the variability present in the differences of observations, i.e. $Var[X - Y]$, is given by:

$$Var[X - Y] = Var[X] + Var[Y] - 2Cov[X, Y]. \tag{4.2}$$

In addition, since we know the co-variance can be defined in terms of the correlation between the random variables under observation, then equation (4.2) becomes:

$$Var[X - Y] = Var[X] + Var[Y] - 2\rho(X, Y)\sqrt{Var[X]Var[Y]}. \tag{4.3}$$

However, if we let $Var[X] \approx Var[Y]$, i.e. the variability is held constant between FECs obtained between Days 0 and Day 14 (this is generally assumed as part of paired study designs, though it is important to note that there would be no reason to expect this with FEC data), then equation (4.3) simplifies to

$$\begin{aligned} Var[X - Y] &\approx 2Var[Y] - 2\rho(X, Y)\sqrt{Var[Y]^2} \\ &\approx 2Var[Y] - 2Var[Y]\rho(X, Y) \\ &\approx 2Var[Y](1 - \rho(X, Y)) \\ &\approx 2s_Y^2(1 - r). \end{aligned} \tag{4.4}$$

where s_Y^2 is the sample variation of the FECs obtained on Day 14 and r is *Pearson's* correlation coefficient.

So, with reference to Equation (4.4), what can be said about $Var[X - Y]$ for given values of r?

If we consider $r > \frac{1}{2}$:

$$\begin{aligned} r &> \frac{1}{2} \\ \Rightarrow 1 - r &< \frac{1}{2} \\ \Rightarrow 2(1-r) &< 1 \\ \Rightarrow 2(1-r)s_Y^2 &< s_Y^2 \\ \Rightarrow Var[X-Y] &< s_Y^2. \end{aligned}$$

Thus, if the correlation is greater than $\frac{1}{2}$ (a positive correlation), then the variability in the differences between measured responses (i.e. the variability associated with a paired study design) will be approximately less than the variability of the single end of study measured responses, i.e. Day 14 FECs, in which case, considering animals as their own controls is favoured as an experimental design. For example if $r = 0.7$, then $Var[X-Y] = 0.6s_Y^2$, i.e. the variation of the differences between Day 0 and Day 14 FECs is approximately 40% less than the variability of the Day 14 FECs.

By a similar argument, if $0 < r < \frac{1}{2}$, then the variability in the differences of between measured responses will be approximately greater than the variability of the single end of study measured responses. For instance, if $r = 0.2$, then $Var[X-Y] = 1.6s_Y^2$, i.e. the variation of the differences between Day 0 and Day 14 FECs is approximately 60% more than the variability of the Day 14 FECs alone. The result being that including baseline, i.e. Day 0, measured responses would be adding to the variability of the study, if a paired study design were to be considered. In fact, if $r \leq 0$ we find that the variability in the differences of the measured responses would (approximately) at least be as twice as much as the variability associated with the Day 14 measured responses but will at most be four times as much as the variability associated with Day 14 responses since $r \geq -1$. Mathematically, we are saying that if $-1 \leq r \leq 0$, then $2s_Y^2 \leq Var[X-Y] \leq 4s_Y^2$. This can be shown as follows: for correlations r_1 and r_2 satisfying

$$
\begin{aligned}
-1 \leq r_1 \leq -\frac{1}{2} \leq r_2 \leq 0 \\
\Rightarrow 2 \geq 1 - r_1 \geq 1.5 \geq 1 - r_2 \geq 1 \\
\Rightarrow 4 \geq 2(1 - r_1) \geq 3 \geq 2(1 - r_2) \geq 2 \\
\Rightarrow 4s_Y^2 \geq 2s_Y^2(1 - r_1) \geq 3s_Y^2 \geq 2s_Y^2(1 - r_2) \geq 2s_Y^2.
\end{aligned}
$$

It is clear that investigating and evaluating correlations between two sets of paired measured responses, e.g. Day 0 and Day 14 FECs from individual cattle, allows us to infer about the variability present in paired study designs. As a result, if we are able to evaluate correlations and incorporate these into our simulation study then we incorporate information on the variability of the paired study/parallel group designs as required.

4.3.2 Investigating correlations of paired FEC data

The FEC data described in Section 1.6.1, for the mentioned diagnostic sensitivities, were used in order to investigate the correlations and, by extension, the variability in the differences of the measured responses betweeen Days 0 and 14 associated with the available 15 negative control and 61 positive treatment groups, on average. Treatment groups were categorised into those that had pre-treatment group means greater than 150 epg, those which had pre-treatment group means between 100 epg and 150 epg (inclusive) and those with pre-treatment group means less than 100 epg (Coles *et al.* 1992; Coles *et al.* 2006). As a result, 149, 62 and 93 out of the 304 available data sets were classed into these respective categories.

These data were further split into three categories based on treatment group sample sizes. For the negative control groups, FEC data were categorised into: $12 \leq n_{control} \leq 20$ (Small), $20 < n_{control} \leq 40$ (Medium) and $40 < n_{control} \leq 54$ (Large). For the positive treatment groups, FEC data were categorised into: $12 \leq n_{treat} \leq 20$ (Small), $20 < n_{treat} \leq 40$ (Medium) and $40 < n_{treat} \leq 61$ (Large).

Once these were evaluated, weighted averages of the average correlations for negative control and positive treatment groups were obtained for a particular diagnostic sensitivity grouping using the following formula:

$$\frac{\sum_i w_i \bar{r}_i}{\sum_i w_i},$$

where $\bar{r}_i$ is the *i-th* average correlation for a given treatment group and sample size range, and w_i is the the number of data sets for which the *i-th* average correlation was based on. By using a weighted average, we are able to attach more importance (i.e. weight) to some average correlations than others (Upton and Cook 2011) with respect to the number of studies that the average correlations were based on. The value of the average correlation was taken to be either the mean or median valued correlation coefficient, depending on whether or not the distribution of given correlation coefficients was of a normal nature or not (assessed using a *Shapiro-Wilk* normality test as described in Section 3.2.3). In addition, the number of groups for which $\bar{r}$ was based on, was evaluated.

4.3.3 Results

The average Pearson correlation coefficient, denoted as $\bar{r}$, was evaluated for each treatment group, given the appropriate group size range, type of diagnostic sensitivity grouping and type of pre-treatment group means, as shown in Table 4.1.

By observing Table 4.2, we see that the correlations between Day 0 and Day 14 FEC data for the positive treatment groups, irrespective of the diagnostic sensitivities and the range of the pre-treatment group mean ranges observed, have a correlation ranging between the values of 0 and 0.5, on average. As a result, FECs obtained between Days 0 and 14 have a weak, positive linear association with each other.

For the negative control groups we observe that, on average, when these groups have pre-treatment group means less than 100 epg or have a pre-treatment group mean lying between 100 epg and 150 epg (inclusive), the correlation between

Day 0 and Day 14 FECs lies between 0.5 and 1 and hence have a strong positive linear association. However, when the negative control groups have pre-treatment group means greater than 150 epg, we observe, on average, correlation values lying between 0 and 0.5.

As a result of gaining insight into the correlations of Day 0 and Day 14 cattle FEC data, we are able to reflect the variability of paired FEC data and can now utilise these conclusions to inform a simulation study. These correlations will be used in the simulation of FEC data to conclude on the performance of coverage probabilities offered by confidence intervals using different percentage estimates, offered by the paired study designs, as mentioned before.

Field Study Data	Sample Size Ranges	Treatment Group	$\bar{r}$ based on treatment data with pre-treatment means >150 epg	No. of treatment data with pre-treatment means >150 epg	$\bar{r}$ based on treatment data with 100 epg $\leq$ pre-treatment means $\leq$ 150 epg	No. of treatment data with 100 epg $\leq$ pre-treatment means $\leq$ 150 epg	$\bar{r}$ based on treatment data with pre-treatment means <100 epg	No. of treatment data from pre-treatment means <100 epg
	Small	Control	0.49	4	0.84	2	0.56	3
	Small	Positive Treatment	0.16	9	0.14	3	0.06	8
30EPG_McM1	Medium	Control	0.14	2	NA	0	NA	0
	Medium	Positive Treatment	0.26	16	0.07	7	0.06	8
	Large	Control	0.78	1	NA	0	0.63	1
	Large	Positive Treatment	0.34	7	0.32	1	0.11	4
	Small	Control	0.46	3	0.66	2	0.73	3
	Small	Positive Treatment	0.32	8	0.15	6	0.11	7
30EPG_McM2	Medium	Control	0.23	2	0.22	1	NA	0
	Medium	Positive Treatment	0.13	14	0.12	6	0.15	1
	Large	Control	0.69	1	NA	0	0.73	9
	Large	Positive Treatment	0.24	7	0.21	2	0.06	4
	Small	Control	0.48	3	0.73	2	0.78	3
	Small	Positive Treatment	0.26	8	0.14	5	0.19	7
15EPG_McM	Medium	Control	0.2	3	NA	0	NA	0
	Medium	Positive Treatment	0.26	16	0.19	7	0.07	8
	Large	Control	0.76	1	NA	0	0.74	1
	Large	Positive Treatment	0.32	6	0.36	2	0.11	4
	Small	Control	0.44	3	0.79	3	0.71	2
	Small	Positive Treatment	0.28	8	0.08	5	0.14	7
15EPG_McM_SCFT	Medium	Control	0.19	3	NA	0	NA	0
	Medium	Positive Treatment	0.24	16	0.16	7	0.06	8
	Large	Control	0.76	1	NA	0	0.73	1
	Large	Positive Treatment	0.35	7	0.38	1	0.34	4

Table 4.1: Average Correlations of treatment groups' field study data

Pre-treatment Group Mean Range	Type of Data	Type of Treatment Group	Weighted Average of Average Correlations
means>150 epg	30EPG_McM1	Control	0.43
	30EPG_McM1	Positive Treatment	0.25
	30EPG_McM2	Control	0.42
	30EPG_McM2	Positive Treatment	0.21
	15EPG_McM	Control	0.4
	15EPG_McM	Positive Treatment	0.27
	15EPG_McM_SCFT	Control	0.38
	15EPG_McM_SCFT	Positive Treatment	0.28
100 epg≤means≤150 epg	30EPG_McM1	Control	0.84
	30EPG_McM1	Positive Treatment	0.11
	30EPG_McM2	Control	0.51
	30EPG_McM2	Positive Treatment	0.15
	15EPG_McM	Control	0.73
	15EPG_McM	Positive Treatment	0.2
	15EPG_McM_SCFT	Control	0.79
	15EPG_McM_SCFT	Positive Treatment	0.15
means<100 epg	30EPG_McM1	Control	0.58
	30EPG_McM1	Positive Treatment	0.07
	30EPG_McM2	Control	0.73
	30EPG_McM2	Positive Treatment	0.1
	15EPG_McM	Control	0.77
	15EPG_McM	Positive Treatment	0.12
	15EPG_McM_SCFT	Control	0.72
	15EPG_McM_SCFT	Positive Treatment	0.15

Table 4.2: Weighted Averages of the Average Correlations

4.4 Simulation studies: a general overview

There are many percentage estimates that can be used to conclude on apparent anthelmintic efficacy/resistance, such as percentage estimates (2.2), (2.3), (2.4), (2.5), (2.6) and (2.7) mentioned in Section 2.3 (with $i = 14$). As a result, the only

sensible approach to determine what type of percentage estimate offers robust statistical analyses would be to conduct a *simulation study* based on real cattle FEC data (Burton *et al.* 2006). In this case, we would be interested in observing the performance of the confidence intervals associated with the various percentage estimates, via coverage probabilities. The performance of associated confidence intervals for the relative percentage estimates is of interest, since these intervals are used in the decision-making process of classifying apparent efficacy/resistance in cattle herds.

According to Burton *et al.* (2006), simulation studies are considered to be computer intensive procedures to assess the performance of a variety of statistical methods in relation to a known *truth/target value* and such evaluation cannot be achieved with studies of real data alone. As a result, an empirical estimation of the sampling distribution of the parameters of interest is provided that could not be achieved by means of a single study and the estimation of accuracy measures can be compared to the known truth/target value as well.

For the purposes of this study, simulations will involve repeated random sampling from probability distributions and are referred to as *Monte Carlo* simulations and these can produce a vast number of scenarios, i.e. *iterations* (Diaz-Emparanza 2002; Burton *et al.* 2006). These types of simulations are considered to be a relatively new area of Mathematics with a variety of applications and specialist packages being developed for their implementation. For example, Monte-Carlo simulation is easily implemented in *R*/*RStudio* using built in functions, which generate random values from probability distributions. As a result, these functions can be utilised within user-defined functions, which code the simulation.

The design of simulation studies however, in general, is considered to be a complex process and presents itself with numerous issues to be considered. For example, justifications have to be supplied for all decisions made prior to the simulation study being carried out (these would be detailed in a *study protocol*), a description of the simulation procedures, statistical methods to be evaluated, the storing of data and estimates, the number of simulations to be performed, the particular type of accuracy measures to be used, etc. (Burton *et al.* 2006).

Two issues in particular that are worth further discussion are those surrounding

the questions: *"How many simulations do I carry out?"* and *"What are the different types of performance measures available for me to use in my simulation study?"*

4.4.1 Number of simulations to carry out as part of a study

The number of simulations a researcher is required to carry out is an important consideration, because the quality of the results is directly influenced by the number of simulations carried out (Diaz-Emparanza 2002).

Burton *et al.* (2006) tell us that the number of simulations, B say, can be determined by using generic formula such as

$$B = \left(\frac{Z_{(1-\frac{\alpha}{2})}\sigma}{\delta} \right)^2 \tag{4.5}$$

where δ is the specified level of accuracy of the estimate of interest the researcher is willing to accept, i.e. the permissable difference from the "true"/target value, $Z_{(1-\frac{\alpha}{2})}$ is the $1-\frac{\alpha}{2}$ quantile of the standard normal distribution, α is the significance level set by the researcher and σ^2 is the variance of the parameter of interest.

Alternatively, B could be determined based on the power $1-\beta$ of a study, i.e. the probability of detecting a significant difference from the "true"/target value, where the following formula is considered:

$$B = \left(\frac{(Z_{(1-\frac{\alpha}{2})} + Z_{(1-\beta)})\sigma}{\delta} \right)^2 \tag{4.6}$$

and in fact, (4.6) is equivalent to (4.5) if the power is assumed to be 50%.

These formulae require a realistic estimate of the variance being obtained from data, but if the variance is unknown or cannot be estimated reliably then it may be possible to perform an identical pilot/test simulation study to obtain these

realistic estimates.

When one considers the relatively large FEC data set available in this project, the estimation of variability of the positive treatment and negative control FEC data for both Days 0 and 14 would require extensive time - a cost which is often taken into account when carrying out these types of studies (Burton *et al.* 2006). If individual values of B were to be examined for each data set, there is no guarantee that the maximum value of B that would be considered as the number of simulations to carry out would be manageable to achieve given the time constraints of this project - especially when we consider estimating the variability of these data given the different types of treatment groups, diagnostic sensitivities and farms involved. As a result, the number of simulations carried out was based on what could be achieved given the time constraints of the project, as opposed to evaluating the number of simulations based on the use of generic formulae such as those mentioned above.

4.4.2 Performance measures

According to Burton *et al.* (2006), it is necessary to consider the criteria for evaluating the performance of results from the different scenarios/statistical approaches being examined. The comparison of simulated results with the "true"/target values used to simulate the data provides a measure of performance and associated precision of the simulated process. Performance measures often considered are bias, accuracy and coverage. In our case, we are primarily interested in assessing the performance of confidence intervals and the estimation of percentage estimates in a Bootstrapping framework and so the coverage will be adopted as our primary performance measure. As an addition, if a percentage estimate provides adequate coverage consistently, we will then consider observing the *standardised bias* present for the relevant percentage estimates, which is described below.

Burton *et al.* (2006) tell us that the coverage is the proportion of times that the obtained confidence interval contains the true specified parameter value and that it should approximately be equal to the advertised coverage rate. Over-coverage occurs when the coverage rates are above the advertised coverage rates

and this can suggest that the results are too conservative since more simulations will not find a significant result when there is a true effect thus leading to a loss of statistical power with too many type II errors. On the other hand, under-coverage occurs when the coverage rates are lower than the advertised coverage rates and is considered unacceptable according to Burton *et al.* (2006) as it indicates over-confidence in the estimates.

One possible criterion for acceptability of the coverage is that it should not fall outside of approximately two standard errors (SEs) of the normal coverage probability, p say, i.e. $SE(p) = \sqrt{\frac{p(1-p)}{B}}$ (Burton *et al.* 2006). For example, if 95% confidence intervals are calculated using 1000 independent simulations then the coverage should fall between 93.6% and 96.4%. This criterion shall be utilised as part of our simulation study in order to identify those percentage estimates that offer suitable coverage and shall be referred to as *Burton's criterion* for the remainder of this Chapter.

Recall from Section 2.4.2, that the bias (Equation (2.15)) of an estimator, $\hat{\theta}$, for a parameter θ say, is defined as

$$Bias = E[\hat{\theta}] - \theta,$$

where $E[\hat{\theta}]$ is the average value of the distribution for the estimate $\hat{\theta}$ in a set of simulations. According to Burton *et al.* (2006), the amount of bias that is considered troublesome has varied from $\frac{1}{2}se[\hat{\theta}]$ to $2se[\hat{\theta}]$, where $se[\hat{\theta}]$ is the empirical standard error of the estimate $\hat{\theta}$ (estimated as the standard deviation) over all simulations.

An alternative concept is that of the standardised bias:

$$Standardised\ Bias = \frac{E[\hat{\theta}] - \theta}{se[\hat{\theta}]}.$$

According to Collins *et al.* (2001) and Burton *et al.* (2006), the standardised bias

can be more informative, as the consequence of the bias depends on the size of the uncertainty in the parameter estimate. For example, a value of -50% means that the estimate, on average, falls one half of a standard error below the parameter (Collins *et al.* 2001). It is also worth noting that a standardised bias within ±40% has been shown to not have noticeable adverse impact on the coverage probabilities of confidence intervals as well as the efficiency of estimators and error rates (Collins *et al.* 2001; Burton *et al.* 2006). As a result, this criterion shall be utilised when observing the standardised biases for percentage estimates, if required, and shall be referred to as *Collins' criterion* for the remainder of this Chapter.

4.5 Simulating paired data: copula

According to Nelsen (2006), the study of copula and their applications are a modern branch of Statistics. The term *copula* is used to describe functions that join multivariate distribution functions to their marginal distribution functions. In other words, copulas are multivariate distribution functions whose margins are uniform on the domain $(0, 1)$. In fact, Nelsen (2006) also tell us that copulas can offer us a way of studying scale-free measures of dependence and can be thought of as a starting point constructing families of distributions with a view to simulation. Upton and Cook (2011) also tell us that copulas encapsulate the interdependencies between variables and is also referred to as the *dependence* function.

In Mathematical terms, according to Upton and Cook (2011), let F be the multivariate distribution function for the random vector $(X_1, \ldots, X_d)$ and let the cumulative distribution function (CDF) of X_i (for $1 \leq i \leq d$) be F_i.

Define the random vector $(U_1, \ldots, U_d)$, such that $U_i = F_i(X_i)$ for each i, so that the marginal distribution of each U_i has a continuous uniform distribution on the domain (0,1). Assume that for each value u_i there is a unique value $x_i = F_i^{-1}(u_i)$ (i.e. the inverse of F_i exists, and to guarantee this we assume F_i to be continuous) and let the joint CDF of $(U_1, \ldots, U_d)$ be C. Then the copula, C is defined as

$$C(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d) = F[X_1 \leq F_1^{-1}(u_1), \ldots, X_d \leq F_d^{-1}(u_d)], \tag{4.7}$$

for all $u_1, \ldots, u_n$ in (0,1), since $U_i \leq u_i$ if and only if $X_i \leq F_i^{-1}(u_i)$.

Upton and Cook (2011) also tells us that an equivalent form of (4.7) is given by

$$C(F_1(x_1), \ldots F_d(x_d)) = F(x_1, \ldots x_d),$$

for all $x_1, \ldots, x_d$, where $u_i = F_i(x_i)$ for each i.

Copulas are important due to *Sklar's* Theorem, which states that, for a given F, there is a unique C such that (4.7) is satisfied (Nelsen 2006; Upton and Cook 2011).

As a result, the copula C contains all the relevant information on the dependence structure between the components of the random vector $(X_1, \ldots, X_d)$ and the marginal cumulative distribution functions F_i, for $1 \leq i \leq d$, have all the relevant information about the marginal distributions (Nelsen 2006; Upton and Cook 2011).

Equation (4.7) thus tells us that given the individual distribution functions are known but the joint CDF is unknown then a copula can be used to suggest a suitable form for the CDF (Upton and Cook 2011). However, a more powerful use of copulas lies in the reverse of the process described above. That is, given a procedure to generate the random vector $(U_1, \ldots, U_d)$ from the copula distribution, the required random vector $(X_1, \ldots, X_d)$ can be formed as

$$(X_1, \ldots, X_d) = (F_1^{-1}(U_1), \ldots, F_d^{-1}(U_d)).$$

4.5.1 Gaussian copula

Nelsen (2006) explains that there are many families of copulas used in practical applications, however, one of the most popular to be used, particularly in areas of finance (Malevergne and Sornette 2003; Cherubini *et al.* 2004; Renard and Lang 2007), is the Gaussian copula and this is the case where the joint CDF is a multivariate normal distribution function. This distribution function is the extension of the normal distribution of a single random variable to a random vector of elements that are each normally distributed (Upton and Cook 2011) and through its form, it naturally takes account of the correlations of the elements of the random vector.

Recall that, for a random variable Y to follow a normal distribution, then it is said to have probability density function

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}$$

and we say that $Y \sim \mathcal{N}(\mu, \sigma^2)$.

The multivariate normal distribution for a random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ is given by:

$$f(\mathbf{y}) = \frac{1}{2\pi^{\frac{n}{2}}} |\Sigma_d|^{-\frac{1}{2}} exp\left(\frac{(y-\mu_d)^T \Sigma_d^{-1} (y-\mu_d)}{2}\right)$$

where, $|\cdot|$ denotes the determinant, T denotes the transpose of a matrix, $\mu_d = (\mu_{Y_1}, \ldots, \mu_{Y_d})$ are the means for each of the respective components of $\mathbf{Y}$ and Σ_d is defined as the covariance matrix of the d components present in $\mathbf{Y}$ (Rice 2007; Upton and Cook 2011) and we say that $\mathbf{Y} \sim \mathcal{N}(\mu_d, \Sigma_d)$.

For example, the bivariate case of the multivariate normal distribution consists of $\mu_2 = (\mu_{Y_1}, \mu_{Y_2})$ and

$$\Sigma_2 = \begin{pmatrix} \sigma^2_{Y_1} & Cov[Y_1, Y_2] \\ Cov[Y_1, Y_2] & \sigma^2_{Y_2} \end{pmatrix}$$

for the normally distributed random variables $Y_1 \sim \mathcal{N}(\mu_{Y_1}, \sigma^2_{Y_1})$ and $Y_2 \sim \mathcal{N}(\mu_{Y_2}, \sigma^2_{Y_2})$.

In fact, according to Song (2000) and Arbenz (2013), if unit variances are allowed such that Σ_d becomes a correlation matrix, so for instance in the bivariate case we would obtain

$$\Sigma_2 = \begin{pmatrix} 1 & \rho(Y_1, Y_2) \\ \rho(Y_1, Y_2) & 1 \end{pmatrix},$$

then the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma_d)$ induces the Gaussian copula such that $\mathbf{0}$ is defined to be the mean zero vector and Σ_d is the relevant correlation matrix. Arbenz (2013) tells us that a copula is invariant under strictly increasing transformations of the margins, i.e. a positive scaling of the components of a multivariate normal changes the covariance structure but not the copula. Hence, the copula induced by the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma_d)$ (with Σ_d a covariance matrix, but not necessarily a correlation matrix) depends only on the correlation matrix induced by Σ_d. Malevergne and Sornette (2003) also tell us that the Gaussian copula is completely determined by the knowledge of the correlation matrix, which is indeed what we want to incorporate and account for when simulating FEC data as part of our simulation study.

As a result, the Gaussian copula can be defined as

$$C(u_1, \ldots, u_d) = \Phi_{\Sigma_d}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))$$

where Φ^{-1} is the inverse cumulative distribution function of a standard nor-

mal distribution and Φ_{Σ_d} is the joint CDF of a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \Sigma_d)$ where Σ_d is a correlation matrix with order $d \times d$. It is worth noting however, that due to a non-linear transformation, essentially, being applied to the uniformly distributed random variables, it is the case that the correlations of the simulated required samples will *approximately* be of the same value to the correlations used when defining the joint CDF.

4.6 Simulation study methodology

It was decided and agreed by the project team that three farms would be selected on which to base our simulation study, in order to compare the performance of the coverage probabilities of 95% Bootstrapped percentile confidence intervals. These were farms E32, D20 and D05. These farms were selected because they, respectively, had treatment groups that had pre-treatment group means greater than 150 epg, pre-treatment group means between 100 epg and 150 epg (inclusive) and pre-treatment group means less than 100 epg (with respect to a 15 epg sensitivity). As a result, we would be able to infer on the coverage of the confidence intervals for different percentage estimates for farms with different pre-treatment group mean thresholds based on published guidelines (Coles *et al.* 1992; Coles *et al.* 2006). These farms also had one negative control and at least one positive treatment group to utilise. For farm E32, the fenbendazole treatment group and for farms D20 and D05, the ivermectin injectable treatment groups' data were considered as part of this simulation study. These field study data were then used to evaluate the six percentage estimates: (2.2), (2.3), (2.4), (2.5), (2.6) and (2.7) (with $i = 14$), as described in Section 2.3.

For percentage estimates constructed using central tendencies, these were evaluated for each set of field study data, using the central tendencies/location parameters of the distributions of best fit, which were obtained from work presented in Chapter 3. As a result, we obtain the known "true"/target percentage estimates based on field study data for a particular diagnostic sensitivity grouping.

With respect to simulating FEC data, treatment group sample sizes of 15, 20, 30, 40 and 50 were considered and based on the correlation and variability results

described in Section 4.3.3, Day 0 and Day 14 data were not simulated independently for each type of treatment group; instead these data were simulated from a Gaussian copula to ensure specific correlation values, $r_{control}$ and r_{treat} say, were reflected in the simulated data. For farm E32, $r_{control} = 0.4$ and $r_{treat} = 0.3$, for farm D20, $r_{control} = 0.7$ and $r_{treat} = 0.2$ and for farm D05, $r_{control} = 0.7$ and $r_{treat} = 0.15$, given the information available from Table 4.2.

As mentioned in Section 4.5.1, due to a non-linear transformation being applied in the process of utilising the Gaussian copula, it is the case that the correlations of the simulated samples will be of approximately the same value to the correlations used when defining the joint CDF as part of the Gaussian copula. As a result, data were simulated till the first 1000 negative control and positive treatment group simulated sets of data, for farm E32, satisfied $0 < r_{control} < 0.5$ and $0 < r_{treat} < 0.5$, respectively. For farms D20 and D05, data were simulated till the first 1000 negative control and positive treatment group simulated sets of data, for each farm, satisfied $0.5 < r_{control} < 1$ and $0 < r_{treat} < 0.5$, respectively. These boundaries on the correlation values were chosen since they distinguish between different conclusions on the variability of paired data, as mentioned in Section 4.3.1. Also, the value of 1000 was chosen because this was the number of simulations that could be achieved given the time constraints of the project.

In the scenario where a correlation value of *NA* was obtained and stored in *RStudio*, this value is obtained due to at least one of either the negative control or positive treatment group Day 0 or Day 14 FEC data all being of the same value upon simulation. These were kept and considered as part of the simulation study, particularly since the scenario of positive treatment group Day 14 data all being of a value of zero represents the scenario of 100% (apparent) efficacy.

The field study data were also used to inform the parameter estimates of the required distributions for which the Gaussian copula aided in simulating. In particular, the field 15EPG_McM_SCFT data were used to inform the parameters of the Negative Binomial distribution (parameterised and denoted by NBII as described in Chapter 3), such that 1000 Day 0 and Day 14 negative control and positive treatment group data (satisfying the boundaries on the correlations mentioned above) were simulated (refer to Appendix A.1 to view the code utilised

for farm E32 data using treatment group sample sizes of 15, as an example). The 30EPG_McM1, 30EPG_McM2 and 15EPG_McM field study data each informed the parameters of a zero inflated Poisson inverse-Gaussian (ZIPIG) distribution (as described in Chapter 3), from which Day 0 and Day 14 negative control and positive treatment groups data were simulated 1000 times (again, satisfying the boundaries on the correlation values as stated above) as well (refer to Appendix A.2 to view the code utilised for farm E32 data using treatment group sample sizes of 15, as an example). These distributions were selected on the basis that they are recommended for representing data obtained from sensitive (i.e. 1 epg) and what could be considered less sensitive (e.g. 30 or 15 epg) counting techniques based on work and results described in Chapter 3.

Using the 1000 simulated Day 0 and Day 14, negative control and positive treatment group data sets; 95% Bootstrapped percentile confidence intervals (explained in Section 2.5.2) for the six percentage estimates considered, were evaluated based on 1000 iterations for each central tendency associated with the Day 0 and Day 14 negative control and positive treatment group data. When constructing percentage estimates, every possible combination of the obtained central tendencies for a given form of percentage estimate was evaluated; resulting in a sampling distribution of $1x10^6$ estimates to base the percentile intervals on. In the instance of considering the percentage estimates (2.6) and (2.7), percentile intervals were constructed based on $1x10^6$ re-samplings of the individual-based egg count percentage reductions/changes from the simulated data.

As a result, 1000 Bootstrapped percentile confidence intervals were evaluated for each of the six percentage estimates considered, for treatment groups associated with a given sample size and belonging to a particular diagnostic sensitivity grouping. The coverage probabilities associated with these percentage estimates was then obtained using the proportion of times that the target percentage estimates obtained using the field study FEC data were found to be in each of the considered 1000 Bootstrapped percentile confidence intervals based on the simulated data. Figure 4.1 highlights the methodology of this simulation study. The coverage probabilities, including those which satisfy *Burton's criterion* are highlighted in Section 4.6.1.

Only upon observing percentage estimates that consistently provided adequate coverage across treatment group sample sizes for a particular type of diagnostic sensitivity and classifications of pre-treatment means, were the associated standardised biases evaluated. With respect to the standardised bias, the value of the average of the estimates obtained from the simulations was taken to be the mean, since the standardised bias is dependent on the standard deviation of these estimates and this statistic gives inference on the measure of spread of the data about the mean (Upton and Cook 2011). These results are highlighted in Section 4.6.1.

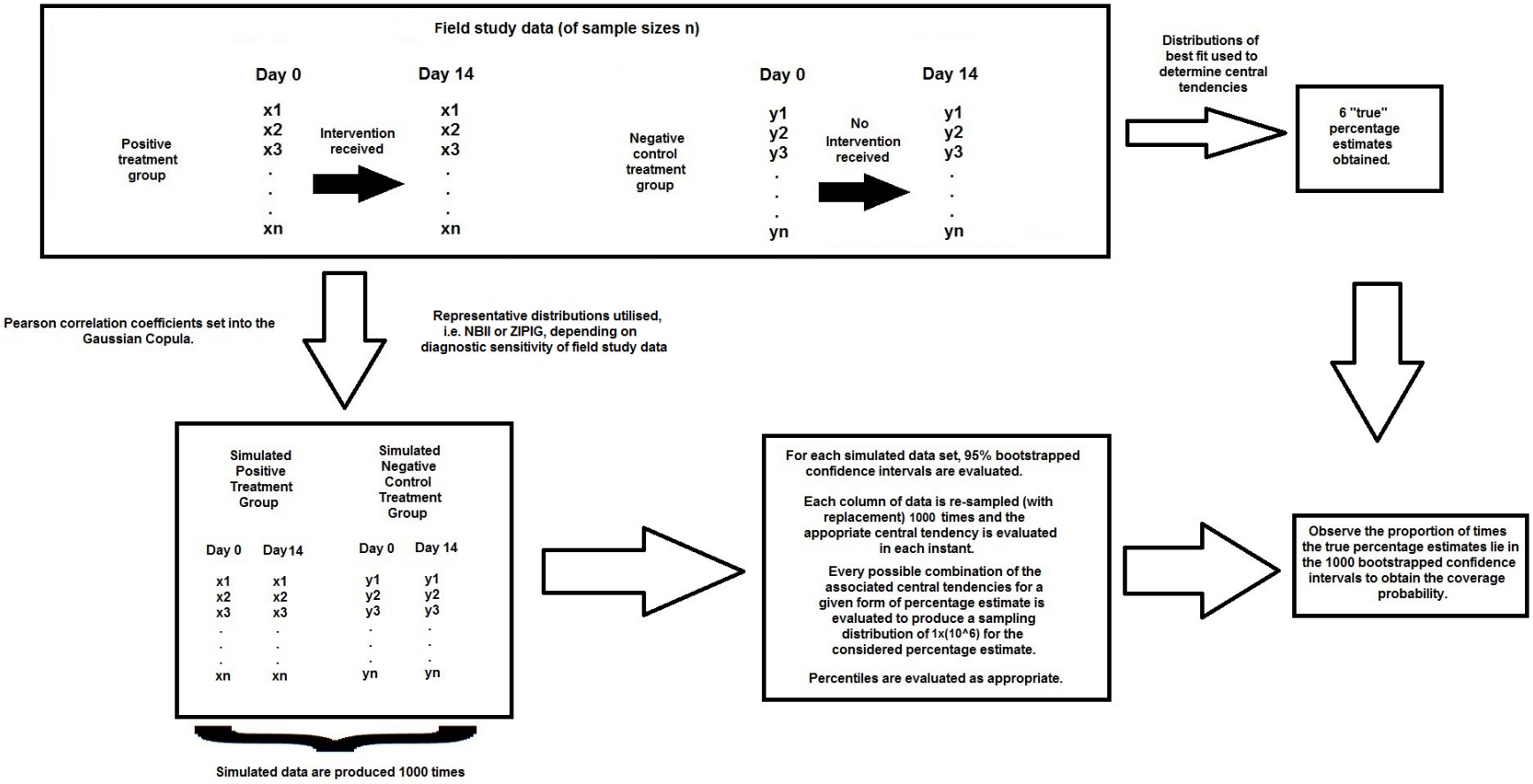

Figure 4.1: Bootstrap Simulation Study Methodology

4.6.0.1 Robustness of code used to obtain Bootstrapped percentile intervals

Due to the presence of paired data, the appropriate estimates that were produced were not obtained as part of a *boot*$(\cdot)$ object; but with a *sample*$(\cdot)$ function being utilised across the paired sets of FECs from the simulated data, in order to obtain Bootstrapped percentile intervals for each percentage estimate. When constructing these percentage estimates, every possible combination of the associated central tendencies for a given form of percentage estimate was evaluated; resulting in a sampling distribution of $1\text{x}10^6$ estimates to base the percentile intervals on. In the instance of considering the percentage estimates (2.6) and (2.7), *boot*$(\cdot)$ objects could be produced and percentile intervals were constructed based on $1\text{x}10^6$ re-samplings of the percentage estimates obtained from the simulated data.

When considering the original field study data, in the possible, yet unlikely scenario where one of the 15EPG_McM, 30EPG_McM1 or 30EPG_McM2 sets of treatment group data did not contain at least one zero count on a particular day: the value ν was set to $1\text{x}10^{-16}$ since the zero inflated distributions available in the GAMLSS package in *R/RStudio* require this parameter to satisfy $0 < \nu < 1$. As a result, $\mu_1 \approx \mu$. Similarly, in the scenario that one of the 15EPG_McM, 30EPG_McM1 or 30EPG_McM2 sets of treatment group data consisted of all zero counts on a particular day: the value ν was set to 0.999... to 16 decimal places, since the zero inflated distributions available in the GAMLSS package in *R/RStudio* require this parameter to satisfy $0 < \nu < 1$.

In the possible instance where any simulated set of data consisted of FECs all of a value of zero, then the arithmetic mean μ was used as the central tendency for this data and to construct the relevant percentage estimates.

When dealing with percentage estimate (2.6), when a pairing of two zero values for the Day 0 and Day 14 data occurred in the positive treatment group simulated data, the estimate, based on this pairing, was stored as a *NaN* in *R/RStudio* and removed before constructing the relevant Bootstrapped percentile interval, since the estimate is undefined in this scenario. Also, in the case of the denominator

of an individual percentage reduction/change being of a value of zero but the numerator being a positive value (this corresponds to a zero valued Day 0 data point with respect to the positive treatment group) this would be recognised as *-Inf* in the *R/RStudio* environment and was stored as an *NA* value and removed before constructing the relevant Bootstrapped percentile interval.

When dealing with the percentage estimate (2.7), when a pairing of two zero values for the Day 0 and Day 14 data occurred in the positive treatment group data; the percentage reduction estimate, based on this pairing, was stored as a *NaN* value in *R/RStudio* and removed since the estimate is undefined in this scenario.

4.6.1 Results

Tables 4.3-4.5 highlight the coverage probabilities associated with the Bootstrapped confidence intervals for different percentage estimates based on farm E32 D20 and D05 data, respectively. In particular, acceptable coverage probabilities, with respect to *Burton's criterion* are highlighted in the shaded grey cells of the tables. We observe that these acceptable coverage probabilities are associated with data simulated from the Negative Binomial distribution, for which the 15EPG_McM_SCFT field study data were utilised. In fact, the majority of these adequate coverage probabilities are associated with percentage estimates based on a paired study design involving a positive treatment group only, namely estimates $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$. Hence, Tables 4.6 and 4.7 highlight the associated standardised biases present for these percentage estimates respectively, for a given classification of pre-treatment group mean and treatment group sample sizes considered. The estimated standardised biases satisfying *Collins' criterion* are highlighted in the shaded grey cells of these Tables as well.

All of the coverage probabilities that are associated with data simulated from the ZIPIG distribution, for which field study data were obtained using a 15/30 epg sensitivity were utilised, are considerably less than the advertised coverage probability of 95% and do not satisfy *Burton's criterion*.

Field study data	Percentage Estimate (PEs)[1]	15	20	30	40	50 [2]
15EPG_McM_SCFT	PE1	91.5	92.1	92.1	93.3	93.4
	PE2	94.2	94.4	94.8	94.1	95.3
	PE3	65.4	64.6	70.4	67.2	71.1
	PE4	93.9	94.4	94.7	93.9	95.2
	PE5	71.6	71	74.9	73.2	76.9
	PE6	92.3	92.3	93.6	93.4	94.4
15EPG_McM	PE1	50.8	60.9	59.7	58.3	51.8
	PE2	54.8	64.1	61.1	61.5	60
	PE3	29.8	44.9	5.4	0.1	0
	PE4	54.8	64.1	61.1	61.5	60
	PE5	18.8	40.7	3.7	0.1	0.1
	PE6	72.8	76.2	68.4	73	72.2
30EPG_McM1	PE1	69.1	66.4	69.1	63.9	64
	PE2	72.2	68.5	74.5	68.1	74.9
	PE3	50.6	54.5	10.5	1.8	0.4
	PE4	72.2	68.5	74.5	68.1	74.9
	PE5	46.2	49.9	8.6	1.9	0.3
	PE6	75.8	72.4	77.6	76.2	78.5
30EPG_McM2	PE1	66.8	54.1	59.7	58.8	59.7
	PE2	72.4	54.4	72.1	67.1	69.6
	PE3	40.1	81.3	8.7	1.8	0.8
	PE4	72.4	54.4	72.1	67.1	69.6
	PE5	35.9	80.5	9.6	1.2	0.8
	PE6	78.9	54.7	69.2	75.2	76.7

Table 4.3: Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with pre-treatment group means greater than 150 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%)

[1]PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[2]Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [3]	15	20	30	40	50 [4]
15EPG_McM_SCFT	PE1	84.8	84.7	79.9	76.9	79.3
	PE2	79.5	75	63.8	55.4	64.5
	PE3	83.8	85.4	87.1	83.6	83.9
	PE4	79.5	75	63.8	55.4	64.5
	PE5	86	88.1	93.1	92.5	91.5
	PE6	87.4	89.3	85.8	82.6	85
15EPG_McM	PE1	78.1	84.4	86.1	88.4	87.4
	PE2	57.9	61.1	60.1	67.3	70.5
	PE3	83.1	45.2	25.1	66.5	25.2
	PE4	57.9	61.2	60.1	67.3	70.5
	PE5	81.3	30.3	16.4	31.8	4.9
	PE6	75.6	76.6	77.7	80.8	81.8
30EPG_McM1	PE1	75.6	86.8	83.8	85.7	87.2
	PE2	68	62.5	61	63.2	71
	PE3	76.1	44.9	19.9	21.6	5.2
	PE4	68	62.5	61	63.2	71
	PE5	66.4	33.7	15.7	13	2.8
	PE6	77.1	78.1	78.3	79.8	82.1
30EPG_McM2	PE1	82.9	49.1	90	88.9	89.5
	PE2	65.2	49.1	67.8	59.2	71.4
	PE3	81	72.9	37.7	22.2	27.9
	PE4	65.1	49.1	67.8	59.2	71.4
	PE5	89.6	68.4	45.5	79	39
	PE6	86.2	49.3	84.9	77.7	83.5

Table 4.4: Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%)

[3] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[4] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [5]	15	20	30	40	50 [6]
15EPG_McM_SCFT	PE1	92.5	92.6	93.8	93.8	92.7
	PE2	92.3	93.1	93.6	94.3	94.7
	PE3	72.6	61.3	36.8	21.3	12
	PE4	92.3	93.1	93.6	94.3	94.7
	PE5	73.8	62.3	45.9	31.1	21.7
	PE6	89.3	89.9	90.5	92.8	91.6
15EPG_McM	PE1	80.1	86.2	85.9	79	88.4
	PE2	86	83.1	85.1	86.8	86
	PE3	77.9	71.2	49.2	38.7	17.2
	PE4	86.1	83.1	85.1	86.8	86
	PE5	90.6	88.4	55.2	78	68.9
	PE6	82.8	83.2	84.1	80.5	84.2
30EPG_McM1	PE1	78.2	87.6	88.1	83.5	89.3
	PE2	85.1	84.7	85.3	87.7	84.8
	PE3	79.9	68.2	38.3	61.3	12.2
	PE4	85.1	84.7	85.3	87.7	84.8
	PE5	90.3	87.1	70.2	91.6	86.8
	PE6	78.3	84.7	83.9	82	84.7
30EPG_McM2	PE1	82.2	85.3	87.2	80.6	88.5
	PE2	83	84.5	82.7	86.3	83.7
	PE3	87	88.9	85.1	67.2	79.2
	PE4	83.6	84.6	82.7	86.3	83.7
	PE5	89.1	84.5	80.7	75.6	71.7
	PE6	83.9	84.8	84.8	78.8	84.7

Table 4.5: Coverage probabilities (%) for the associated 95% percentile intervals of percentage estimates from a farm with pre-treatment group means less than 100 epg (highlighted cells have coverage probabilities that lie between 93.6% and 96.4%)

[5] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[6] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Pre-treatment Group Mean Range (Farm)	$n_{treat} = 15$	$n_{treat} = 20$	$n_{treat} = 30$	$n_{treat} = 40$	$n_{treat} = 50$
mean>150 epg (Farm E32)	-8.62	-5.84	-8.47	2.6	-5.49
100 epg≤mean≤150 epg (Farm D20)	90.97	133.57	192.95	232.81	185.62
mean<100 epg (Farm D05)	-27.43	-26	-31.9	-34.63	-35.37

Table 4.6: Standardised biases (%) of the percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)\%$ for pre-treatment group mean classifications and treatment group sample sizes, based on 15EPG_McM_SCFT field study data (highlighted cells have standardised biases between ±40%)

Pre-treatment Group Mean Range (Farm)	$n_{treat} = 15$	$n_{treat} = 20$	$n_{treat} = 30$	$n_{treat} = 40$	$n_{treat} = 50$
mean>150 epg (Farm E32)	-7.51	-4.72	-7.27	3.88	-4.1
100 epg≤mean≤150 epg Farm (D20)	97.79	135.31	191.53	230.07	184.97
mean<100 epg (Farm D05)	-19.52	-18.31	-26.42	-29.88	-30.64

Table 4.7: Standardised biases (%) of the percentage estimate $100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)\%$ for pre-treatment group mean classifications and treatment group sample sizes, based on 15EPG_McM_SCFT field study data (highlighted cells have standardised biases between ±40%)

4.7 Sensitivity analysis

A *sensitivity analysis* can be thought of as a systematic approach of changing the parameters involved in a model or simulation study and viewing the results under the given changes, or more commonly known as *scenarios*. As a result, one is able to view the impact that these changes in certain parameters may have on a model or simulation study's final conclusions.

In this case, we would be interested to see how the performance of the coverage probabilities of the associated 95% Bootstrapped confidence intervals for each of the percentage estimates react to changes in the correlation values that are used

to simulate the appropriate data. In general, the simulation study carried out so far involves observing the coverage of confidence intervals based on the first 1000 simulated Day 0 and 14 negative control and positive treatment group FEC data satisfying $0 < r_{control}, r_{treat} < 1$, for a given farm.

As part of this sensitivity analysis we set $r_{control}$ and r_{treat} both to the value of zero for any given farm's data, which represents the scenario that there would be no linear relationship between the Day 0 and Day 14 FECs of the positive treatment and negative control groups. However, as mentioned in Section 4.5.1, due to a non-linear transformation essentially being applied in the process of utilising the Gaussian copula, it is the case that the correlations of the simulated required samples will *approximately* be of the same value to the correlations used when defining the joint CDF as part of the Gaussian copula. As a result, data were simulated till the first 1000 negative control and positive treatment group simulated sets of data satisfied $-0.1 < r_{control} < 0.1$ and $-0.1 < r_{treat} < 0.1$ for any given farm.

4.7.1 Results

Tables 4.8-4.10 display similar information as in Tables 4.3-4.5, but with correlations, for simulating the negative control and positive treatment groups' FEC data, being set to the value of zero. We observe fewer acceptable coverage probabilities, with respect to *Burton's criterion*, associated with data simulated from a Negative Binomial distribution in comparison to what was obtained as part of the original simulation study. As well as this, all of the coverage probabilities that are associated with data simulated from the ZIPIG distribution are still considerably less than the advertised coverage probability of 95% and do not satisfy *Burton's criterion*.

Field study data	Percentage Estimates (PEs) [7]	15	20	30	40	50 [8]
15EPG_McM_SCFT	PE1	91.7	93	93	93.3	92.6
	PE2	91.1	92.4	92.1	92.8	93.9
	PE3	81.6	85.2	87	90.3	92.9
	PE4	90.9	92.3	92	92.6	93.8
	PE5	84.2	88.1	89.2	92.3	93.6
	PE6	92.2	92.7	93.5	93.7	93.8
15EPG_McM	PE1	65.5	77.4	73.1	65.2	63.9
	PE2	64.2	79.2	71.4	68.4	70.3
	PE3	36.5	53.6	8.5	0.4	0.2
	PE4	64.2	79	71.4	68.4	70.3
	PE5	20.8	45.1	4.6	0.2	0.2
	PE6	83.4	91.6	81.1	81.8	81.8
30EPG_McM1	PE1	79.2	83.5	83.4	71	73.1
	PE2	79.8	84.2	82.5	74.3	80.8
	PE3	53.3	57.6	15.9	2.7	0.2
	PE4	79.8	84.2	82.5	74.3	80.8
	PE5	44.8	48.8	9.8	1.2	0.2
	PE6	83.6	86.8	87.6	83.8	85.6
30EPG_McM2	PE1	81.6	73.1	72.7	66.6	69.9
	PE2	83.2	74.8	80.9	74.3	76
	PE3	48.4	81.1	11.5	2.3	0.7
	PE4	83.2	74.8	80.9	74.3	76
	PE5	36	78.1	9.5	1.4	0.5
	PE6	90.7	75.2	82.3	83.8	84.5

Table 4.8: Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with pre-treatment group means greater than 150 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%)

[7] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[8] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [9]	15	20	30	40	50 [10]
15EPG_McM_SCFT	PE1	82.7	85.5	80.2	77.3	72.4
	PE2	77.8	74.6	63.6	55.8	48.3
	PE3	86.1	87.3	82.7	74.1	64.9
	PE4	77.8	74.6	63.5	55.8	48.3
	PE5	90	92.2	92.2	90.5	90.4
	PE6	91.4	92.7	90.5	89.8	88.5
15EPG_McM	PE1	80.5	89	87.4	88.2	85.2
	PE2	69.1	74.6	72.4	77	76.8
	PE3	82.2	53.1	38.2	78.5	34.5
	PE4	69.1	74.6	72.4	76.9	76.7
	PE5	83.3	32.7	18.5	33	4
	PE6	90.5	92.1	90.5	90.1	89.1
30EPG_McM1	PE1	81.7	88.9	82.2	84.9	84.7
	PE2	80.1	77.3	73.1	74.4	76.9
	PE3	84.5	58.4	30.5	33.2	9
	PE4	79.9	77.3	73.1	74.4	76.9
	PE5	68.1	34.2	15.7	14.4	2.7
	PE6	88.5	92.5	89.8	89.6	89.5
30EPG_McM2	PE1	86.7	72.4	87.2	90	89
	PE2	75.9	76.7	75.2	71.1	77.9
	PE3	77.8	79.8	47.9	40.1	43.7
	PE4	76.1	76.7	75.2	71.1	77.9
	PE5	90.7	67.5	45.8	82.5	39.5
	PE6	90.8	75.8	92.1	87.3	90

Table 4.9: Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%)

[9] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[10] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [11]	15	20	30	40	50 [12]
15EPG_McM_SCFT	PE1	92.6	92	93.3	94	94.1
	PE2	91.7	92	92.8	94.5	94.9
	PE3	65.3	51.6	27.7	12.7	5.6
	PE4	91.7	92	92.8	94.5	94.9
	PE5	67.5	58.7	40.6	27.6	16.7
	PE6	92.8	92.2	92.7	93.4	93.9
15EPG_McM	PE1	88.9	86.9	87.9	76.2	86.7
	PE2	90.6	90	90	90.7	88.7
	PE3	83.3	81.3	61.8	44.7	21.6
	PE4	90.6	90	90	90.7	88.7
	PE5	90.9	90.2	56.3	79.3	70.3
	PE6	93	93.3	92.8	88.1	91.1
30EPG_McM1	PE1	85.3	92.8	89.5	81.5	88.3
	PE2	91.1	92.1	91	90	88.4
	PE3	86.5	82.1	50.1	72.8	19.1
	PE4	91.2	92.1	91	90	88.4
	PE5	92.9	90.7	72.5	92.7	87.8
	PE6	89.8	93.6	93.6	89.6	91.4
30EPG_McM2	PE1	87.1	87.3	88.2	76.4	88.1
	PE2	86.5	88.7	85.9	90.4	86.7
	PE3	86	91	88.3	73.6	86.9
	PE4	87.1	88.7	85.9	90.4	86.7
	PE5	90.1	87.9	77.7	77.2	73.8
	PE6	91.6	92.2	90.2	87.9	90.8

Table 4.10: Coverage probabilities (%) for 95% percentile intervals of % estimates from a farm with pre-treatment group means less than 100 epg as part of sensitivity analysis (highlighted cells have coverage probabilities between 93.6% and 96.4%)

[11] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[12] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

4.8 Discussion

Results based on farm with pre-treatment group means greater than 150 epg:

For data simulated by the Negative Binomial distribution (based on 15EPG_McM _SCFT field study data) which is often utilised and widely accepted to represent and simulate FEC data from livestock due to FEC data being over-dispersed/ aggregated (Wilson *et al.* 1996; Shaw and Dobson 1995; Morgan *et al.* 2005; Levecke *et al.* 2012), the SPC:

$$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)\%,$$

had acceptable coverage probabilities associated with it for all treatment group sample sizes considered. The SPC can offer many theoretical advantages according to Berry and Ayers (2006), such as being bounded (i.e. between $\pm 100\%$).

Adequate coverage was also achieved when the following percentage estimate was considered:

$$100\left(1 - \frac{T_{14}}{T_0}\right)\%.$$

These acceptable coverage probabilities were associated with all treatment group sample sizes as well. Published guidelines on establishing apparent anthelmintic efficacy/resistance, such as the thresholds provided by Coles *et al.* (1992) and Coles *et al.* (2006), have been adopted in the instance of this percentage estimate being used. On the other hand, it could be argued that the SPC provides more theoretical benefits. In addition, the standardised biases associated with these two percentage estimates satisfied *Collins' criterion* for all treatment group sample sizes considered.

It can be observed however, that acceptable coverage probabilities were achieved in the instance of the following percentage reduction being adopted for the more

larger treatment group sample sizes of 30 and 50:

$$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%.$$

Lyndal-Murphy *et al.* (2014) advocates the use of this percentage estimate, but this recommendation is based on simulated data for animal group sample sizes of 15 only and confidence intervals derived using the Delta method were used, which did involve correlations of FEC data; but not correlations of ln-transformed data as required. However, our study is able to assess the coverage probabilities of confidence intervals which do not depend on large sample normal approximations, use appropriate estimators associated with the distributions utilised in simulating FEC data, consider treatment groups of various sample sizes and various percentage estimates and also consider the correlations between Day 0 and Day 14 FEC data appropriately.

One could argue however, that the practicability of this percentage estimate may not be favourable, particularly since percentage estimates such as those involving pre- and post-treatment counts from a positive treatment group only (i.e. those mentioned above), simultaneously provide adequate coverage probabilities - for all of the treatment group sample sizes considered. These percentage estimates are calculated in the instance of a paired study design being adopted involving a positive treatment group only, and not a parallel group study design involving a negative control group. This type of design has proved popular in the veterinary parasitology community and has been widely adopted due to the convenience of not having to include a negative control group (Kochapakdee *et al.* 1995; Lyndal-Murphy *et al.* 2010; Levecke *et al.* 2012; Vidyashankar *et al.* 2012; Lester *et al.* 2013; Stratford *et al.* 2014; Geurden *et al.* 2015; George *et al.* 2017).

The sensitivity analysis revealed fewer acceptable coverage probabilities being obtained in comparison to those obtained as part of positive correlations being set in the original simulation study. In fact, the majority of acceptable coverage probabilities were associated with numerous percentage estimates where treatment group sample sizes of 50 animals were considered. This indicates that the

coverage probabilities are influenced by the correlations, and by extension, the variability present between negative control and positive treatment group Days 0 and 14 FEC data when counting techniques of a 1 epg sensitivity are utilised and pre-treatment group means are greater then 150 epg.

It is worth noting that no acceptable coverage probabilities were obtained for simulated data based on the 15EPG_McM, 30EPG_McM1 and 30EPG_McM2 field study data as part of the original simulation study, or as part of the sensitivity analysis.

Results based on farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg:

When considering data simulated by the Negative Binomial distribution, i.e. those based on 15EPG_McM_SCFT field study data, acceptable coverage probabilities were not obtained for any of the treatment group sample sizes considered. No acceptable coverage probabilities were obtained for simulated data based on the 15EPG_McM, 30EPG_McM1 and 30EPG_McM2 field study data either and this is a consistent result for these data with results obtained from farm data with pre-treatment group means greater than 150 epg.

No acceptable coverage probabilities were obtained as part of the sensitivity analysis, irrespective of the diagnostic sensitivity for which field study data were based on.

One could argue at this point that a limitation of this simulation study is that one representative farm is only being utilised for each classifications of pre-treatment group means to draw conclusions from. However, given the consistency of the results between the different types of farm data considered and the time constraints of the project that could be invested into carrying out this simulation study, which is indeed a constraint that has to be taken into account when carrying out these types of studies according to Burton *et al.* (2006), three farms from which to draw conclusions from was achievable. As well as this, the value in utilising examples of field study data, no matter how few, supersedes this limitation, since we are able to gain informative conclusions based on data, which is indeed representative of what is collected in the field.

Results based on farm with pre-treatment group means less than 100 epg:

For simulated data based on the 15EPG_McM_SCFT field study data, acceptable coverage probabilities were obtained for the SPC and the percentage estimate involving pre- and post treatment counts for a positive treatment group mentioned above - but these probabilities are associated with treatment group sample sizes of 30, 40 and 50 in this instance. The standardised biases associated with these percentage estimates for these treatment group sample sizes also satisfy *Collins' criterion*. On the other hand, one would have to consider the feasibility and practicability of this result and ask if it is worth running a FECRT with a large positive treatment group sample size of 30, 40 or 50 cattle (in order to ensure adequate coverage), with pre-treatment group means are less than 100 epg?

Moreover, the following percentage estimate:

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%$$

also had acceptable coverage probabilities associated with it, for treatment group sample sizes of 30 and 40. Published guidelines, such as Coles *et al.* (1992) and Coles *et al.* (2006), recommend the use of this percentage estimate for treatment group sizes of 15 animals using arithmetic means, and these are indeed the estimators being considered here. But again, one would have to consider the feasibility and practicability of using this percentage estimate with a parallel group design involving a negative control group and a positive treatment group with sample sizes of 30 and 40 cattle, where pre-treatment group means are less than 100 epg? It is also worth mentioning that adequate coverage probabilities were obtained with the percentage estimates, mentioned above, that involve a paired study design with a positive treatment group only - for the same treatment group sample sizes of 30 and 40. As a result, these percentage estimates may appear to be more favourable to utilise, as opposed to the percentage estimate based on a parallel group design being adopted, with respect to practicability and feasibility given that pre-treatment group means are less than 100 epg.

Over and above this, no acceptable coverage probabilities were obtained for any

of the percentage estimates based on data simulated from a zero inflated distribution, and this is a consistent result with the other two types of farm data that have been considered.

Interestingly, when we consider the sensitivity analysis: the SPC and the percentage estimate involving pre- and post treatment counts for a positive treatment group only had accepted coverage probabilities associated with them for sample sizes of 40 and 50. This could suggest that for these treatment group sample sizes in particular, these percentage estimates could be perceived as robust in light of the different values of correlations, and by extension variability, set as part of the simulation study and sensitivity analysis, in the instance where we observe pre-treatment group means less than 100 epg.

No acceptable coverage probabilities were obtained, again, for simulated data based on the 15EPG_McM, 30EPG_McM1 and 30EPG_McM2 field study data as part of the sensitivity analysis.

Overall conclusions:

This simulation study, along with the sensitivity analysis, revealed that the acceptable coverage probability offered by the associated 95% Bootstrapped percentile intervals can be influenced by treatment group sample sizes, the diagnostic sensitivity of the counting techniques used in obtaining the original FEC data, the correlations (and by extension the variability) between FEC data obtained at baseline and end of study and the classifications of the pre-treatment group means obtained.

By considering this study as a whole, very few scenarios consisted of 95% Bootstrapped percentile intervals with adequate coverage probabilities being achieved. Although, it could be recommended that the SPC or the percentage estimate using pre- and post-treatment group FECs from a positive treatment group only, both using arithmetic means, be considered as percentage estimates to be used in the conduct of a FECRT when FEC data are obtained using sensitive counting techniques (1 epg sensitivity) with Bootstrapping methodologies being utilised. However, depending on the classification of the pre-treatment group means will influence the treatment group sample sizes for these percentage estimates to be

utilised. Based on this study, when dealing with treatment groups with pre-treatment group means greater than 150 epg, a minimum positive treatment group sample size of 15 animals could be utilised to achieve adequate coverage, whereas we are unable to recommend any treatment group sample sizes, based on this study, to obtain percentage estimates with adequate coverage where pre-treatment group means are between 100 epg and 150 epg (inclusive) and a minimum positive treatment group sample size of 30 animals would be required if pre-treatment group means are less than 100 epg to obtain adequate coverage.

If either of these percentage estimates were utilised as part of the statistical calculations for a FECRT, given the appropriate treatment group sample sizes in the various scenarios discussed above, then a paired study design involving a positive treatment group would only be required, as opposed to using a parallel group design involving a negative control group.

No acceptable coverage probabilities were obtained in any of the scenarios considered based on field study data being obtained with less sensitive counting techniques (i.e. 30 or 15 epg). Therefore, it would not be recommended to conduct a FECRT with counting techniques having these sensitivities in the case where confidence intervals are obtained in the Bootstrapping framework - the only alternative would be to obtain credible intervals in a Bayesian framework as a means of interval estimation.

Since very few scenarios consisted of 95% Bootstrapped percentile intervals with adequate coverage probabilities being achieved, Bayesian methods will now be considered in Chapter 5, where we will aim to identify appropriate and *robust* percentage estimates, and by extension experimental designs, to investigate apparent anthelmintic efficacy/resistance.

Chapter 5

Identifying a robust design of experiment via a simulation study involving Bayesian methodology

5.1 Introduction

Research has been invested into using Bayesian methods and obtaining credible intervals when investigating apparent anthelmintic efficacy and resistance, but mainly with respect to equine FEC data (Denwood 2010; Denwood *et al.* 2010); though are being employed to analyse cattle and sheep FEC data in more recent studies (Denwood *et al.* 2008; Dobson *et al.* 2012; Busin *et al.* 2013; Geurden *et al.* 2015; Wang *et al.* 2017).

As mentioned in Section 2.4.3, a Bayesian approach to analysing data offers the benefit of not necessarily assuming data to be normal for estimating intervals for population parameters of interest - an assumption which has been found not to be reasonable with regards to cattle FEC data, even upon transformation (refer to Section 3.2 for further information).

Even though one could also consider constructing intervals for parameters of interest via a Bootstrapping framework, in which we do not necessarily need data to be of a normal nature to generate intervals; results from Chapter 4 indicated

that some types of cattle FEC data, e.g. those obtained using less sensitive counting techniques (15 or 30 epg) do not provide confidence intervals, for percentage estimates considered in this project, with adequate coverage.

As a result, it was of interest to conduct another simulation study similar to what was conducted and described in Chapter 4, i.e. investigate the robustness of percentage estimates and, by extension, associated experimental designs, but with Bayesian methodologies being utilised to construct posterior distributions for point estimates and credible intervals for considered percentage estimates.

As mentioned in Section 2.4.3, the mean/median of the posterior distribution is usually quoted as the point estimate for a parameter of interest, i.e. percentage estimate, and so how *accurate* this estimate is with respect to a known "true"/target value of a percentage estimate will be examined as part of this simulation study in order to investigate the robustness of percentage estimates and, by extension, associated experimental designs.

It is also worth noting that, as explained in Section 2.5.4, a credible interval tells us that given the observed data, there would be a $100(1-\alpha)\%$ probability that the true value of the parameter, θ say, lies within the credible region, i.e. we can be $100(1-\alpha)\%$ sure that the parameter lies in the credible region obtained. As a result, we do not consider coverage as a performance measure as part of this simulation study since, by definition, a credible interval is assumed to contain the true parameter value with a $100(1-\alpha)\%$ probability.

This simulation study will be carried out using *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.). The methodology adopted for this simulation study is as follows.

5.2 Simulation study methodology

The farms, percentage estimates, treatment group sample sizes, diagnostic sensitivities, probability distributions, methodologies and simulated FEC data, which were presented in Section 4.6, were considered and utilised for this simulation study.

For each of the 1000 simulated Day 0 and 14 negative control and positive treatment group Day 0 and 14 data, likelihoods were considered, namely the Negative Binomial distribution (i.e. the NBII distribution, as described in Chapter 3) for simulated data based on 15EPG_McM_SCFT data or a zero inflated Poisson inverse-Gaussian distribution (i.e. the ZIPIG, as described in Chapter 3) for simulated data based on field study data collected with 15/30 epg sensitivities. Posterior distributions were derived for the parameters μ and σ associated with the NBII likelihood or μ_1, σ and ν associated with the ZIPIG likelihood. The posterior distributions for the parameters μ and μ_1 were subsequently used (depending on the type of simulated data being considered) for constructing the following percentage estimates: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, $100\left(1-\frac{T_{14}}{T_0}\right)\%$, $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$ and $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$. The prior specifications used to obtain the posterior distributions for each of the parameters described above are described in Section 5.2.1.

For each of the 1000 simulated Day 0 and 14 positive treatment group Day 0 and 14 data, individual-based egg count percentage reduction/change data of the following forms were also considered: $100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%$ and $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (where $T_{0,j}$ and $T_{14,j}$ are pre- and post-treatment FECs from host j, respectively). For these data, the truncated normal distribution (further information on this distribution can be found in Appendix B.3) was considered as a likelihood for both forms of data and posterior distributions were derived for the associated parameters: μ and σ. The posterior distribution for the parameter μ was then used to construct the following percentage estimates: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$ and $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$. It is worth noting that the prior specifications used to obtain the posterior distributions for each of the parameters involved with the truncated normal likelihood are independent of the priors specified above for the parameters involved in the NBII and ZIPIG likelihoods, and further information on these are available in Section 5.2.1.

To obtain posterior distributions, Metropolis-Hastings (MH) algorithms (further information on MH algorithms is available in Appendix B.1) were developed us-

ing *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.) and implemented based on a total of 10,000 iterations. An initial discard of 5,000 iterations was considered (i.e. the burnin period). Normal proposal distributions centred around current parameter values of the chains were used, with a proposal variance of one, i.e. $\mathcal{N}(\theta, 1)$. Each parameter considered was updated in turn. To ensure each of the parameters' chains were tuned and explored their respective posterior distributions efficiently (i.e. achieved acceptance rates between 20-40% after the initial burnin period) the MH algorithms were programmed to be adaptive, such that at every $200th$ iteration in the burnin period: if the acceptance rate of the chain was above 40% then the variance of the proposal distribution was doubled and if the acceptance rate of the chain was below 20% then the variance of the proposal distribution was halved. In developing these algorithms, convergence diagnostics, such as trace plots, acceptance rates, autocorrelations and the Gelman-Rubin statistic were applied to assess and verify that there was no indication of *lack of convergence* to the stationary, posterior distributions post-burnin and how efficient the algorithm was when traversing the posterior distributions, the results of which are available in Appendix B.2. For constructing the percentage estimates mentioned above, the remaining 5,000 iterations associated with the location parameters μ or μ_1 (depending on the likelihoods being considered), were thinned-out so every $5th$ estimate contributed towards the sample from the posterior distribution. The resulting sample for each central tendency/location parameter considered, consisted of 1000 sampled estimates from the posterior distributions. The relevant percentage estimates were then constructed; resulting in a sampling distribution of 1000 estimates for each relevant percentage estimate.

From the sampling distributions for each of the six percentage estimates considered, the medians of these distributions were obtained and the "true"/target values obtained from field study data were used to derive the *Root Mean Square Error* (RMSE) for each percentage estimate (further information on this statistic can be found in Appendix B.3). It is also worth noting that 95% credible intervals, i.e. Highest Posterior Density intervals, were obtained for each percentage estimate. The percentage estimates associated with the lowest RMSEs evaluated out of the six obtained values, for a given simulation, are of interest here, since these indicate the percentage estimates that appear to be the most accurate and

these results are presented in Section 5.3. Figure 5.1 highlights the methodology of this simulation study.

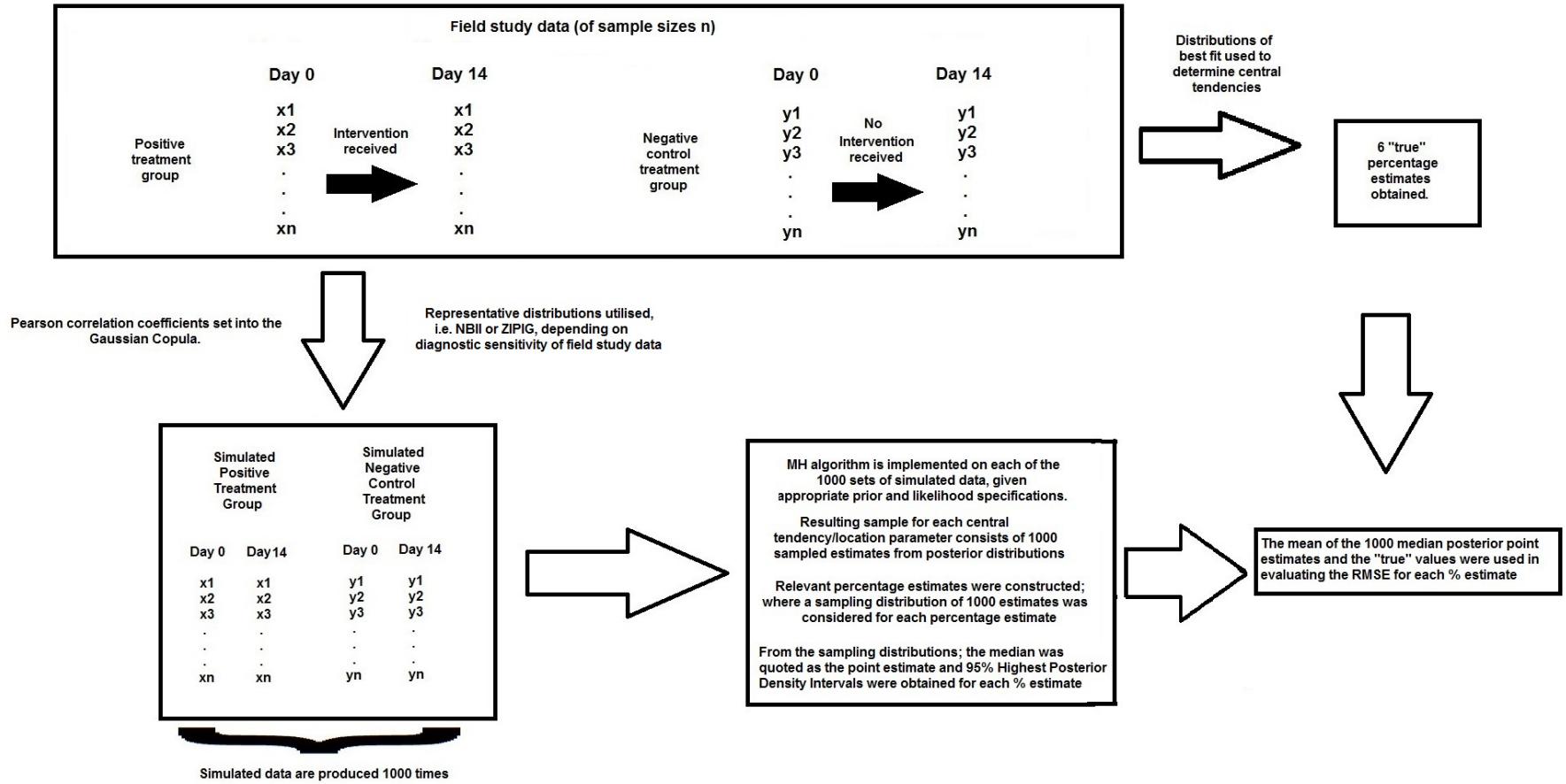

Figure 5.1: Bayesian Simulation Study Methodology

5.2.1 Prior and likelihood specifications

For positive treatment and negative control Day 0 and Day 14 data based on being obtained by a sensitive counting technique (i.e. 1 epg), and based on the results presented in Chapter 3, the likelihood was defined to be the product of probabilities from a Negative Binomial Type II distribution, denoted as $NBII(\mu, \sigma)$. Therefore, with respect to Equation 2.18:

$$f(x|\mu, \sigma) = \prod_{i=1}^{n} NBII(x_i, \mu, \sigma).$$

When considering the approaches of averaging over individual-based egg count percentage reductions/changes of a positive treatment group, given by percentage estimates (2.6) and (2.7), the likelihood was defined to be the product of probabilities from a truncated normal distribution. In this case, the distribution is truncated at the value of 100 since, under the assumption of treatment working effectively we would expect individual-based egg count percentage reductions/changes to have a left-tailed skewed distribution to feature with the majority of values located at the value of 100. This distribution will be denoted as $\mathcal{N}_{trunc}(\mu, \sigma)$ and further information on this distribution can be found in Appendix B.3. As a result, with respect to Equation 2.18:

$$f(x|\mu, \sigma) = \prod_{i=1}^{n} \mathcal{N}_{trunc}(x_i, \mu, \sigma).$$

For μ and σ, prior distributions, namely continuous uniform distributions, were adopted to obtain posterior distributions and were defined over different domains in light of the different types of data considered as part of this study:

Prior Distributions for parameters for data based on 1 epg sensitivities:

Day 0 Negative Control Group Data :

$$\mu \sim Unif(1\text{x}10^{-16}, 1000) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

Day 14 Negative Control Group Data :

$$\mu \sim Unif(1\text{x}10^{-16}, 1000) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

Day 0 Positive Treatment Group Data :

$$\mu \sim Unif(1\text{x}10^{-16}, 1000) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

Day 14 Positive Treatment Group Data :

$$\mu \sim Unif(1\text{x}10^{-16}, 300) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

$$100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)\% \text{ Data} :$$

$$\mu \sim Unif(-100, 100) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

$$100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)\% \text{ Data} :$$

$$\mu \sim Unif(-100, 100) \text{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

When considering the continuous uniform distribution as a prior for each of the parameters, this distribution could be considered as unusual, on the grounds that there are more commonly used distributions utilised for certain parameters. For example, the inverse-Gamma distribution can often be adopted when considering the dispersion parameter σ. However, from discussions with veterinarian and parasitological collaborators on the project, the domains stated within each of

the continuous uniform priors, for each given parameter, were agreed upon as they are of a conservative nature and the parameters would not exceed the values stated in the domains of the continuous uniform priors.

The domains chosen to define the continuous uniform distributions for the parameter μ, for Day 0 and 14 negative control and Day 0 positive treatment group data are the same. This is because FECs at baseline would be assumed to come from the same population (as they are collected prior to any treatments given to the positive treatment groups on Day 0). As well as this, due to the purpose of a control group we would not expect FECs to change drastically on Day 14 in the absence of an intervention, however the value of 1000 for these central tendencies is of a high, conservative nature. Under the assumption that positive treatments have worked effectively, the continuous uniform distribution for the parameter μ based on the Day 14 positive treatment group data being considered is defined over a smaller, but could still be considered as a conservative range between 0 to 300 for these data. However, the lower limits of the domains for the continuous uniform distributions defining the central tendencies are of a value $1\text{x}10^{-16}$, since we require $\mu > 0$ with respect to the Negative Binomial distribution.

Based on utilising individual-based egg count percentage reductions/changes of the form $100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)\%$, the continuous uniform distribution is defined over the range of -100 to 100 for the parameter μ. We know that at most, any individual animal can obtain a 100% reduction in their FECs between Day 0 and Day 14. However, there is the possibility that some animals may not present a reduction in their FECs between this time; in fact their shedding of egg counts increases between Day 0 and Day 14 and so in this scenario, the individual's percentage "reduction" would present itself as a negative value, and this scenario must be accounted for when defining prior distributions for these data. We also know that if the individual-based egg count percentage reductions/changes are of the form represented by the SPC; the average of these individual changes must be bounded by -100 to 100 since the SPC itself as a percentage estimate is bounded between -100 and 100 Berry and Ayers (2006). Therefore, the continuous uniform distribution $Unif(-100, 100)$ was defined as a prior distribution for the parameter μ for both types of data considered.

It is also worth mentioning that the $Unif(1\text{x}10^{-16}, 500)$ distribution has been defined for all σ parameters considered. Again, the value of 500 is of a high, conservative nature and we require $\sigma > 0$ for all likelihoods considered thus far. In terms of variability, the domain being considered is approximately $\sigma^2 \sim Unif(0, 250000)$ in these cases which is reasonable, since we know that FECs are over-dispersed/aggregated (Wilson *et al.* 1996; Shaw and Dobson 1995; Morgan *et al.* 2005; Levecke *et al.* 2012). It is worth noting however, that these domains can be modified as part of a prior-sensitivity analysis, and this will be presented later.

For positive treatment and negative control Day 0 and Day 14 data based on being obtained by a less sensitive counting technique (i.e. 15 or 30 epg), and based on the results presented in Chapter 3, the likelihood was defined to be the product of probabilities from a zero inflated Poisson inverse-Gaussian distribution, denoted as $ZIPIG(\mu_1, \sigma, \nu)$. Therefore, we obtain the following likelihood:

$$f(x|\mu_1, \sigma, \nu) = \prod_{i=1}^{n} ZIPIG(x_i, \mu_1, \sigma, \nu).$$

For μ_1, σ and ν prior distributions, were adopted again to obtain posterior distributions and were defined over different domains in light of the different types of data considered:

Prior Distributions for parameters for data based on 15/30 epg sensitivities:

$$\textit{Day 0 Negative Control Group Data} : \mu_1 \sim Unif(1\text{x}10^{-16}, 1500),$$
$$\sigma \sim Unif(1\text{x}10^{-16}, 500) \textit{ and } \nu \sim Unif(1\text{x}10^{-16}, 0.99...(16.d.p.)).$$

$$\textit{Day 14 Negative Control Group Data} : \mu_1 \sim Unif(1\text{x}10^{-16}, 1500),$$
$$\sigma \sim Unif(1\text{x}10^{-16}, 500) \textit{ and } \nu \sim Unif(1\text{x}10^{-16}, 0.99...(16.d.p.)).$$

$$\textit{Day 0 Positive Treatment Group Data} : \mu_1 \sim Unif(1\text{x}10^{-16}, 1500),$$
$$\sigma \sim Unif(1\text{x}10^{-16}, 500) \textit{ and } \nu \sim Unif(1\text{x}10^{-16}, 0.99...(16.d.p.)).$$

$$\textit{Day 14 Positive Treatment Group Data} : \mu_1 \sim Unif(1\text{x}10^{-16}, 500),$$
$$\sigma \sim Unif(1\text{x}10^{-16}, 500) \textit{ and } \nu \sim Unif(1\text{x}10^{-16}, 0.99...(16.d.p.)).$$

$$100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)\% \textit{ Data} : \textit{Defined as before.}$$
$$100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)\% \textit{ Data} : \textit{Defined as before.}$$

It is worth noting that the domains considered for the continuous uniform distributions for the parameter μ_1 have increased, in light of Day 0 and 14 positive treatment and negative control data being considered. This is due to the work presented in Chapter 3, since we know the location parameter/central tendency of a zero inflated distribution is greater than the value of the parameter μ.

It is also worth noting that the parameter ν is defined using the $Unif(1\text{x}10^{-16}$, $0.99...(16.d.p.))$ distribution, which is defined over the domain from $1\text{x}10^{-16}$ to 0.9999999999999999, since we require for the ZIPIG distribution parameter ν to be defined as $0 < \nu < 1$.

5.3 Results

Tables 5.1-5.3 highlight the RMSE values associated with the different percentage estimates considered as part of this study, based on farms E32, D20 and D05 data, respectively. In particular, RMSE values closest to the value of zero, for a given scenario, are highlighted in the shaded grey cells in these Tables.

For data simulated from the 15EPG_McM_SCFT field study data based on farms with pre-treatment group means either between 100-150 epg (inclusive) or greater than 150 epg, we observe that the majority of RMSE values closest to the value of zero are associated with the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$. However, this is not the case when these field study data are obtained from a farm with pre-treatment group means less than 100 epg - in this case we see that the majority of RMSE values closest to the value of zero are associated with percentage estimates involving a negative control group (Table 5.3), but when we consider treatment group sample sizes of 15 animals; the lowest RMSE value is associated with $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$.

For data simulated with 15 and 30 epg data based on farms with pre-treatment group means either between 100-150 epg (inclusive) or greater than 150 epg, we observe that the majority of RMSE values closest to zero are associated with the percentage estimate $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$. Although, for these same field study data based on a farm with pre-treatment group means less than 100 epg, we observe that the majority of RMSE values closest to zero are associated with the percentage estimate: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$.

Field study data	Percentage Estimates (PEs) [1]	15	20	30	40	50 [2]
15EPG_McM_SCFT	PE1	3.95	2.44	1.51	1.08	0.90
	PE2	2.21	1.35	0.71	0.53	0.44
	PE3	2.55	2.66	2.80	2.90	2.95
	PE4	4.13	2.56	1.45	1.02	0.84
	PE5	4.32	4.57	4.84	5.01	5.10
	PE6	3.23	1.96	1.16	0.83	0.69
15EPG_McM	PE1	46.86	49.05	51.29	53.87	56.01
	PE2	30.57	32.55	34.86	39.18	40.86
	PE3	46.02	49.55	51.54	54.09	56.89
	PE4	41.20	42.94	44.99	47.83	49.77
	PE5	6.83	4.53	3.91	2.87	2.20
	PE6	42.80	45.04	48.64	51.25	53.69
30EPG_McM1	PE1	41.44	46.56	49.59	51.73	52.97
	PE2	28.48	30.32	32.86	35.63	38.33
	PE3	40.37	42.89	45.46	47.21	49.93
	PE4	36.76	38.71	40.09	42.66	45.70
	PE5	5.77	3.63	3.15	2.57	2.19
	PE6	43.13	45.15	48.17	50.79	51.96
30EPG_McM2	PE1	46.77	49.76	51.86	53.94	56.02
	PE2	29.10	30.76	32.83	35.65	37.94
	PE3	45.86	44.98	42.82	40.17	39.01
	PE4	37.83	39.38	42.16	44.85	46.03
	PE5	6.41	3.86	3.28	2.89	2.24
	PE6	45.15	44.81	43.56	41.75	39.95

Table 5.1: RMSE values of percentage estimates based on a farm with pre-treatment group means greater than 150 epg (highlighted cells have the lowest RMSE values in a given scenario)

[1] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[2] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [3]	15	20	30	40	50 [4]
15EPG_McM_SCFT	PE1	6.54	4.05	3.45	3.22	3.41
	PE2	4.50	2.65	2.74	2.84	2.73
	PE3	9.44	8.68	9.03	9.59	8.96
	PE4	7.23	4.74	4.98	5.15	4.96
	PE5	19.59	20.95	22.85	23.79	22.83
	PE6	8.72	4.32	3.10	2.86	3.15
15EPG_McM	PE1	29.83	33.91	36.54	38.78	40.21
	PE2	47.76	52.99	54.99	56.31	58.77
	PE3	54.61	55.75	56.98	58.22	60.33
	PE4	40.00	44.22	48.54	49.76	51.11
	PE5	16.73	16.44	15.02	13.38	12.66
	PE6	56.70	61.01	62.73	65.35	67.74
30EPG_McM1	PE1	30.90	35.49	37.84	39.29	41.26
	PE2	48.05	55.32	56.61	58.55	59.92
	PE3	59.51	59.16	57.21	55.89	54.37
	PE4	42.88	45.08	46.29	48.08	50.22
	PE5	15.73	10.77	8.73	7.02	6.53
	PE6	55.93	63.62	62.92	64.94	66.36
30EPG_McM2	PE1	26.00	32.84	34.95	36.89	39.57
	PE2	47.76	51.62	52.82	55.31	57.49
	PE3	66.17	60.98	58.49	56.26	54.14
	PE4	39.28	42.45	44.24	46.86	49.99
	PE5	51.06	11.56	10.03	8.81	6.98
	PE6	59.26	65.32	66.29	68.77	70.21

Table 5.2: RMSE values of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg (highlighted cells have the lowest RMSE values in a given scenario)

[3] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0 T_{14}}{T_0 C_{14}}\right)\%$.

[4] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [5]	15	20	30	40	50 [6]
15EPG_McM_SCFT	PE1	23.96	17.46	13.31	11.53	9.58
	PE2	21.56	17.66	14.55	11.57	10.05
	PE3	35.32	31.94	32.24	29.62	27.46
	PE4	20.28	17.55	14.72	12.26	10.99
	PE5	18.96	21.06	24.03	25.99	27.62
	PE6	22.72	16.75	12.03	9.41	8.23
15EPG_McM	PE1	36.27	41.95	43.68	45.49	47.86
	PE2	23.65	26.74	29.31	31.29	34.55
	PE3	62.30	67.50	69.96	70.59	72.03
	PE4	20.08	20.24	19.84	17.56	15.84
	PE5	30.51	32.82	34.54	36.72	39.88
	PE6	53.49	60.42	62.42	65.55	67.64
30EPG_McM1	PE1	31.52	38.15	40.01	42.35	45.25
	PE2	31.60	33.29	35.65	37.62	39.89
	PE3	69.51	71.26	72.45	74.65	76.66
	PE4	27.01	26.37	24.32	22.61	20.08
	PE5	34.78	37.65	39.48	40.26	42.74
	PE6	54.04	68.94	70.62	72.82	74.98
30EPG_McM2	PE1	29.15	38.64	40.94	42.63	45.45
	PE2	26.95	32.03	35.67	38.97	40.44
	PE3	54.87	58.41	60.91	62.37	64.42
	PE4	21.61	22.42	19.86	18.78	17.54
	PE5	24.99	24.70	26.86	29.29	31.51
	PE6	54.43	69.19	71.54	73.63	75.42

Table 5.3: RMSE values of percentage estimates based on a farm with pre-treatment group means less than 100 epg (highlighted cells have the lowest RMSE values in a given scenario)

[5] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[6] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

5.4 Prior sensitivity analysis

As mentioned in Chapter 2, the dependence of the posterior distribution on the prior distribution should always be assessed through a prior sensitivity analysis to confirm whether or not the posterior distribution is *data-driven* (i.e. a posterior distribution that is insensitive to the choice of prior) or *prior-driven* (i.e. a posterior distribution that is sensitive to the choice of prior), and this is typically carried out by considering a range of prior specifications and comparing the posterior distributions obtained.

In our case, the priors on the parameter σ in Section 5.2.1 will be modified such that each σ parameter will follow a continuous uniform distribution $Unif(1\text{x}10^{-16}, 1000)$. As a result, the prior for this parameter in each instance has an extended domain and with respect to variability, the domain being considered is approximately $\sigma^2 \sim Unif(0, 1\text{x}10^{6})$. Therefore, we are being even more conservative in our beliefs about the values that the parameter σ can take than before.

In light of any results and the priors considered as part of this study, it was also of interest to observe the average standard deviation estimates of Day 0 and Day 14 FEC data from negative control and positive treatment groups. It was also of interest to observe average standard deviations associated with different forms of individual-based egg count percentage reductions/changes being considered from positive treatment groups, i.e. $100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%$ and $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$. The value of the average was taken to be either the mean or median valued standard deviation, depending on whether or not the distribution of given standard deviations was of a normal nature or not (assessed using a *Shapiro-Wilk* normality test as described in Section 3.2.3).

5.4.1 Results

Tables 5.4-5.6 display similar information as in Tables 5.1-5.3, but as part of our priors, each σ parameter follows a continuous uniform distribution $Unif(1\text{x}10^{-16}, 1000)$. Again, RMSE values closest to the value of zero, for given types of field study data and treatment group sample sizes are highlighted in the shaded grey

cells in the Tables, as before.

For data simulated from the 15EPG_McM_SCFT field study data, similar results and conclusions are observed as before.

For data simulated based on 15 or 30 epg field study data based on a farm with pre-treatment group means greater than 150 epg, we observe no change with respect to the percentage estimates for which the lowest valued RMSE values are associated with. However, this is not the case when these field study data are obtained from a farm with pre-treatment group means either between 100-150 epg (inclusive) or less than 100 epg (Tables 5.5 and 5.6) - in this case we see that all of the RMSE values closest to the value of zero are associated with the percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$.

By observing the average standard deviation estimates for different types of FEC data (Table 5.7) and those associated with different forms of individual-based egg count percentage reductions/changes being considered (Table 5.8) we observe that all of the average standard deviation estimates, irrespective of the type of data being considered, lie within the range between 0 and 500.

Field study data	Percentage Estimates (PEs) [7]	15	20	30	40	50 [8]
15EPG_McM_SCFT	PE1	4.05	2.05	1.11	0.81	0.67
	PE2	2.42	1.38	0.66	0.48	0.40
	PE3	2.56	2.66	2.80	2.90	2.95
	PE4	4.94	2.61	1.26	0.92	0.77
	PE5	4.32	4.57	4.84	5.01	5.10
	PE6	3.52	1.80	0.94	0.68	0.57
15EPG_McM	PE1	31.07	30.95	28.76	22.07	19.97
	PE2	82.89	24.73	20.23	18.01	16.07
	PE3	36.74	37.95	39.42	42.33	45.44
	PE4	35.83	35.69	30.66	27.85	22.75
	PE5	6.87	4.50	2.98	1.75	1.10
	PE6	74.54	29.26	27.54	23.82	20.04
30EPG_McM1	PE1	25.04	26.38	28.72	31.24	33.63
	PE2	20.02	20.91	23.02	26.66	29.12
	PE3	32.77	33.79	34.73	36.98	39.56
	PE4	28.38	29.41	32.07	35.05	38.68
	PE5	5.78	3.56	2.31	1.08	0.87
	PE6	24.18	25.35	27.72	29.69	32.27
30EPG_McM2	PE1	29.73	30.63	33.34	35.26	37.67
	PE2	21.74	43.82	45.96	48.24	50.62
	PE3	35.42	34.98	33.62	34.51	33.01
	PE4	30.91	31.65	30.98	32.73	34.24
	PE5	6.48	3.82	2.76	1.32	1.05
	PE6	27.68	40.43	42.67	45.15	47.72

Table 5.4: RMSE values of percentage estimates based on a farm with pre-treatment group means greater than 150 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario)

[7] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[8] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [9]	15	20	30	40	50 [10]
15EPG_McM_SCFT	PE1	9.71	4.53	3.48	3.22	3.43
	PE2	6.17	3.01	3.01	2.99	2.98
	PE3	9.49	8.76	9.07	9.59	9.00
	PE4	9.48	5.41	5.47	5.43	5.43
	PE5	19.61	20.93	22.83	23.78	22.84
	PE6	11.79	4.80	3.36	3.03	3.38
15EPG_McM	PE1	12.63	13.59	12.43	10.02	8.75
	PE2	24.78	26.55	28.74	30.65	31.27
	PE3	41.71	41.08	39.28	37.86	34.54
	PE4	25.68	27.67	29.63	31.72	33.68
	PE5	16.83	16.48	14.07	13.24	11.57
	PE6	22.91	24.78	26.87	27.29	29.95
30EPG_McM1	PE1	15.13	9.75	8.62	6.57	5.88
	PE2	25.79	27.73	30.37	31.28	34.24
	PE3	47.06	45.89	42.01	39.43	38.08
	PE4	28.30	29.27	31.67	35.65	37.00
	PE5	15.72	10.89	9.76	7.21	6.83
	PE6	22.94	25.46	27.24	29.35	32.02
30EPG_McM2	PE1	19.33	17.24	15.64	13.47	11.32
	PE2	26.79	28.92	30.71	32.62	34.58
	PE3	53.23	50.21	48.31	45.26	43.34
	PE4	27.25	29.89	31.21	33.58	35.61
	PE5	31.08	22.28	20.09	18.57	16.28
	PE6	24.98	26.76	28.45	30.66	32.33

Table 5.5: RMSE values of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario)

[9] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0 T_{14}}{T_0 C_{14}}\right)\%$.

[10] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Field study data	Percentage Estimates (PEs) [11]	15	20	30	40	50 [12]
15EPG_McM_SCFT	PE1	23.70	17.12	12.85	11.48	9.55
	PE2	25.35	19.30	14.99	11.53	10.00
	PE3	24.08	20.61	18.49	18.18	19.30
	PE4	22.90	18.79	15.06	12.26	10.94
	PE5	18.99	21.00	24.07	26.03	27.62
	PE6	22.72	15.84	11.07	8.82	7.85
15EPG_McM	PE1	16.76	19.07	15.21	13.21	11.04
	PE2	20.47	19.76	18.04	15.64	14.02
	PE3	48.30	50.05	49.37	50.22	48.76
	PE4	20.21	19.26	17.24	15.28	13.62
	PE5	30.51	32.78	34.28	36.75	38.98
	PE6	21.71	23.10	25.75	26.86	29.99
30EPG_McM1	PE1	18.76	20.51	17.92	15.48	13.75
	PE2	33.04	32.41	30.65	28.79	25.45
	PE3	55.39	54.56	53.62	50.06	49.59
	PE4	29.74	29.05	28.61	27.28	25.44
	PE5	34.74	37.68	39.29	40.06	41.46
	PE6	23.95	27.43	29.76	31.28	33.63
30EPG_McM2	PE1	14.22	17.89	16.54	13.22	12.67
	PE2	23.62	23.48	21.47	20.08	19.42
	PE3	44.55	45.24	46.26	47.88	48.92
	PE4	23.10	22.46	21.01	19.86	18.77
	PE5	24.99	24.65	23.05	22.01	21.17
	PE6	25.00	29.84	32.64	33.78	34.47

Table 5.6: RMSE values of percentage estimates based on a farm with pre-treatment group means less than 100 epg as part of prior sensitivity analysis (highlighted cells have the lowest RMSE values in a given scenario)

[11] PE1: $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, PE2: $100\left(1-\frac{T_{14}}{T_0}\right)\%$, PE3: $\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%/n_{treat}$, PE4: $100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$, PE5: $\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%/n_{treat}$ and PE6: $100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%$.

[12] Values represent the treatment group sample sizes n_{treat} and $n_{control}$ considered.

Type of Data	Day	Treatment Group	Average standard deviations from groups with pre-treatment means >150 epg	Average standard deviations from groups with 100 epg ≤ pre-treatment means ≤ 150 epg	Average standard deviations from groups with pre-treatment group means <100 epg
	0	Control	272.29	127.76	84.74
30EPG_McM1	14	Control	257.07	104.38	126.31
	0	Positive Treatment	235.86	181.99	65.37
	14	Positive Treatment	35.3	43.39	15.43
	0	Control	274.91	169.05	100.42
30EPG_McM2	14	Control	261.77	112.54	145.67
	0	Positive Treatment	264.69	160.08	65.56
	14	Positive Treatment	44.56	27.44	11.21
	0	Control	258.65	133.25	93.2
15EPG_McM	14	Control	235.89	135.74	134.8
	0	Positive Treatment	242.82	156.71	60.44
	14	Positive Treatment	35.55	37.01	12.18
	0	Control	259.66	144.81	70.71
15EPG_McM_SCFT	14	Control	227.68	123.36	150.49
	0	Positive Treatment	239.8	158.6	61.84
	14	Positive Treatment	32.87	36.77	2.77

Table 5.7: Average standard deviation estimates of FEC data from treatment groups

Pre-treatment Group Mean Range	Type of Data	Form of individual % reductions (%)	Average standard deviation estimates
	30EPG_McM1	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	24.14
	30EPG_McM1	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	34.1
	30EPG_McM2	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	26.1
mean>150 epg	30EPG_McM2	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	31.28
	15EPG_McM	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	17.39
	15EPG_McM	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	30.45
	15EPG_McM_SCFT	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	16.04
	15EPG_McM_SCFT	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	27.39
	30EPG_McM1	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	66.53
	30EPG_McM1	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	52.13
	30EPG_McM2	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	42.3
100 epg≤mean≤150 epg	30EPG_McM2	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	43.91
	15EPG_McM	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	55.57
	15EPG_McM	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	46.36
	15EPG_McM_SCFT	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	36.64
	15EPG_McM_SCFT	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	36.35
	30EPG_McM1	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	33.65
	30EPG_McM1	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	43.21
	30EPG_McM2	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	24.65
mean<100 epg	30EPG_McM2	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	49.83
	15EPG_McM	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	30.92
	15EPG_McM	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	41.25
	15EPG_McM_SCFT	$100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)$	22.73
	15EPG_McM_SCFT	$100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$	38.88

Table 5.8: Average standard deviation estimates associated with individual-based egg count percentage reductions/changes

5.5 Are there any indistinguishable RMSE values?

With the results presented in Section 5.3, one may notice that some of the lowest obtained RMSE values may be of similar value to other RMSE values in a given scenario, presented in Tables 5.1-5.3. As a result, the question of "*Are there any lowest RMSE values that are indistinguishable to other RMSE values for a given scenario?*" is one of interest and, by extension, one would be able to conclude on whether or not the accuracies of different percentage estimates be significant (or not significant) from one another.

To investigate this, we explore the theory of assessing whether or not two variances are significantly different from one another and examine how this theory can be adapted when considering RMSE values.

5.5.1 The confidence interval for the ratio of two population RMSE values

It can be shown that, assuming data are independent and the two populations under consideration are normally distributed, then the $100(1-\alpha)\%$ confidence interval for the ratio of two population variances is given by

$$\frac{s_1^2}{s_2^2}F_{(1-\alpha)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2}F_{(\alpha)}, \tag{5.1}$$

where σ_1^2 and σ_2^2 are the population variances for populations 1 and 2 respectively, s_1^2 and s_2^2 are the sample variances from the samples of populations 1 and 2 respectively, $F_{(1-\alpha)}$ is the F-statistic on $n_1 - 1$ and $n_2 - 1$ degrees of freedom and $F_{(\alpha)}$ is the F-statistic on $n_2 - 1$ and $n_1 - 1$ degrees of freedom, where n_1 and n_2 are the sample sizes of the samples from populations 1 and 2, respectively. It is also worth noting that this is based on a one-tailed test being considered for our purposes since we will be examining if any RMSE values are significantly *greater*

than the lowest RMSE values that were obtained for given scenarios, but this can easily be edited to consider a two-tailed test, by replacing α with $\frac{\alpha}{2}$.

Upon taking the square root of confidence interval (5.1), one can obtain a confidence interval for the ratio of two population standard deviations given by (5.2):

$$\frac{s_1}{s_2}\sqrt{F_{(1-\alpha)}} \leq \frac{\sigma_1}{\sigma_2} \leq \frac{s_1}{s_2}\sqrt{F_{(\alpha)}}. \tag{5.2}$$

Now we know that, with respect to estimators, Equation (B.1) tells us that the $MSE = Var[\hat{\theta}] + (Bias)^2$ (essentially saying that the MSE is a variance and the RMSE is a standard deviation). So we can re-arrange Equation (B.1) and obtain

$$Var[\hat{\theta}] = MSE - (Bias)^2 \tag{5.3}$$

and by substituting Equation (5.3) into the confidence interval (5.1) and assuming unbiasedness of the two estimators under consideration, we obtain:

$$\frac{\hat{MSE}_1}{\hat{MSE}_2}F_{(1-\alpha)} \leq \frac{MSE_1}{MSE_2} \leq \frac{\hat{MSE}_1}{\hat{MSE}_2}F_{(\alpha)}, \tag{5.4}$$

where, for estimators $\hat{\theta_1}$ and $\hat{\theta_2}$, MSE_1 and MSE_2 are the population mean squared errors for these estimators respectively and $\hat{MSE}_1$ and $\hat{MSE}_2$ are the estimated mean squared errors based on samples for the estimators $\hat{\theta_1}$ and $\hat{\theta_2}$, respectively. Upon taking the square root of the confidence interval (5.4) we then obtain (5.5):

$$\frac{\hat{RMSE}_1}{\hat{RMSE}_2}\sqrt{F_{(1-\alpha)}} \leq \frac{RMSE_1}{RMSE_2} \leq \frac{\hat{RMSE}_1}{\hat{RMSE}_2}\sqrt{F_{(\alpha)}}. \tag{5.5}$$

The interpretation accompanying these intervals is simple: if the confidence in-

terval spans the value of one then one population variance, MSE or RMSE is not significantly greater from the other, i.e. indistinguishable from one another. Otherwise, one is significantly greater than the another with significance level α.

The ratio of values and confidence interval are able to tell us how many times one population measure is greater than the other. As an example, if $\frac{RMSE_1}{RMSE_2}$ was estimated to be 1.80 with a 95% confidence interval of (1.67,1.98) then we conclude that the $RMSE_1$ (associated with estimator $\hat{\theta_1}$) would be significantly greater than the $RMSE_2$ value (associated with estimator $\hat{\theta_2}$), and it is estimated as being 1.80 times greater than $RMSE_2$. It is also highly likely that the population $RMSE_1$ value associated with $\hat{\theta_1}$ can be at least 1.67 times greater, but at most be 1.98 times greater than the population $RMSE_2$ value associated with $\hat{\theta_2}$, upon repeated sampling of the population.

5.5.2 A Bootstrapped version of the confidence interval for the ratio of two population RMSE values

Two limitations presented with the confidence intervals discussed thus far, is that we assume that the two populations are normally distributed and the estimators considered are unbiased, which may not always be satisfied.

As a result, a Bootstrapping approach for obtaining a confidence interval for the ratio of two RMSE values could be considered. In essence if we were to consider our simulation study methodology, we are able to obtain 1000 squared errors, i.e. one thousand $(\hat{\theta}-\theta)^2$, for each given percentage estimate. We propose that if one were to randomly sample the 1000 squared errors for one percentage estimate and also randomly sample the 1000 squared errors of the percentage estimate that has the lowest RMSE value (in a given scenario), for say 1000 iterations, then one would be able to take the $\sqrt{mean(\cdot)}$ of each of these 1000 Bootstrapped samples of squared errors for each percentage estimate. Then one could take the ratio of each of these RMSE values for each iteration in turn (with those associated with the percentage estimate having the lowest RMSE value being assigned to the denominator of the ratio). As a result, one would obtain a sampling distribution for the ratio of two RMSE values and could take the $\frac{\alpha}{2}$ and $(1-\frac{\alpha}{2})$ percentiles

of the sampling distribution for a significance level α. Hence, one could obtain a 95% Bootstrapped percentile interval for the ratio of two RMSE values.

Now we will check to see if using the Bootstrapped approach presented in this Section gives rise to similar results with the theoretical, explicit intervals presented earlier.

5.5.3 Bootstrap approach vs. theoretical approach

We know it can be shown that if data are independent and if two populations under consideration are normally distributed, then the theoretical, explicit confidence intervals can be used to give inference on the ratio of two population variances, standard deviations, etc. Here, we shall use *R/RStudio* software to compare the results obtained from using explicit confidence interval formulae presented and the Bootstrapping approach proposed, using large sampled, normally distributed data, to investigate if both approaches give similar results.

As a means of examining the two approaches, we can simulate two sets of 1000 samples from a normal distribution (one could say these were samples of estimators, for instance) having defined location parameters μ_1 and μ_2 and standard deviations σ_1 and σ_2, respectively. In addition, σ_1 and σ_2 will be chosen such that the values of the ratio $\frac{\sigma_1}{\sigma_2}$ that will be considered are 1, 1.1, 1.5 and 2. We shall utilise confidence interval (5.2) as a means of convenience and gaining a confidence interval for two population RMSE values, since under the assumption of unbiasedness, these confidence intervals and estimates are essentially equivalent. The following R code was used to calculate the explicit confidence interval formulae for the ratio of two population standard deviations:

```
#Two independent samples from normal distribution,
#mu1, mu2, sd1 & sd2 to be defined below:
sample1<-rnorm(1000,mean=mu1,sd=sd1)
sample2<-rnorm(1000,mean=mu2,sd=sd2)
alpha<-0.05

Ratio_and_CI<-function(s1,s2){
 perc<-1-alpha
  x<- (var(s1)/var(s2))*(1/qf(perc,length(s1)-1,length(s2)-1))
  y<- (var(s1)/var(s2))*(qf(perc,length(s2)-1,length(s1)-1))
  Ratio<-sd(s1)/sd(s2)
  LCL <- sqrt(x)
  UCL <- sqrt(y)
  vec<-c(round(Ratio,2),round(LCL,2),round(UCL,2))
  return(vec)}
Ratio_and_CI(sample1,sample2)
```

The following results were then obtained using the above code, with $\mu_1 = 10$ and $\mu_2 = 5$:

Ratio considered	Ratio Estimate obtained	95% Confidence Interval (CI)	CI Width
1	0.98	(0.93, 1.03)	0.1
1.1	1.13	(1.08, 1.19)	0.11
1.5	1.42	(1.35, 1.50)	0.15
2	2.03	(1.93, 2.14)	0.21

Table 5.9: Ratios of RMSE values obtained using theoretical calculations

Now that we have results obtained from theoretical calculations, we can consider the coding for the Bootstrapping approach discussed earlier. Below is the

R/RStudio code which makes use of the already available two sets of 1000 normally distributed samples, with $\mu_1 = 10$ and $\mu_2 = 5$, where the standard deviation is the estimate of interest to be evaluated from Bootstrapped samples, from which the ratio of the two would be considered:

```
a<-rep(NA,1000)
b<-rep(NA,1000)

#Re-sampling:
for(i in 1:length(b)){
  newd<-sample(sample1,length(sample1),replace=TRUE)
  b[i]<-sd(newd)}

for(i in 1:length(a)){
  newd<-sample(sample2,length(sample2),replace=TRUE)
  a[i]<-sd(newd)}

ratio1<-round(sd(sample1)/sd(sample2),2)
names(ratio1)<-c("Ratio")

z<-as.vector(b/a)
z<-subset(z,is.infinite(z)==FALSE)

CI1<-c(round(quantile(z,c(alpha/2),na.rm=TRUE),digits=2),
round(quantile(z,c(1-(alpha/2)),na.rm=TRUE),digits=2))
names(CI1)<-c(paste(100*(1-alpha),"%", "LCL",sep=""),
paste(100*(1-alpha),"%", "UCL",sep=""))
results1<-c(ratio1,CI1)
print(results1)
```

Utilising the Bootstrapping approach presented above, the following results were

then obtained:

Ratio considered	Ratio Estimate obtained	95% Confidence Interval (CI)	CI Width
1	0.98	(0.92, 1.04)	0.12
1.1	1.13	(1.07, 1.21)	0.14
1.5	1.42	(1.34, 1.51)	0.17
2	2.03	(1.91, 2.17)	0.26

Table 5.10: Ratios of RMSE values obtained using developed Bootstrapping approach

From Tables 5.9 and 5.10, we observe very similar confidence intervals being obtained for the various ratios of standard deviations considered, from both approaches. In fact, using the proposed Bootstrapping approach seems to have produced only slightly wider confidence intervals being obtained.

In the Section to follow, the ratios of RMSE values and the lowest RMSE values (for a given scenario) are presented, as well as 95% Bootstrapped percentile intervals using our proposed approach, for those RMSE values obtained and presented in Section 5.3.

5.5.4 Results

From Tables 5.11-5.13, we observe that the majority of the ratios of RMSE values are significantly greater than the lowest RMSE value obtained, for a given scenario. In fact, there are only two instances where the 95% Bootstrapped percentile intervals span the value of one and these are highlighted in the shaded grey cells in these Tables.

An example of how to interpret the results in the Tables is as follows: if we consider the first entry in Table 5.11 which has the value 1.79 - this value is based on the ratio of the RMSE value 3.95 and the lowest RMSE value 2.21 obtained in this scenario and highlighted in Table 5.1, which are associated with the percentage estimates $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and $100\left(1-\frac{T_{14}}{T_0}\right)\%$, respectively and the

confidence interval (1.71, 1.87) is the associated 95% Bootstrapped percentile interval for this ratio. Subsequently, all other ratios are obtained in this scenario with respect to the lowest RMSE value 2.21 for treatment group sample sizes of 15, where 15EPG_McM_SCFT have been used to simulate data. Similar interpretations can be obtained for ratios associated with other treatment group sample sizes, diagnostic sensitivities considered and the lowest RMSE values obtained.

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
15EPG_McM_SCFT	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.79 (1.71, 1.87)	1.81 (1.72, 1.89)	1.99 (1.89, 2.07)	2.04 (1.94, 2.15)	2.05 (1.97, 2.16)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	-	-	-	-	-
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	1.15 (1.10, 1.21)	1.97 (1.87, 2.08)	3.68 (3.48, 3.90)	5.47 (5.18, 5.81)	6.70 (6.42, 7.12)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	1.87 (1.87, 1.88)	1.90 (1.89, 1.90)	1.91 (1.91, 1.91)	1.92 (1.91, 1.92)	1.91 (1.90, 1.92)
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	1.95 (1.87, 2.06)	3.39 (3.21, 3.57)	6.37 (6.03, 6.75)	9.45 (8.96, 10)	11.59 (11.07, 12.28)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	1.46 (1.40, 1.51)	1.45 (1.40, 1.51)	1.53 (1.48, 1.59)	1.57 (1.51, 1.63)	1.57 (1.53, 1.65)
15EPG_McM	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	6.86 (6.13, 7.62)	10.83 (10.10, 11.69)	13.12 (12.34, 13.85)	18.77 (17.84, 20.21)	25.46 (24.37, 27.60)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	4.48 (4.08, 4.93)	7.19 (6.77, 7.67)	8.92 (8.37, 9.45)	13.65 (12.74, 14.43)	18.57 (17.80, 20.19)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	6.74 (6.08, 7.45)	10.94 (9.92, 12.03)	13.18 (12.48, 14.15)	18.85 (18.28, 20.65)	25.86 (24.04, 27.19)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	6.03 (5.45, 6.60)	9.48 (8.94, 10.06)	11.51 (10.68, 12.04)	16.67 (15.48, 17.46)	22.62 (21.12, 24.07)
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	-	-	-	-	-
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	6.27 (5.58, 6.98)	9.94 (9.24, 10.64)	12.44 (11.77,13.40)	17.86 (17.64, 20.06)	24.40 (22.56, 25.30)
	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	7.18 (6.48, 7.92)	12.83 (11.78, 14.05)	15.74 (14.86, 16.78)	19.94 (19.48, 22.05)	24.19 (22.88, 25.81)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	4.94 (4.51, 5.40)	8.35 (7.91, 8.84)	10.43 (9.82, 11.20)	13.86 (13.72, 16.64)	17.50 (16.87, 18.96)

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
30EPG_McM1	$\Sigma_j 100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	7.00 (6.23, 7.73)	11.82 (10.55, 13.10)	14.43 (13.92, 15.88)	18.37 (17.88, 20.25)	22.80 (21.70, 24.51)
	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)$	6.37 (5.81, 6.95)	10.66 (10.09, 11.28)	12.73 (11.92, 13.49)	16.60 (16.03, 17.94)	20.87 (19.58, 22.35)
	$\Sigma_j 100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)/n_{treat}$	-	-	-	-	-
	$100\left(1 - \frac{C_0 T_{14}}{T_0 C_{14}}\right)$	7.47 (6.66, 8.26)	12.44 (11.47, 13.55)	15.29 (14.52, 16.56)	19.76 (18.61, 21.06)	23.73 (23.17, 26.09)
30EPG_McM2	$100\left(1 - \frac{T_{14}}{C_{14}}\right)$	7.30 (6.54, 8.04)	12.89 (12.14, 13.74)	15.81 (14.33, 16.21)	18.66 (16.60, 18.74)	25.01 (22.92, 25.82)
	$100\left(1 - \frac{T_{14}}{T_0}\right)$	4.54 (4.17, 4.94)	7.97 (7.54, 8.38)	10.01 (9.45, 10.72)	12.34 (11.55, 13.09)	16.94 (15.50, 17.61)
	$\Sigma_j 100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	7.15 (6.45, 7.95)	11.65 (10.55, 12.81)	13.05 (12.29, 13.76)	13.90 (13.31, 15.07)	17.42 (16.96, 19.20)
	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)$	5.90 (5.49, 10.66)	10.20 (9.71, 10.67)	12.85 (12.32, 13.93)	15.52 (14.44, 16.22)	20.55 (18.64, 20.98)
	$\Sigma_j 100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)/n_{treat}$	-	-	-	-	-
	$100\left(1 - \frac{C_0 T_{14}}{T_0 C_{14}}\right)$	7.04 (6.34, 7.85)	11.61 (10.73, 12.56)	13.28 (12.46, 14.26)	14.45 (13.40, 15.17)	17.83 (16.93, 19.17)

Table 5.11: Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with pre-treatment means >150 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one)

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
15EPG_McM_SCFT	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.45 (1.28, 1.63)	1.53 (1.44, 1.62)	1.26 (1.21, 1.30)	1.13 (1.10, 1.17)	1.25 (1.21, 1.28)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	-	-	-	-	-
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	2.10 (1.86, 2.34)	3.28 (3.11, 3.45)	3.30 (3.22, 3.37)	3.38 (3.32, 3.44)	3.28 (3.20, 3.36)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	1.61 (1.56, 1.66)	1.79 (1.78, 1.79)	1.82 (1.81, 1.82)	1.81 (1.81, 1.82)	1.82 (1.81, 1.82)
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	4.35 (3.89, 4.86)	7.91 (7.65, 8.18)	8.34 (8.16, 8.51)	8.38 (8.22, 8.56)	8.36 (8.18, 8.55)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	1.94 (1.83, 2.06)	1.63 (1.53, 1.73)	1.13 (1.09, 1.17)	1.01 (0.98, 1.04)	1.15 (1.10, 1.20)
15EPG_McM	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.78 (1.59, 1.96)	2.06 (1.91, 2.24)	2.43 (2.40, 2.73)	2.90 (2.70, 3.07)	3.18 (3.12, 3.52)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	2.85 (2.63, 3.08)	3.22 (3.02, 3.42)	3.66 (3.37, 3.79)	4.21 (4.07, 4.56)	4.64 (4.18, 4.75)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	3.26 (3.02, 3.51)	3.39 (3.16, 3.62)	3.79 (3.49, 4.00)	4.35 (4.14, 4.67)	4.77 (4.38, 4.97)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	2.39 (2.25, 2.54)	2.69 (2.56, 2.82)	3.23 (3.13, 3.52)	3.72 (3.46, 3.89)	4.04 (3.92, 4.43)
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	-	-	-	-	-
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	3.39 (3.00, 3.80)	3.71 (3.36, 4.02)	4.18 (3.79, 4.27)	4.88 (4.54, 5.16)	5.35 (5.07, 5.73)
	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.96 (1.71, 2.25)	3.30 (2.95, 3.68)	4.33 (4.10, 4.61)	5.60 (5.28, 6.03)	6.32 (5.80, 6.62)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	3.05 (2.64, 3.45)	4.95 (4.59, 5.31)	6.48 (6.22, 7.03)	8.34 (7.82, 8.83)	9.18 (8.38, 9.42)

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
30EPG_McM1	$\Sigma_j 100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	3.78 (3.34, 4.25)	5.49 (4.98, 6.04)	6.55 (5.79, 6.58)	7.96 (7.31, 8.23)	8.33 (7.80, 8.86)
	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)$	2.73 (2.44, 3.06)	4.19 (3.86, 4.54)	5.30 (5.01, 5.71)	6.85 (6.41, 7.25)	7.69 (7.11, 8.03)
	$\Sigma_j 100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)/n_{treat}$	-	-	-	-	-
	$100\left(1 - \frac{C_0 T_{14}}{T_0 C_{14}}\right)$	3.56 (3.07, 4.07)	5.91 (5.52, 6.62)	7.21 (6.50, 7.39)	9.25 (8.61, 9.84)	10.16 (9.15, 10.35)
30EPG_McM2	$100\left(1 - \frac{T_{14}}{C_{14}}\right)$	-	2.90 (2.73, 3.08)	3.48 (3.24, 3.70)	4.19 (4.07, 4.58)	5.67 (5.14, 5.82)
	$100\left(1 - \frac{T_{14}}{T_0}\right)$	1.84 (1.63, 2.08)	4.35 (4.06, 4.64)	5.27 (5.05, 5.69)	6.28 (5.94, 6.70)	8.24 (7.55, 8.49)
	$\Sigma_j 100\left(1 - \frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	2.54 (2.28, 2.85)	5.28 (4.97, 5.63)	5.83 (5.30, 6.01)	6.39 (6.25, 7.09)	7.76 (6.99, 7.86)
	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)$	1.51 (1.37, 1.69)	3.84 (3.63, 4.10)	4.41 (4.08, 4.65)	5.32 (4.92, 5.56)	7.16 (6.91, 7.78)
	$\Sigma_j 100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)/n_{treat}$	1.19 (1.08, 1.33)	-	-	-	-
	$100\left(1 - \frac{C_0 T_{14}}{T_0 C_{14}}\right)$	2.28 (2.02, 2.58)	5.55 (5.24, 5.90)	6.61 (6.27, 7.04)	7.81 (7.21, 8.19)	10.06 (9.56, 10.76)

Table 5.12: Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with 100 epg $\leq$ pre-treatment means $\leq$ 150 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one)

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
15EPG_McM_SCFT	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.26 (1.15, 1.38)	1.04 (0.97, 1.12)	1.11 (1.02, 1.20)	1.23 (1.13, 1.34)	1.16 (1.09, 1.25)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	1.14 (1.06, 1.22)	1.05 (1.01, 1.10)	1.21 (1.16, 1.26)	1.23 (1.18, 1.29)	1.22 (1.17, 1.27)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	1.86 (1.7, 2.03)	1.91 (1.69, 2.13)	2.68 (2.36, 3)	3.15 (2.81, 3.52)	3.34 (2.98, 3.70)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	1.07 (1.01, 1.12)	1.05 (1.01, 1.09)	1.22 (1.17, 1.28)	1.30 (1.25, 1.36)	1.34 (1.28, 1.39)
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	-	1.26 (1.17, 1.35)	2 (1.85, 2.15)	2.76 (2.59, 2.95)	3.36 (3.16, 3.59)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	1.20 (1.10, 1.31)	-	-	-	-
15EPG_McM	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.81 (1.63, 1.99)	2.07 (1.87, 2.28)	2.20 (2.08, 2.33)	2.59 (2.43, 2.73)	3.02 (2.75, 3.10)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	1.18 (1.13, 1.23)	1.32 (1.25, 1.39)	1.48 (1.40, 1.60)	1.78 (1.63, 1.84)	2.18 (2.10, 2.37)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	3.10 (2.91, 3.32)	3.33 (3.11, 3.57)	3.53 (3.13, 3.53)	4.02 (3.81, 4.29)	4.55 (4.10, 4.62)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	-	-	-	-	-
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	1.52 (1.47, 1.58)	1.62 (1.55, 1.70)	1.74 (1.59, 1.79)	2.09 (2.04, 2.30)	2.52 (2.45, 2.77)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	2.66 (2.43, 2.92)	2.99 (2.73, 3.25)	3.15 (2.85, 3.22)	3.73 (3.40, 3.85)	4.27 (4.07, 4.59)
	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.17 (1.05, 1.28)	1.45 (1.28, 1.62)	1.65 (1.56, 1.77)	1.87 (1.75, 1.97)	2.25 (2.11, 2.40)

Field study data simulations were based on	Percentage Estimates (%)	$n_{control}$, $n_{treat} = 15$	$n_{control}$, $n_{treat} = 20$	$n_{control}$, $n_{treat} = 30$	$n_{control}$, $n_{treat} = 40$	$n_{control}$, $n_{treat} = 50$
30EPG_McM1	$100\left(1-\frac{T_{14}}{T_0}\right)$	1.17 (1.14, 1.21)	1.26 (1.22, 1.31)	1.47 (1.32, 1.50)	1.66 (1.57, 1.76)	1.99 (1.76, 1.99)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	2.57 (2.41, 2.73)	2.70 (2.53, 2.89)	2.98 (2.80, 3.16)	3.30 (3.12, 3.54)	3.82 (3.47, 3.93)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	-	-	-	-	-
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	1.29 (1.24, 1.33)	1.43 (1.39, 1.47)	1.62 (1.57, 1.78)	1.78 (1.71, 1.92)	2.13 (2.01, 2.26)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	2.00 (1.77, 2.24)	2.61 (2.19, 3.09)	2.90 (2.63, 3.01)	3.22 (3.03, 3.42)	3.73 (3.52, 3.96)
30EPG_McM2	$100\left(1-\frac{T_{14}}{C_{14}}\right)$	1.35 (1.21, 1.51)	1.72 (1.55, 1.92)	2.06 (1.92, 2.19)	2.27 (2.16, 2.45)	2.59 (2.36, 2.67)
	$100\left(1-\frac{T_{14}}{T_0}\right)$	1.25 (1.18, 1.32)	1.43 (1.31, 1.55)	1.80 (1.66, 1.86)	2.08 (1.88, 2.14)	2.31 (2.13, 2.40)
	$\Sigma_j 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)/n_{treat}$	2.54 (2.37, 2.72)	2.61 (2.42, 2.78)	3.07 (2.88, 3.27)	3.32 (3.11, 3.56)	3.67 (3.47, 3.90)
	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)$	-	-	-	-	-
	$\Sigma_j 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)/n_{treat}$	1.16 (1.09, 1.24)	1.10 (1.05, 1.15)	1.35 (1.24, 1.40)	1.56 (1.41, 1.62)	1.80 (1.70, 1.92)
	$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)$	2.52 (2.30, 2.74)	3.09 (2.72, 3.49)	3.60 (3.56, 4.03)	3.92 (3.70, 4.18)	4.30 (3.86, 4.39)

Table 5.13: Ratios of RMSE values (95% Bootstrapped percentile intervals) of percentage estimates based on a farm with pre-treatment means <100 epg (highlighted cells have 95% Bootstrapped percentile intervals that span the value of one)

5.6 Discussion

Results based on farm with pre-treatment group means greater than 150 epg:

For simulations based on 15EPG_McM_SCFT field study data, the percentage estimate using pre- and post-treatment group FECs from a positive treatment group only:

$$100\left(1-\frac{T_{14}}{T_0}\right)\%,$$

using a $NBII$ likelihood, had the lowest RMSE values associated with it, irrespective of the treatment group sample sizes considered. This is also the case when we consider the results from the prior sensitivity analysis that was conducted. As a result, it would appear that the posterior distributions obtained, for the percentage estimates considered for these data, are insensitive to the choice of prior being set for the parameter σ.

This percentage estimate for apparent efficacy/resistance has appeared in many studies for several livestock species (Kochapakdee *et al.* 1995; McKenna 2006; Denwood 2010; Levecke *et al.* 2012; Vidyashankar *et al.* 2012; Lester *et al.* 2013; Geurden *et al.* 2015; George *et al.* 2017) and is obtained through use of a paired study design involving a positive treatment group only. This design of experiment is often adopted, particularly in the instance of not being able to facilitate a negative control group due to having a smaller number of animals on farm. This design also has the advantage that all animals receive the considered treatment, effectively increasing the sample size to be featured as part of the experiment and decreases the cost of the investigation by requiring fewer animals for the same diagnostic sensitivity being utilised (Denwood 2010). Our simulation study is able to compliment these studies and justify the use of this percentage estimate with respect to it being the most accurate using the Bayesian methodologies adopted, based on FEC data being obtained using sensitive counting techniques (i.e. 1 epg) with pre-treatment group means being greater than 150 epg.

For data simulated with 15 or 30 epg field study data, all of the lowest RMSE

values were associated with

$$\frac{\sum_{j}^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}},$$

irrespective of the treatment group sample sizes considered and utilising a $\mathcal{N}_{trunc}$ likelihood. To consider this type of approach, we would only require to work with FECs from the same individual animal as part of a paired study design involving a positive treatment group only, though not much research has been dedicated to this concept (Cabaret and Berrag 2004). Our simulation study is able to compliment/support this earlier work by Cabaret and Berrag (2004), since the approach of averaging over individual-based egg count percentage reductions/changes has been observed to be the most accurate for various treatment group sample sizes for field study data obtained by, what could be considered as, less sensitive counting techniques (i.e. 15 or 30 epg). In our case, the individual-based egg count percentage reductions/changes are of a form which involves using the *Symmetrised Percentage Change* (SPC). According to Berry and Ayers (2006) this percentage estimate can present several theoretical benefits such as being bounded (i.e. between $\pm 100\%$). Having a bounded range means that the influence of outliers, is greatly reduced. In the case of FEC data being considered, it is often the case that a small number of individual animals will be shedding high numbers of eggs in their faeces, and consequently, these outliers can detrimentally affect the mean values of treatment groups.

It is also worth noting that the lowest RMSE values obtained for data simulated with 15 or 30 epg field study data as part of the prior sensitivity analysis, were associated with the same percentage estimate. As a result, it would appear that the posterior distributions obtained, for the percentage estimates based on 15 or 30 epg data, are insensitive to the choice of prior being set for the parameter σ. Overall, it would appear that FEC data obtained as part of a farm with pre-treatment group means greater than 150 epg are able to consistently produce, accurate percentage estimates based on the Bayesian methodology adopted as part of our study, irrespective of the priors on the dispersion/variability.

That being said, when wanting to apply these methodologies one would have to adopt one prior in practice. Given the average standard deviation estimates for different types of field study FEC data and individual-based egg count percentage reductions/changes (presented in Tables 5.7 and 5.8), all lie within the range of 0 to 500; these estimates would be better reflected and potentially achieved (with respect to convergence) with a prior of $\sigma \sim Unif(1\text{x}10^{-16}, 500)$, as opposed to a wider range $Unif(1\text{x}10^{-16}, 1000)$.

Results based on farm with 100 epg $\leq$ pre-treatment group means $\leq$ 150 epg:

For simulations based on 15EPG_McM_SCFT field study data, the percentage estimate using pre- and post-treatment group FECs from a positive treatment group only,

$$100\left(1 - \frac{T_{14}}{T_0}\right)\%$$

using a $NBII$ likelihood, had the lowest RMSE values associated with it for all treatment group sizes considered. This is also the case when we consider the results from the prior sensitivity analysis that was conducted. Again, it would appear that the posterior distributions obtained, for the percentage estimates considered for these data, are insensitive to the choice of prior being set for the parameter σ.

For simulations based on 15 or 30 epg field study data, the majority of the lowest RMSE values were associated with averaging over individual-based egg count percentage reductions/changes based on utilising the SPC. The results mentioned thus far, are consistent with those based on data from a farm with pre-treatment group means greater than 150 epg.

However, all of the lowest RMSE values obtained for data simulated with 15 or 30 epg field study data as part of the prior sensitivity analysis were associated with the following percentage reduction (where a $ZIPIG$ likelihood was utilised):

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%.$$

As a result, it would appear that when considering FEC data obtained using counting techniques with a 15 or 30 epg sensitivity, from a farm with pre-treatment group means between 100 epg and 150 epg (inclusive); the accuracy of the percentage estimates obtained are sensitive to the beliefs about the parameter σ, to an extent where a negative control group is required to obtain accurate percentage estimates, the more conservative one is about the variability of FEC data.

But it is worth remembering that, given the average standard deviation estimates for different types of field study FEC data and individual-based egg count percentage reductions/changes, these estimates would be better reflected and potentially achieved (with respect to convergence) with a prior of $\sigma \sim Unif(1\text{x}10^{-16}, 500)$, as opposed to a wider range $Unif(1\text{x}10^{-16}, 1000)$. Hence, any conclusions derived from the prior sensitivity analysis should be interpreted with this in mind.

Results based on farm with pre-treatment group means less than 100 epg:

For simulated data based on 15EPG_McM_SCFT field study data, we observe that the percentage estimates with the lowest RMSE values associated with them vary across different treatment group sample sizes. For treatment group sample sizes of 20 or above, the percentage estimate accurately estimated in the majority of cases, utilising a $NBII$ likelihood, was

$$100\left(1-\frac{C_0T_{14}}{T_0C_{14}}\right)\%.$$

This is the first instance, as part of this simulation study, that a percentage estimate based on a parallel group design has been considered the most accurate for data obtained using sensitive counting techniques (i.e. 1 epg). Lyndal-Murphy *et al.* (2014) advocates the use of this percentage estimate, but this recommen-

dation is based on simulated data for animal group sample sizes of 15 only and confidence intervals derived using the Delta method were used, which did involve correlations of FEC data; but not correlations of ln-transformed data as required. However, our simulation study is able to assess the robustness of percentage estimates by observing RMSE values via Bayesian methodologies, which do not depend on large sample normal approximations for FEC data, consider treatment groups of various sample sizes and various percentage estimates and also account for the distributions of different FEC data and correlations between Day 0 and Day 14 FEC data, upon simulation.

One would have to consider, however, the feasibility and practicability of this result and ask if it is worth running a FECRT with both positive treatment and negative control group sample sizes of 20 (or greater) cattle, where pre-treatment group means are less than 100 epg? One possible reason as to why this percentage estimate may have resulted in being the most accurate is that as part of the Defra project, composite screening FECs were performed, and only those herds with a composite count greater than 150 epg were enrolled onto a subsequent FECRT. However, despite these composite screening attempts, repeated individual baseline faecal egg sampling on some of these farms subsequently revealed mean egg counts less than 100 epg at the baseline sampling point before treatment. This highlights the variability in results that can be obtained between composite and individual samples etc. and this may have affected the results presented here. Indeed, various research groups have investigated the most appropriate methodology for conducting composite samples (Morgan *et al.* 2005; Calvete and Uriarte 2013; George *et al.* 2017). In fact, it would be considered unlikely that egg counts less than 100 epg would be associated with sufficient worm burdens and/or pasture contamination to justify anthelmintic use and, in such cases, would not be consistent with best practice recommendations (Coles *et al.* 1992; Coles *et al.* 2006).

For the same type of data, but involving treatment group sample sizes of 15 animals, we observe that the percentage estimate with the lowest RMSE value is (utilising a $\mathcal{N}_{trunc}$ likelihood):

$$\frac{\sum_{j}^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}},$$

which may be more favourable in terms of feasibility and practicability. Indeed, these are the same conclusions obtained for these data under the prior sensitivity analysis. This means that the posterior distributions obtained, for the percentage estimates considered using FEC data obtained using sensitive counting techniques (i.e. 1epg), are insensitive to the choice of prior being set for the parameter σ.

It is worth noting however, that one limitation to our study is that the smallest number of animals considered was 15 (i.e. a total of 30 animals in a given scenario). Upon reflection of the results, it would have been interesting to investigate the accuracy of percentage estimates associated with treatment group sample sizes of 10 animals. Published guidelines, such as Coles *et al.* (1992) and Coles *et al.* (2006), recommend the use of treatment group sample sizes between 10-15. In the scenario of having, what could be considered as, small treatment group sample sizes, one must ask if these sizes are able to provide sufficient power for these types of studies. Denwood (2010) comments on this, saying that prospective power calculations are rarely, performed prior to undertaking a FECRT study and, with respect to equine data, are usually found to be underpowered. As a result, it was more of interest to consider larger treatment group sample sizes as part of our simulation study, but the accuracy of percentage estimates with regards to smaller treatment group sample sizes, of say 10 animals, could be considered as part of future investigations.

For simulations based on 15 or 30 epg field study data being considered, all of the lowest RMSE values were associated with the SPC, which involved utilising a $ZIPIG$ likelihood. In comparison to the other farm data with different pre-treatment group mean ranges considered as part of this study, this is the first instance where averaging over individual-based egg count percentage reductions/changes based on utilising the SPC has not obtained the lowest RMSE values for 15 or 30 epg field study data. This could possibly be because when we consider farm data with pre-treatment group means less than 100 epg, zero

counts are more likely to feature as part of Day 0 and 14 data, and for averaging over individual-based egg count percentage reductions/changes, this is more likely to lead to individual-based egg count percentage reductions/changes being removed before evaluating the average of these or with lower individual-based egg count percentage reductions/changes being evaluated, i.e. possibly resulting in a larger difference being obtained between the average of the sampling distribution of percentage estimates and the known true/target percentage estimate, as part of evaluating the RMSE. However, evaluating the SPC and averaging over individual-based egg count percentage reductions/changes of the SPC form can still both be obtained using only a paired study design with a positive treatment group.

As part of the prior sensitivity analysis, however, the lowest RMSE values obtained based on 15 or 30 epg field study data were associated with the percentage reduction:

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%,$$

indicating that when considering FEC data obtained using counting techniques with a 15 or 30 epg sensitivity, from a farm with pre-treatment group means less than 100 epg; the accuracy of the percentage estimates obtained are sensitive to the beliefs about the variability of FEC data, to an extent where a negative control group is required.

It is worth remembering however, that any conclusions derived from the prior sensitivity analysis should be interpreted with the fact that the average standard deviation estimates, for different types of field study FEC data and individual-based egg count percentage reductions/changes, would be better reflected and potentially achieved (with respect to convergence) with a prior of $\sigma \sim Unif(1\text{x}10^{-16}, 500)$, as opposed to a wider range $Unif(1\text{x}10^{-16}, 1000)$.

Overall conclusions:

This simulation study, along with the prior sensitivity analysis, revealed that the

accuracy of the estimation of the percentage estimates considered can be influenced by treatment group sample sizes, the diagnostic sensitivity of the counting techniques used in obtaining the original FEC data, the prior beliefs about the dispersion and variability of the FEC data considered and the classifications of the pre-treatment group means obtained.

By considering this study as a whole, it could be recommended that a prior of $Unif(1\text{x}10^{-16}, 500)$ be adopted for the dispersion parameter σ for the different types of FEC and individual-based egg count percentage reduction/change data considered as part of this study. Subsequently, the following recommendations are based on the above prior being utilised as part of the Bayesian methodology considered as part of this study.

For FEC data obtained using a sensitive counting technique (1 epg sensitivity), when pre-treatment group means are greater than or equal to 100 epg, the percentage estimate based on pre- and post-treatment group FECs from a positive treatment group only (using a $NBII$ likelihood) was estimated the most accurate, for all treatment group sample sizes considered.

When considering FEC data obtained using a sensitive counting technique (1 epg sensitivity) but from a farm with pre-treatment group means less than 100 epg, for treatment group sample sizes of 20 or above the percentage estimate involving pre- and post-counts from both a negative control and positive treatment group (utilising a $NBII$ likelihood) was estimated the most accurate. For the same FEC data under consideration, but with treatment group sample size of 15 animals; one could utilise a paired study design involving a positive treatment group only, to average over individual-based egg count percentage reductions/changes utilising the SPC estimate (utilising a $\mathcal{N}_{trunc}$ likelihood), since this was the most accurate percentage estimate in this scenario.

For FEC data obtained using less sensitive counting techniques (i.e. 15 or 30 epg), when pre-treatment group means are greater than or equal to 100 epg, the approach of averaging over individual-based egg count percentage reductions/changes utilising the SPC estimate (utilising a $\mathcal{N}_{trunc}$ likelihood), from a positive treatment group only, was estimated the most accurate, for all treatment group sample sizes considered. When considering this same type of FEC data but

Sensitivities & Pre-treatment Mean Ranges \ Treatment Group Sizes		$n_{control} = 15$, $n_{treat} = 15$	$n_{control} \geq 20$, $n_{treat} \geq 20$
Hybrid Data (involving 1epg and 15epg McMaster Counts)	Pre-treatment means ≥100epg	$100\left(1 - \frac{T_{14}}{T_0}\right)\%$	
	Pre-treatment means < 100epg	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)\%$	$100\left(1 - \frac{T_{14}}{C_{14}}\right)\%$
15 and 30epg McMaster Data	Pre-treatment means ≥ 100epg	$\frac{\sum 100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)\%}{n_{treat}}$	
	Pre-treatment means < 100epg	$100\left(\frac{T_0 - T_{14}}{T_0 + T_{14}}\right)\%$	

Table 5.14: A summary of the most accurate percentage estimates to be estimated for a given diagnostic sensitivity and treatment group sample size

from a farm with pre-treatment group means being less than 100 epg, the SPC (utilising a $ZIPIG$ likelihood) was estimated the most accurate for all treatment group sample sizes considered as part of this study.

It is worth noting however, that in the majority of scenarios considered, there is an opportunity to adopt a paired study design involving a positive treatment group only, in order to obtain those percentage estimates that were most accurate. Table 5.14 highlights a summary of the above mentioned results.

With respect to answering the question "*Are there any percentage estimates whose accuracies are indistinguishable?*", we observed that the majority of the ratios of RMSE values were significantly greater than the lowest RMSE value obtained, for a given scenario, at the 5% significance level. Table 5.15 highlights a summary of these results, in which the most accurate percentage estimates are recommended (i.e. those with the lowest RMSE values) as well as those percentage estimates whose RMSE values were indistinguishable from the lowest RMSE values in a given scenario (these are represented by ✓).

			Percentage Estimates considered					
Diagnostic Sensitivity	Treatment Group Sample Sizes considered for $n_{control}$ and n_{treat}	Pre-treatment mean ranges	$100\left(1-\frac{T_{14}}{C_{14}}\right)\%$	$100\left(1-\frac{T_{14}}{T_0}\right)\%$	$\frac{\sum 100\left(1-\frac{T_{14,j}}{T_{0,j}}\right)\%}{n_{treat}}$	$100\left(\frac{T_0-T_{14}}{T_0+T_{14}}\right)\%$	$\frac{\sum 100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%}{n_{treat}}$	$100\left(1-\frac{T_{14}C_0}{T_0C_{14}}\right)\%$
Hybrid Data (involving 1epg and 15epg McMaster Counts)	15	Pre-treatment means >150epg	×	✓	×	×	×	×
	20							
	30							
	40							
	50							
	15	100epg ≤ Pre-treatment means ≤ 150epg	×	✓	×	×	×	×
	20							×
	30							×
	40							✓
	50							×
	15	Pre-treatment means < 100epg	×	×	×	×	✓	×
	20		✓				×	✓
	30		×				×	✓
	40		×				×	✓
	50		×				×	✓
15 and 30epg McMaster Data	15	Pre-treatment means >150epg	×	×	×	×	✓	×
	20							
	30							
	40							
	50							
	15	100epg ≤ Pre-treatment means ≤ 150epg	×	×	×	×	✓	×
	20							
	30							
	40							
	50							
	15	Pre-treatment means < 100epg	×	×	×	✓	×	×
	20							
	30							
	40							
	50							

Table 5.15: A summary of the most accurate percentage estimates (inclusive of those indistinguishable from % estimates with the lowest RMSE values) for a given scenario

Given the recommendations above, we will now consider using these to analyse various cattle FEC data available, as part of this project, to investigate the apparent anthelmintic efficacy/resistance status of cattle herds in the UK.

Chapter 6

Applying Bayesian robust methodologies to conclude on apparent anthelmintic efficacy and resistance in UK cattle populations

6.1 Introduction

In Chapter 5, we were able to identify robust percentage estimates (i.e. most accurate) and, by extension, associated experimental designs utilising our developed Bayesian methodologies. We shall now apply some of our developed robust Bayesian methodologies identified in Chapter 5, to conclude on the apparent anthelmintic efficacy/resistance status of cattle herds in the UK, using the available data provided by *Westpoint Farm Vets* (WFV) and the *Veterinary Medicines Directorate* (VMD). We shall also investigate the agreement between these classifications and those which were evaluated using the methodology adopted in the Defra project (Defra 2015; Defra 2018b).

6.2 Data and methodologies used

It was decided by the project team that 15 epg (i.e. 15EPG_McM) data and the hybrid sets of data (i.e. 15EPG_McM_SCFT), which involved counts obtained using the SCFT with a 1 epg sensitivity, as describe in Section 1.6, be considered for analysis.

Treatment groups with pre-treatment means less than 100 epg were not to be further considered as part of this analysis. This is because it is considered unlikely that egg counts less than 100 epg would be associated with sufficient worm burdens and/or pasture contamination to justify anthelmintic use (and associated efficacy testing). Anthelmintic use in such cases would not be consistent with best practice recommendations (Coles *et al.* 1992; Coles *et al.* 2006).

It is worth noting that as part of the Defra project (Defra 2015; Defra 2018b), composite screening FECs were performed, and only those herds with a composite count greater than 150 epg were enrolled onto a subsequent FECRT. However, despite these composite screening attempts, repeated individual baseline faecal egg sampling on some of these farms subsequently revealed mean egg counts less than 100 epg at the baseline sampling point before treatment. This highlights the variability in results that can be obtained between composite and individual samples etc. Indeed, various research groups have investigated the most appropriate methodology for conducting composite samples (Morgan *et al.* 2005; Calvete and Uriarte 2013; George *et al.* 2017). As a result, 53 positive treatment groups were considered for analysis.

For the 15EPG_McM_SCFT data, the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ was evaluated using $NBII(\mu,\sigma)$ likelihoods for the Day 0 and Day 14 data, and for the 15EPG_McM data the average of the individual-based egg count percentage reductions/changes of the SPC form was evaluated utilising a $\mathcal{N}_{trunc}(\mu,\sigma)$ likelihood, along with 95% credible intervals using our developed Bayesian methodologies. These percentage estimates were utilised on the basis that these were the most accurately estimated for data collected using these respective diagnostic sensitivities and for pre-treatment group means greater than or equal to 100 epg, irrespective of treatment group sample size, based on work and findings presented

in Chapter 5.

Priors were specified for each parameter involved from the likelihoods for a specific type of data, and are highlighted below:

Prior Distributions for parameters:

$$\textit{Day 0 Positive Treatment Group Data}:$$
$$\mu \sim Unif(1\text{x}10^{-16}, 1000) \textit{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

$$\textit{Day 14 Positive Treatment Group Data}:$$
$$\mu \sim Unif(1\text{x}10^{-16}, 300) \textit{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

$$100\left(\frac{T_{0,j} - T_{14,j}}{T_{0,j} + T_{14,j}}\right)\% \textit{ Data}:$$
$$\mu \sim Unif(-100, 100) \textit{ and } \sigma \sim Unif(1\text{x}10^{-16}, 500).$$

To obtain posterior distributions for parameters of the likelihoods considered, Metropolis-Hastings (MH) algorithms (further information on MH algorithms is available in Appendix B.1) were developed using *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.) and implemented based on a total of 10,000 iterations. An initial discard of 5,000 iterations was considered (i.e. the burnin period). Normal proposal distributions centred around current parameter values of the chains were used, with a proposal variance of one. Each parameter considered was updated in turn. To ensure each of the parameters' chains were tuned and explored their respective posterior distributions efficiently (i.e. achieved acceptance rates between 20-40% after the initial burnin period) the MH algorithms were programmed to be adaptive, such that at every $200th$ iteration in the burnin period: if the acceptance rate of the chain was above 40% then the variance of the proposal distribution was doubled and if the acceptance rate of the chain was below 20% then the variance of the proposal distribution was

halved. For constructing the percentage estimates mentioned above, the remaining 5,000 iterations associated with the location parameter μ were thinned-out so every $5th$ estimate contributed towards the sample from the posterior distribution. The resulting sample for each location parameter considered, consisted of 1000 sampled estimates from the posterior distributions. The relevant percentage estimates were then constructed; resulting in a sampling distribution of 1000 estimates for each relevant percentage estimate.

From the sampling distributions for each of the percentage estimates considered, the medians of these distributions were obtained as a point estimate for each percentage estimate considered along with 95% credible intervals, i.e. Highest Posterior Density intervals, being obtained for each percentage estimate.

In order to classify the apparent efficacy status of treatment groups as part of this analysis, it was decided that published guidelines' criteria (Coles *et al.* 1992; Coles *et al.* 2006) be adapted in light of the European Medicines Agency (EMA) regarding the FECRT as an estimation of efficacy, and not confirmation of resistance (EMA 2014) and in the interest of being consistent with work that was carried out as part of the Defra project (Defra 2015; Defra 2018b). A treatment was classed as being apparently *efficacious* (Eff) if the relevant percentage estimate was greater than or equal to 95% and the lower limit of the credible interval was greater than or equal to 90%. If only one of these criteria were satisfied, then treatment groups were classified as having *suspected lack of efficacy* (SLOE). If both criteria were not satisfied, i.e. the relevant percentage estimate was less than 95% and the lower limit of the credible interval being less than 90%, then the FECRT result was classified as *lack of efficacy* (LOE).

As part of the Defra study, for 15EPG_McM_SCFT and 15EPG_McM data, percentage estimates

$$100\left(1-\frac{T_{14}}{C_{14}}\right)\%$$

and

$$100\left(1-\frac{T_{14}}{T_0}\right)\%$$

(i.e. percentage estimates (1.2) and (1.4)), were evaluated, respectively.

Confidence intervals (based on normality) were also evaluated using the following respective formulae:

$$100\left(1-\frac{T_{14}}{C_{14}}\exp\left(\pm t_{(n_{treat}+n_{control}-2)}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T^2_{14}}+\frac{s^2_{c.eos}}{n_{control}C^2_{14}}}\right)\right)\%$$

and

$$100\left(1-\frac{T_{14}}{T_0}\exp\left(\pm t_{(n_{treat}-1)}\sqrt{\frac{s^2_{t.eos}}{n_{treat}T^2_{14}}+\frac{s^2_{t.base}}{n_{treat}T^2_{0}}}\right)\right)\%.$$

These percentage estimates and confidence intervals (based on normality) were considered and classified utilising the adapted criteria for classifying apparent efficacy outlined above. With these classifications, based on methods carried out as part of the Defra project, measures of agreement between these sets of classifications and those using our developed Bayesian methodologies were evaluated, for each of the 15EPG_McM_SCFT and 15EPG_McM data. The measures of agreement considered as part of this analysis were the exact agreement, the kappa (κ) and weighted kappa (κ_w) statistics (refer to Appendix B.4 ofr further information on these measures of agreement). The following general guideline for interpreting the κ statistic, provided by Landis and Koch (1977), was adopted for interpretation purposes and is shown below in Table B.2.

It is worth noting that due to a statistical review made as part of the study, negative control groups did not feature as part of the field studies in 2014. With

Value of κ	Strength of Agreement
$<$0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very Good

Table 6.1: Interpreting agreement using κ

this in mind, out of the 53 available positive treatment groups, 23 of these featured in parallel group designs and hence, available to be used to evaluate the percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)$ % as part of the analysis described in this Section. All statistical analyses were carried out using *RStudio* software (version 0.98.994 along with *R* software version 3.1.1.). All visual representation of results, i.e. bar charts, were produced using *Microsoft Excel 2016* software.

6.3 Results

6.3.1 Classification of treatment groups using developed Bayesian methodologies

The results from applying our developed Bayesian methodologies to the 53 available positive treatment group FECRT results from 2012, 2013 and 2014, based on utilising the 15EPG_McM_SCFT data, are presented in Figures 6.1, 6.2 and 6.3, respectively. Included in all Figures are the 95% lower and upper credible intervals represented by the black error bars. The horizontal blue solid line represents the 95% threshold used to classify apparent anthelmintic efficacy status' and the horizontal red dashed line represents the 90% threshold set for the 95% lower limit of the credible interval. Positive treatment groups that received an

injectable or pour-on formulation of doramectin, a fenbendazole treatment, or an injectable or pour-on formulation of ivermectin are highlighted as purple, orange, red, blue and green, in these charts, respectively. Table 6.2 also highlights the breakdown of the classifications with respect to each type of positive treatment group featured.

Based on these results, the majority of treatment groups were classified as having an apparent LOE status. In fact, from Table 6.2 we observe that approximately 73.58%, of the 53 positive treatment groups considered, were classified as having an apparent LOE status.

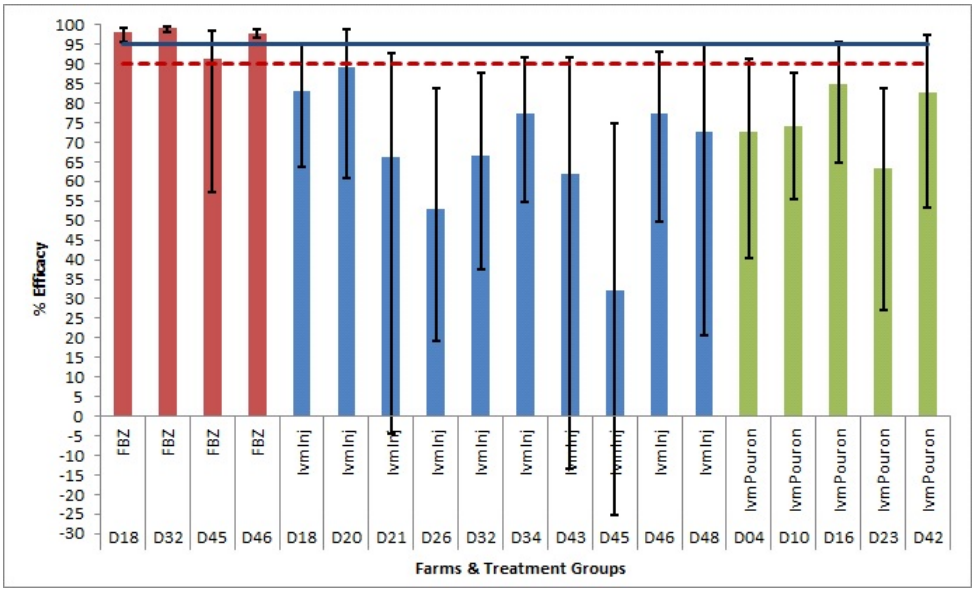

Figure 6.1: 2012 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

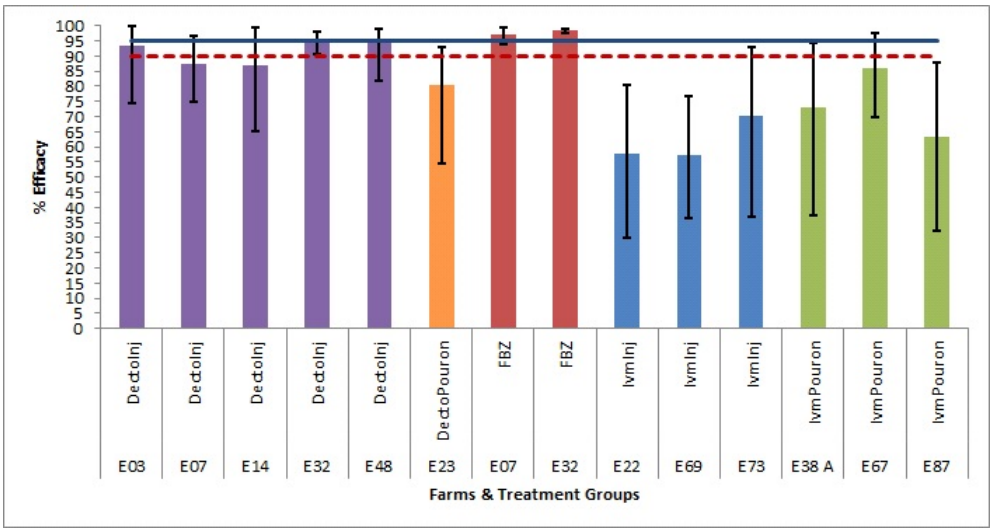

Figure 6.2: 2013 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

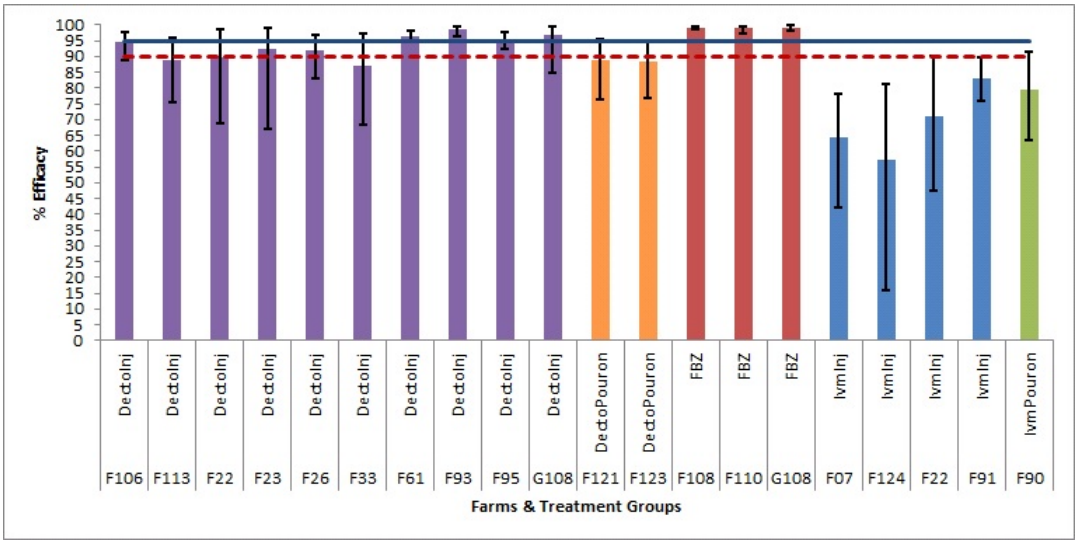

Figure 6.3: 2014 FECRT results, based on 15EPG_McM_SCFT data, utilising Bayesian methodologies with percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Efficacy Status			
Treatment Groups	Eff	SLOE	LOE	Total
DectoInj	4	2	9	15
DectoPouron	0	0	3	3
FBZ	8	0	1	9
IvmInj	0	0	17	17
IvmPouron	0	0	9	9
Total	12	2	39	53

Table 6.2: Classifications of treatment groups, based on 15EPG_McM_SCFT data and utilising Bayesian methodologies with percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

When considering the 15EPG_McM data, the results from applying our developed Bayesian methodologies to the 53 available positive treatment group FECRT results from 2012, 2013 and 2014 for these data, are presented in Figures 6.4, 6.5 and 6.6, respectively. Table 6.3 also highlights the breakdown of these classifications with respect to each type of positive treatment group featured. The majority of treatment groups were classified as having an apparent efficacious or SLOE status. In fact, from Table 6.3, we observe that approximately 47.17% of the 53 positive treatment groups considered, were classified as having an apparent efficacious status and 26.42% were classified as having an apparent SLOE status. Overall, 73.59% of treatment groups were classed as having an apparent efficacious or suspected lack of efficacy status.

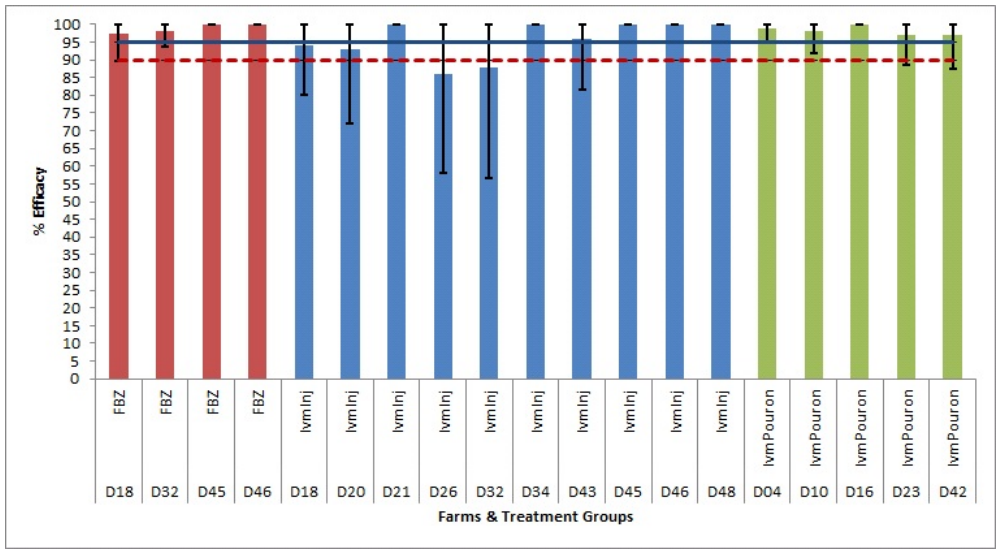

Figure 6.4: 2012 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

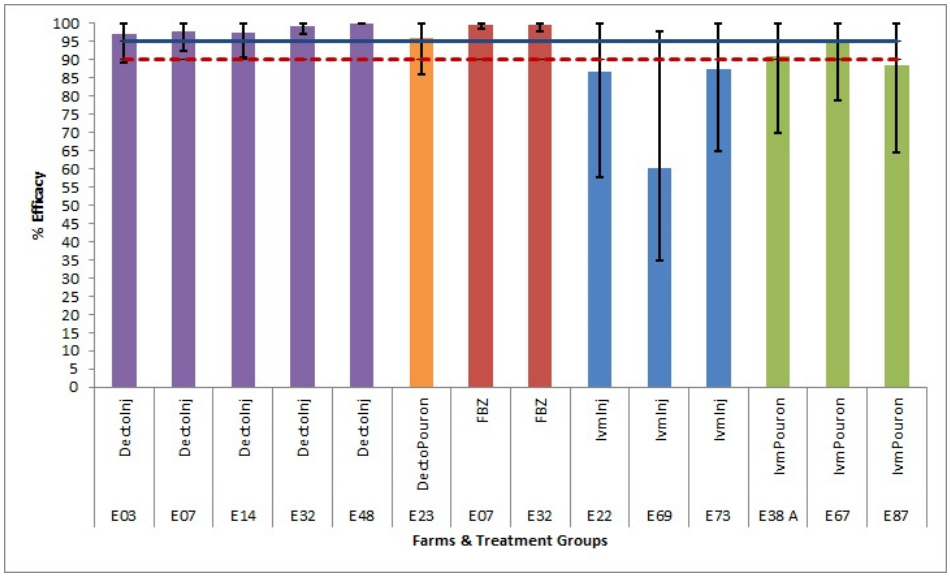

Figure 6.5: 2013 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

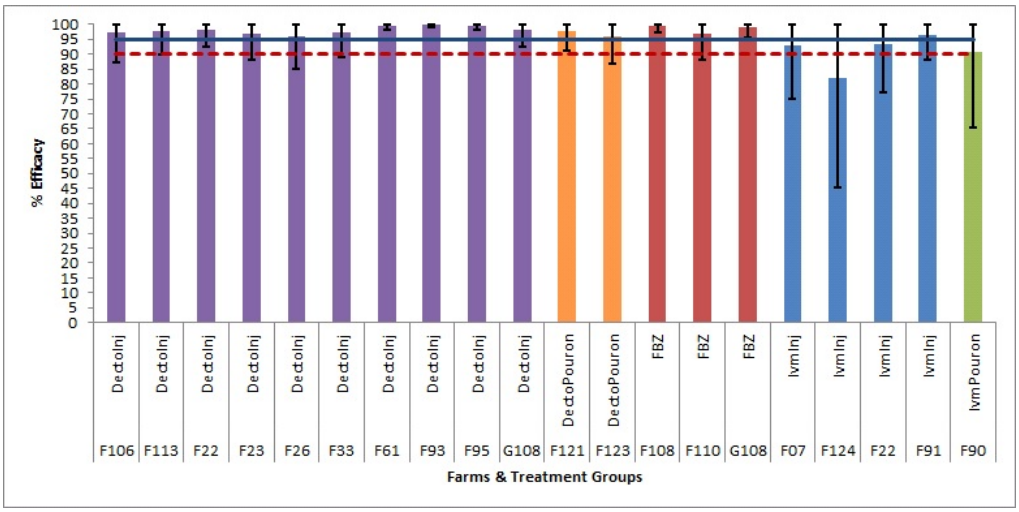

Figure 6.6: 2014 FECRT results, based on 15EPG_McM data, utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Efficacy Status			
Treatment Groups	Eff	SLOE	LOE	Total
DectoInj	9	6	0	15
DectoPouron	1	2	0	3
FBZ	7	2	0	9
IvmInj	5	2	10	17
IvmPouron	3	2	4	9
Total	25	14	14	53

Table 6.3: Classifications of treatment groups based on 15EPG_McM data and utilising Bayesian methodologies with percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

6.3.2 Agreement between classifications using developed Bayesian methodologies and Defra study methodologies

The results from the Defra project involving the application of the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, with its associated 95% confidence interval based on normality, to the 53 available positive treatment group FECRT results from 2012, 2013 and 2014, utilising the 15EPG_McM_SCFT data, are presented in Figures 6.7, 6.8 and 6.9, respectively. Included in all Figures are the 95% lower and upper credible intervals represented by the black error bars. The horizontal blue solid line represents the 95% threshold used to classify apparent anthelmintic efficacy status' and the horizontal red dashed line represents the 90% threshold set for the 95% lower limit of the credible interval. Positive treatment groups that received an injectable or pour-on formulation of doramectin, a fenbendazole treatment,

or an injectable or pour-on formulation of ivermectin are highlighted as purple, orange, red, blue and green, in these charts, respectively.

Utilising the classification criteria in Section 6.2 for these results, along with those classifications concluded from the results based on our developed Bayesian methodologies presented in Figures 6.1 - 6.3, a 3x3 contingency table of the classifications and agreement between the two methodologies considered, is presented in Table 6.4. The exact agreement, i.e. the sum of the diagonal entries divided by the total number of observations considered, is 83.02%. From Table 6.4, the κ statistic was estimated to be $\hat{\kappa} = 0.63$ along with a 95% confidence interval (0.44, 0.83). The weighted κ statistic, i.e. κ_w, based on the results presented in Table 6.4, was estimated as $\hat{\kappa_w} = 0.74$ with a 95% confidence interval (0.58, 0.91).

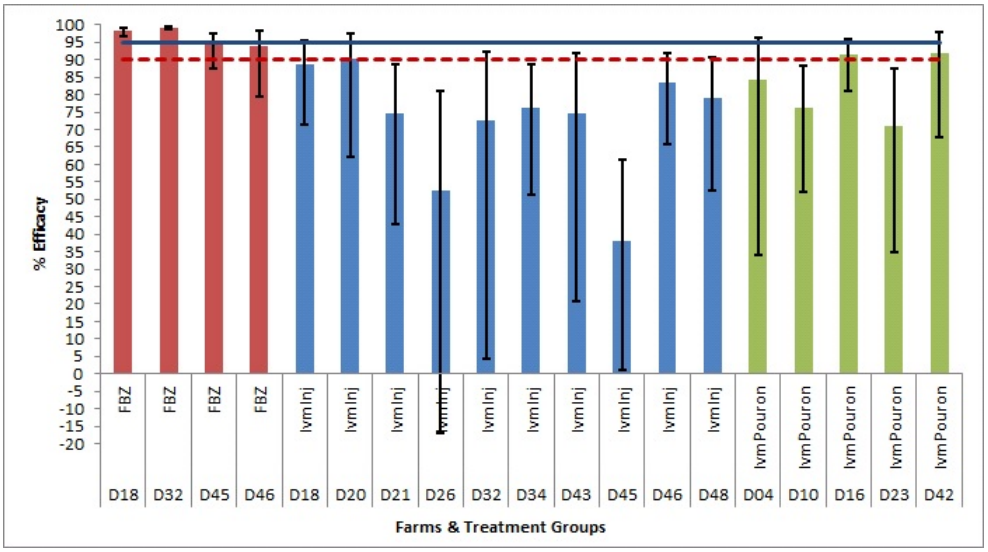

Figure 6.7: 2012 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)$ % and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

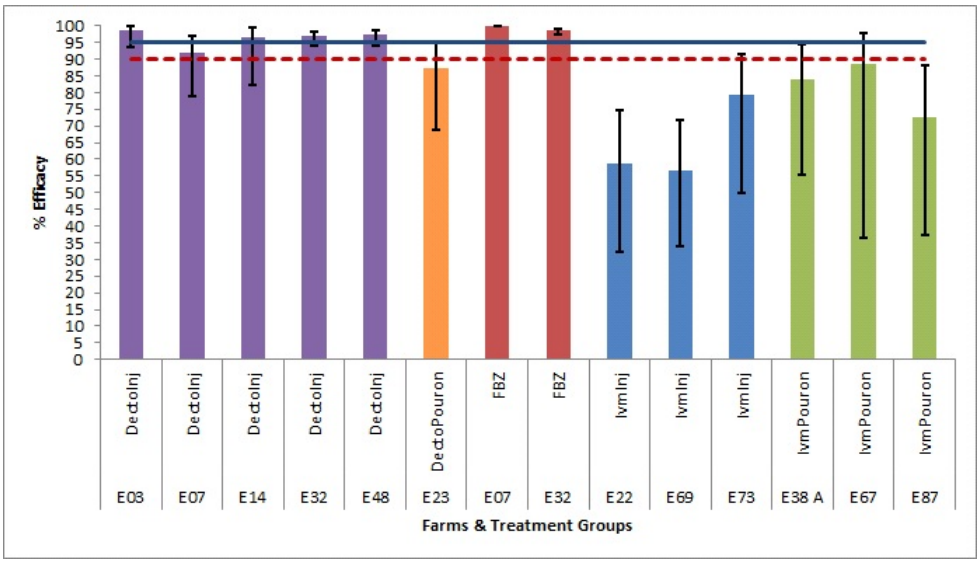

Figure 6.8: 2013 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)$ % and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

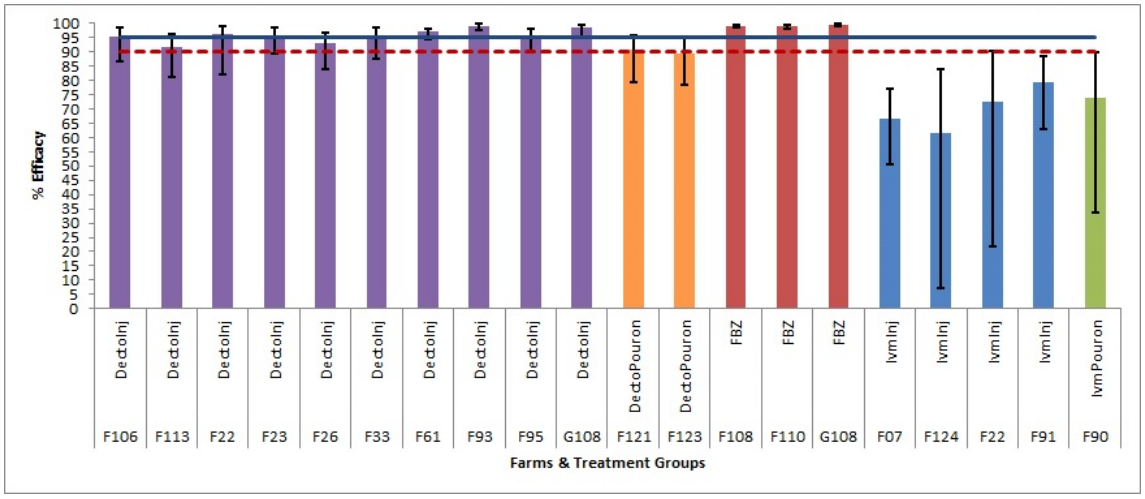

Figure 6.9: 2014 FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)$ % and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Classification using Bayesian method with $100\left(1-\frac{T_{14}}{T_0}\right)\%$			
Classification based on Defra project methods with $100\left(1-\frac{T_{14}}{T_0}\right)\%$	Eff	SLOE	LOE	Total
Eff	11	2	1	14
SLOE	0	0	5	5
LOE	1	0	33	34
Total	12	2	39	53

Table 6.4: 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, based on 15EPG_McM_SCFT data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

Based on utilising the 15EPG_McM_SCFT data, the 23 available positive treatment group FECRT results from 2012 and 2013, involving the percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ along with the associated 95% confidence interval based on normality (from the Defra project) are presented in Figure 6.10. It is worth noting that some of the confidence limits were truncated in this Figure for presentation purposes.

Using the adapted classification criteria in Section 6.2 for these results, along with those classifications concluded from our developed Bayesian methodologies for the relevant 23 positive treatment groups (which can be found in Figures 6.1 and 6.2), a 3x3 contingency table of the classifications and agreement between the classifications from the two methodologies considered, is presented in Table 6.5. The exact agreement is 78.26%, $\hat{\kappa} = 0.51$ with a 95% confidence interval (0.18, 0.85) and $\hat{\kappa_w} = 0.57$ with a 95% confidence interval (0.23, 0.91).

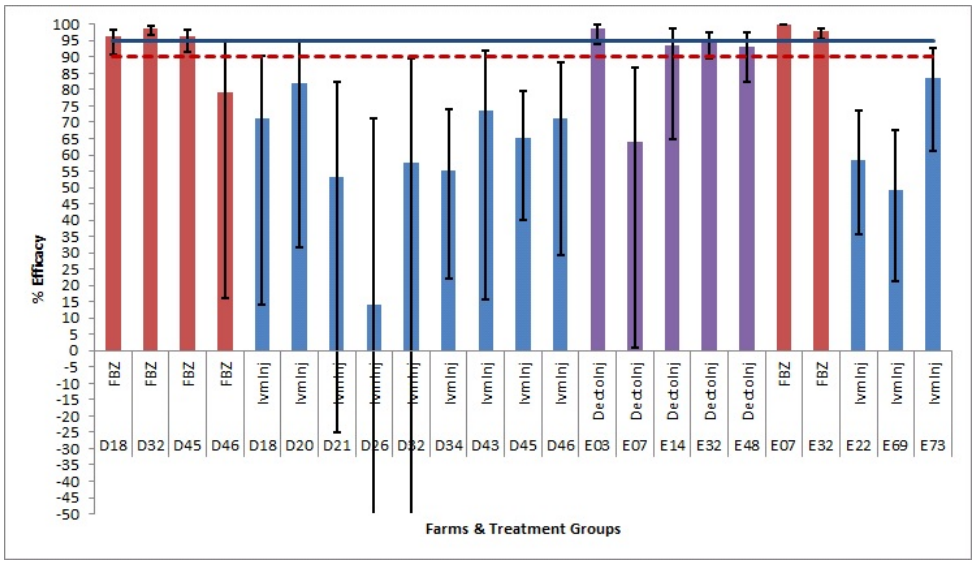

Figure 6.10: FECRT results, based on 15EPG_McM_SCFT data, using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Classification using Bayesian method with $100\left(1-\frac{T_{14}}{T_0}\right)\%$			
Classification based on Defra project methods with $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$	Eff	SLOE	LOE	Total
Eff	4	0	2	6
SLOE	1	0	0	1
LOE	1	1	14	16
Total	6	1	16	23

Table 6.5: 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, based on 15EPG_McM_SCFT data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

From the Defra project, the FECRT results of the 53 available positive treatment groups from 2012, 2013 and 2014, based on utilising the 15EPG_McM data and applying the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ along with its associated 95% confidence interval, based on normality, are presented in Figures 6.11, 6.12 and 6.13, respectively. Again, some of the confidence limits were truncated in these Figures for presentation purposes.

Utilising the classification criteria in Section 6.2 for these results, along with those classifications concluded from our developed Bayesian methodologies presented in Figures 6.4 - 6.6, a 3x3 contingency table of the classifications and agreement between the classifications from the two methodologies considered, is presented in Table 6.6. The exact agreement is 60.38%, $\hat{\kappa} = 0.39$ along with a 95% confidence interval (0.23, 0.56) and $\hat{\kappa_w} = 0.50$ with a 95% confidence interval (0.33, 0.67).

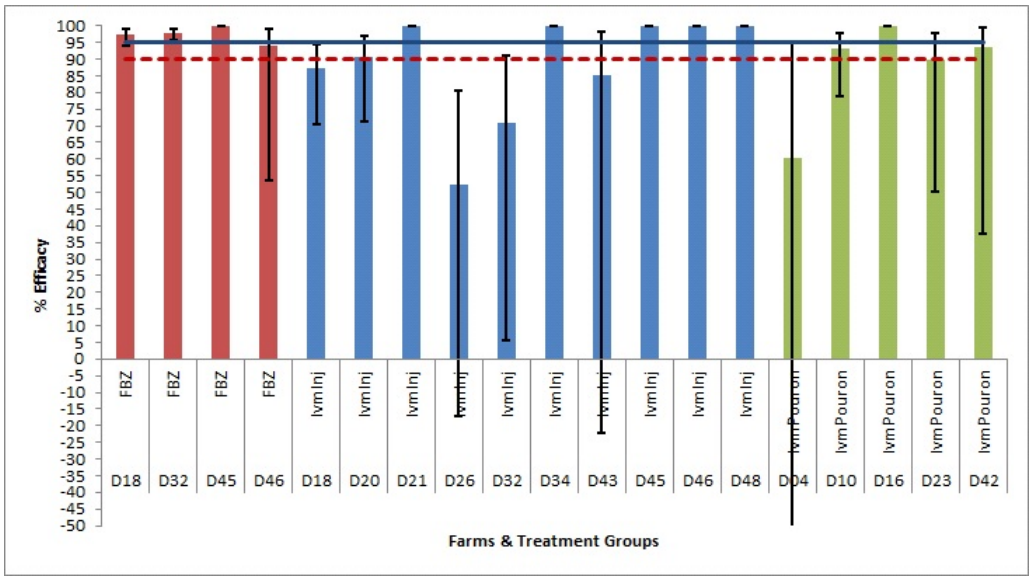

Figure 6.11: 2012 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

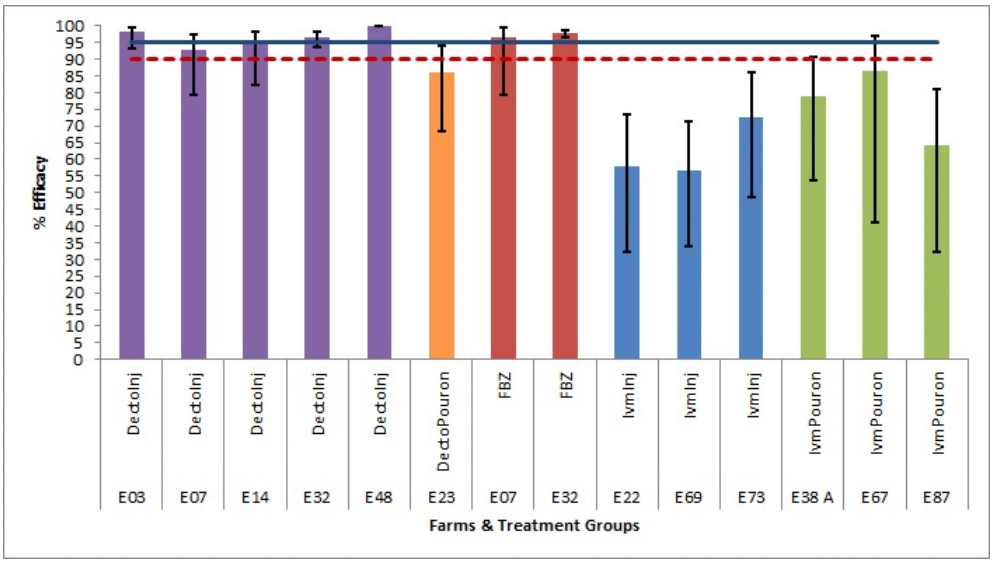

Figure 6.12: 2013 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)$ % and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

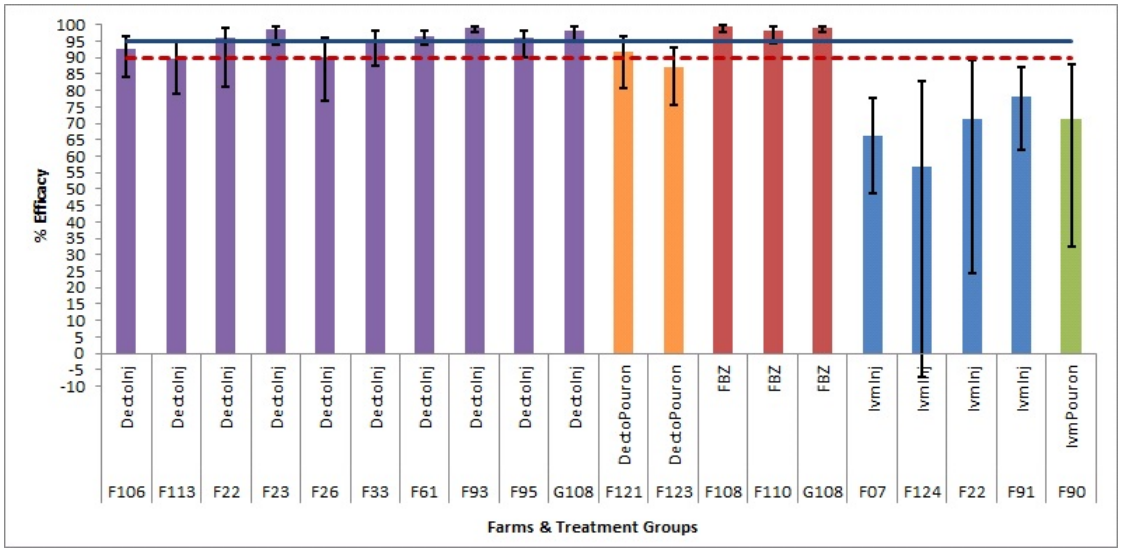

Figure 6.13: 2014 FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)$ % and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Classification using Bayesian method with $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$			
Classification based on Defra project methods with $100\left(1-\frac{T_{14}}{T_0}\right)\%$	Eff	SLOE	LOE	Total
Eff	17	4	0	21
SLOE	2	1	0	3
LOE	6	9	14	29
Total	25	14	14	53

Table 6.6: 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, based on 15EPG_McM data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

The 23 available positive treatment group FECRT results from 2012 and 2013, based on utilising the 15EPG_McM data involving the percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and its associated 95% confidence interval based on normality (from the Defra project) are presented in Figure 6.14. Again, some of the confidence limits were truncated in this Figure for presentation purposes.

Using the adapted classification criteria in Section 6.2 for these results, along with those classifications concluded for the 23 available positive treatment groups using our developed Bayesian methodologies (which can be found in Figures 6.4 and 6.5), a 3x3 contingency table of the classifications and agreement between the classifications from the two methodologies, is presented in Table 6.7. The exact agreement is 60.87%, $\hat{\kappa} = 0.36$ with a 95% confidence interval (0.11, 0.62) and $\hat{\kappa_w} = 0.43$ with a 95% confidence interval (0.15, 0.70).

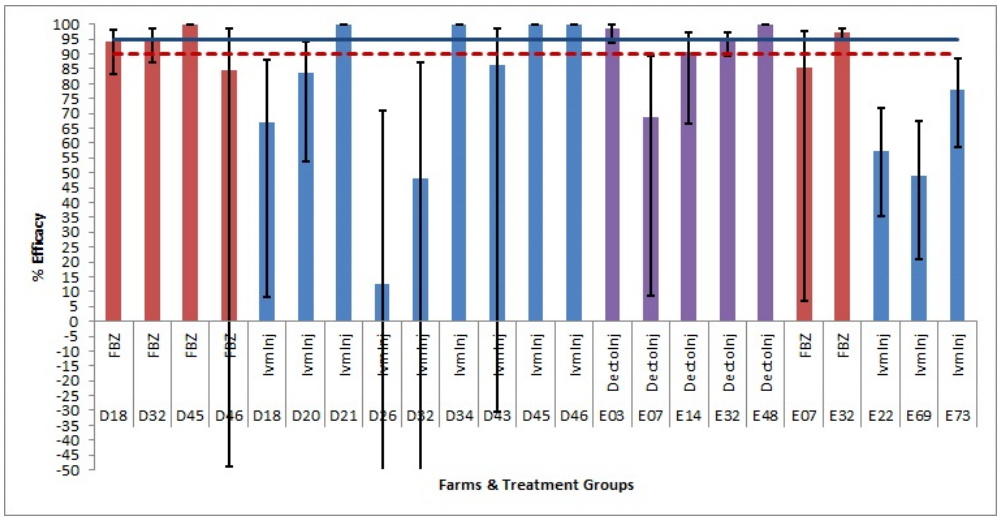

Figure 6.14: FECRT results, based on 15EPG_McM data, using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$ and confidence intervals based on normality, featuring as part of Defra project (Treatments: DectoInj (purple); DectoPouron (orange): FBZ (red); IvmInj (blue) and IvmPouron (green))

	Classification using Bayesian method with $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$			
Efficacy Status based on Defra project methods with $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$	Eff	SLOE	LOE	Total
Eff	7	1	0	8
SLOE	1	0	0	1
LOE	5	2	7	14
Total	13	3	7	23

Table 6.7: 3x3 contingency table of classifications using developed Bayesian methodologies using percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$ and Defra project methods using percentage estimate $100\left(1-\frac{T_{14}}{C_{14}}\right)\%$, based on 15EPG_McM data (Classifications: Efficacious (Eff); Suspected lack of efficacy (SLOE) and Lack of efficacy (LOE))

6.4 Discussion

Based on our developed Bayesian methodologies, for the FECRT results using the 15EPG_McM_SCFT data, we observed that the majority of treatment groups were classified as having an apparent LOE status (39 out of 53 positive treatment groups), whereas based on the 15EPG_McM being utilised, we observed that the majority of treatment groups were classified as being either apparently efficacious or having an apparent SLOE status (25 out of 53 positive treatment groups classed as efficacious and 14 out of 53 positive treatment groups classed as having an apparent suspected lack of efficacy status). As a result, the number of treatment groups classed as having a LOE status increased upon improving the diagnostic sensitivity of the counting techniques that were utilised. This is intuitive, since as part of the FECRT studies conducted (Defra 2015; Defra 2018b), any individual faecal samples that were detected below 120 epg (in 2012) or 60 epg (in 2013 and 2014) were re-analysed with a technique that had a lower limit of detection, i.e. 1 epg using the SCFT, and so it is more likely that these individual egg counts that featured as part of the 15EPG_McM_SCFT data, replaced *false/excess* zeros that may have been present in those individual samples as part of the 15EPG_McM data.

When considering the different positive treatment groups, we observed based on the 15EPG_McM_SCFT data, all of the positive treatment groups who received ivermectin (irrespective of the treatment being a pour-on or injectable formulation) were classed as exhibiting an apparent LOE, and the majority of these treatment groups featured as part of FECRT experiments conducted in 2012.

Based on the 15EPG_McM data, the majority of those who received the ivermectin injectable formulation were classed as having an apparent LOE status as well (10 out of 17 ivermectin injectable groups). In fact, those ivermectin injectable treatment groups that featured in FECRT studies conducted in 2013 were all classified as this. It is worth noting that all the apparent efficacious classifications of the ivermectin injectable treatment groups, based on 15EPG_McM data, featured in FECRT studies that were conducted in 2012. With regards to the pour-on formulation variation of ivermectin and the 15EPG_McM data, the majority of these groups exhibited either an apparently efficacious or SLOE

status (5 out of 9 treatment groups). One can also observe that all of those ivermectin pour-on treatment groups who were classed as having an apparent LOE status, featured in FECRT studies conducted in 2013 and 2014 and all of the apparent efficacious classifications associated with ivermectin pour-on treatment groups were part of FECRT experiments conducted in 2012.

For those treatment groups that received doramectin, based on 15EPG_McM_SCFT data being utilised, all three groups that received a pour-on formulation and the majority of those who received the injectable formulation (9 out of 15 doramectin injectable treatment groups) were classed as having an apparent LOE status. Out of those treatment groups that received some formulation of doramectin, two-thirds of these treatment groups (12 out of 18 in total) were classed as having an apparent LOE and the majority of these treatment groups featured as part of FECRTs conducted in the year 2014.

When considering the 15EPG_McM data, the majority of those groups who received an injectable formulation of doramectin were classed as being apparently efficacious (9 out of 15 treatment groups) with the others being classed as having a suspected lack of efficacy status; however two of the three groups who received a pour-on formulation of doramectin were classed as having an SLOE status, with the other one treatment group being classed as apparently efficacious. Overall though, the majority of those treatment groups who had received doramectin were classified as being apparently efficacious (10 out of 18 in total) and most of these treatment groups featured as part of FECRTs conducted in 2014.

As reported in Defra (2015), for the majority of groups where there was an apparent LOE following treatment with a ML anthelmintic and a 1epg sensitivity being used, *C. oncophora* larvae predominated, which is not a surprising result given that *C. oncophora* is the dose-limiting species for the ML group of anthelmintics (Vercruysse and Rew 2002), and these findings are in agreement with other studies (Sargison *et al.* 2009; McArthur *et al.* 2011; O'Shaughnessy *et al.* 2014). Though this is a limitation of the use of FECs as a measured response for investigating the apparent anthelmintic efficacy status of livestock: FECs cannot distinguish between certain species of cattle nematode, whose eggs look very similar morphologically. As a result, it is usually recommended to carry out larval

speciation as an indication of the species of worms present on farm (Taylor 2010a; Defra 2015). Counting parasitic eggs also gives an indirect measure of worm burden present in cattle herds (Eysker and Ploeger 2000) and can present many interpretational issues. For instance, a high parasitic egg count in faeces may be regarded as an indication of high worm burden, but takes no consideration of the fact that species of nematode vary in their fecundity and pathogenicity. Faecal egg production varies throughout the year and is greatly influenced by a number of factors including levels of parasite challenge (which in turn is influenced by seasonal weather patterns) and the development of protective immunity. On the other hand, a low parasitic egg count cannot be associated with a low worm burden since low egg counts do not take account of parasites being immature or in a hypobiotic state. Additionally, the variability in results that can be obtained between composite and individual samples should also be considered. This is because, as part of the Defra project (Defra 2015), composite screening FECs were performed (i.e. obtaining pre-treatment means), and only those herds with a composite count greater than 150 epg were enrolled onto a subsequent FECRT. It is also worth noting that various research groups have investigated the most appropriate methodology for conducting composite samples (Morgan *et al.* 2005; Calvete and Uriarte 2013; George *et al.* 2017). However, despite these composite screening attempts, repeated individual baseline faecal egg sampling on some of these farms subsequently revealed mean egg counts less than 100 epg at the baseline sampling point before treatment. The only way of estimating worm burden to a higher degree of accuracy than what FECs can provide, however, is through a post-mortem examination of cattle, i.e. slaughter trials Powers *et al.* (1982), but these are not practicable in the field.

For treatment groups who received fenbendazole, i.e. a BZ class of anthelmintic, the majority, based on the 15EPG_McM_SCFT data, were classed as being apparently efficacious, i.e. 88.89% of the 9 groups who received this treatment. Based on the 15EPG_McM data, the majority of these positive treatment groups were also classed as being apparently efficacious, though this majority decreased to 77.78%. As mentioned in Defra (2015), the performance observed using FBZ is surprising given the almost ubiquitous resistance in equine small strongyles to FBZ (Lester and Matthews 2013; Stratford *et al.* 2014) and high prevalence

of FBZ resistance in sheep (Bartley *et al.* 2004; Sargison *et al.* 2005). This could reflect limited use of this anthelmintic in cattle compared to other livestock species. The consistency of classifications between the diagnostic sensitivities (and by extension, the developed Bayesian methodologies) used and performance of FBZ is encouraging as it could be used at pasture to reduce ML anthelmintic use, which are regarded as the most popular among the classes of anthelmintics currently used (Vercruysse and Rew 2002; Omura 2008; Taylor 2010a; Defra 2015), and allow targeted ML anthelmintic use.

When considering the measures of agreement between the sets of classifications obtained using our developed Bayesian methodologies and Defra project methods: the exact agreements were 83.02%, 78.26%, 60.38% and 60.87% based on the results from Tables 6.4, 6.5, 6.6 and 6.7, respectively. As a result, we observe that the majority of the classifications, between any two competing methodologies carried out as part of this analysis, were in agreement. It is worth noting that the exact agreements were higher when utilising the 15EPG_McM_SCFT data in comparison to those exact agreements that were evaluated when the 15EPG_McM data had been utilised. One possible reason for this could be that, as part of our developed Bayesian methodologies for analysing the 15EPG_McM_SCFT data: the percentage estimates being utilised were of a similar form, i.e. involved arithmetic mean group estimates, to those that were involved with the Defra project methodologies that were utilised.

Calculating the exact agreement, however, takes no account of where in the contingency Tables 6.4-6.7 the agreement is and some agreement between any two competing methodologies being used would be expected to occur by chance. One could vote to evaluate the κ statistic as a measurement of agreement, and this statistic along with associated 95% confidence interval has been evaluated for completeness for each set of results. However a weakness of this statistic is that it takes no account of the degree of disagreement present - all disagreements are treated equally. Where the categories are ordered, as is in our case, it is preferable to give different *weights* to disagreements according to the magnitude of the discrepancy. Hence, the weighted κ statistic, κ_w, and associated 95% confidence intervals were evaluated for each set of tables and will be discussed here.

For the classifications presented in Table 6.4 based on 15EPG_McM_SCFT data, the weighted κ statistic, κ_w was estimated as $\hat{\kappa_w} = 0.74$. According to Landis and Koch (1977), this would mean that there is *good* agreement between the two methods used to obtain classifications. An accompanying 95% confidence interval for this estimate was obtained: (0.58, 0.91). As a result, with respect to the two competing methodologies being utilised in obtaining the relevant classifications, it would be highly likely, i.e. 95% of the time, the κ_w statistic lies between 0.58 (i.e. *moderate* agreement) and 0.91 (i.e. *very good* agreement) if we were to repeatedly sample from the same population/carry out the experiment many times. These results are encouraging to observe since, even by accounting for the degree of disagreement between the classifications obtained by the two competing methodologies considered, we are able to obtain good agreement between classifications. As well as this, the developed Bayesian methodology used in this analysis was based on using the percentage estimate $100\left(1 - \frac{T_{14}}{T_0}\right)\%$, which is obtained by utilising a paired study design with a positive treatment group only and is a percentage estimate which has grown in popularity for these types of studies amongst different livestock species (Cabaret and Berrag 2004; Lyndal-Murphy *et al.* 2010; Levecke *et al.* 2012; Lester *et al.* 2013; Lyndal-Murphy *et al.* 2014; Stratford *et al.* 2014; George *et al.* 2017).

Again, based on 15EPG_McM_SCFT data being utilised, for the classifications presented in Table 6.5, κ_w was estimated as $\hat{\kappa_w} = 0.57$. With respect to Landis and Koch (1977), this would mean that there is *moderate* agreement between the two methods used to obtain classifications. The associated 95% confidence interval for this estimate was (0.23, 0.91), meaning that 95% of the time, we would expect the κ_w statistic to lie between 0.23 (i.e. *fair* agreement) and 0.91 (i.e. *very good* agreement) if we were to repeatedly sample from the same population/carry out the experiment many times. It is worth noting that the 95% confidence interval here is wider in comparison to the interval mentioned above, though this is probably due to a smaller group of observations being considered, i.e. 23 positive treatment groups being utilised as opposed to 53 available positive treatment groups. Nevertheless, it is encouraging to see that the agreement between the two competing methodologies considered here have moderate agreement between the classifications obtained, in light of the degree of disagreement being accounted

for.

For the classifications presented in Table 6.6 based on 15EPG_McM data being considered, the weighted κ statistic, κ_w was estimated as $\hat{\kappa_w} = 0.50$. According to Landis and Koch (1977) there is then *moderate* agreement between the two methods used to obtain classifications. An accompanying 95% confidence interval for this estimate was obtained: (0.33, 0.67). As a result, we conclude that it would be highly likely, i.e. 95% of the time, that the κ_w statistic lies between 0.33 (i.e. *fair* agreement) and 0.67 (i.e. *good* agreement) if we were to repeatedly sample from the same population/carry out the experiment many times. For the classifications presented in Table 6.7, κ_w was estimated as $\hat{\kappa_w} = 0.43$. According to Landis and Koch (1977), this would mean that there is *moderate* agreement between the two methods used to obtain classifications. The 95% confidence interval associated with this estimate was (0.15, 0.70), meaning that it would be highly likely, i.e. 95% of the time, that the κ_w statistic in this instance lies between 0.15 (i.e. *poor* agreement) and 0.70 (i.e. *good* agreement) if we were to repeatedly sample from the same population/carry out the experiment many times.

One may notice that the estimates for the κ_w statistic are lower (i.e. more disagreements between the two sets of classifications being present), in the instances where our developed Bayesian methodology has been utilised for the 15EPG_McM data. One possible reason for this is that which was explained for the exact agreement results: the Defra project methodologies depend on arithmetic group mean estimates being utilised, which feature as part of the developed Bayesian methodology for the 15EPG_McM_SCFT data, as opposed to averaging over individual-based egg count percentage reductions/changes of the SPC form, which featured in the analysis for the 15EPG_McM data. Nevertheless, it is encouraging to observe an estimated moderate agreement between the classifications obtained from the developed Bayesian methodology and the Defra project methodologies, applied to the 15EPG_McM data.

One may ask themselves and argue however, *"Why would it be worth going to all the extra effort of carrying out this recommended Bayesian method, when there is a significant moderate agreement between the classifications it obtains with those*

obtained using currently recommended frequentist approaches?" In response to this, the fact that our novel Bayesian method (involving individual-based egg count percentage reductions/changes of animals of the SPC form) is able to obtain a significant moderate agreement in the classifications it obtains compared with classifications from methods currently utilised in anthelmintic studies is encouraging to observe and can be regarded as a positive result. Our novel Bayesian method is statistically sound as well, on the grounds that one does not need to assume normality of FEC data - which is an assumption that has been observed to be valid when dealing with cattle FEC data - in order to produce interval estimates, such as those provided by WAAVP guidelines and other communications (Coles *et al.* 1992; Coles *et al.* 2006; Lyndal-Murphy *et al.* 2014).

Overall conclusions:

The discussion above highlights the importance of the choice of diagnostic sensitivity for the counting techniques utilised for these types of experiments and the interpretations that can be concluded from them. Though there was at least an estimated significant moderate agreement between all sets of comparisons of classifications produced by our developed Bayesian and Defra project methods, for both types of diagnostic sensitivities considered (i.e. 1 epg and 15 epg counts).

The majority of treatment groups were classified as having an apparent LOE status when 1 epg data were utilised, whereas based on FECs that had been collected with a 15 epg sensitivity, we observed that the majority of treatment groups were classified as being either apparently efficacious or having an apparent SLOE status. However, work presented and conclusions by EMA (2014) and the Defra project (Defra 2015) tell us that an apparent lack of efficacy status does not necessarily indicate resistance. In fact, based on clinical grounds, the results presented in this Chapter, the results presented in Defra (2015) and by, what was considered from the project team, to be a better interpretation/representation of what occurred in the field studies as part of the Defra project, it would be recommended to:

- utilise a paired study design using a positive treatment group only;

- determine FECs using a 15 epg sensitivity method;
- conduct with pre-treatment group means/composite samples greater than 150 epg;
- use a minimum treatment group sample size of 15 animals;
- and adopt the approach of obtaining the mean value of individual-based egg count percentage reductions/changes of animals using the *Symmetrised Percentage Change* form (along with its associated 95% credible interval) with our developed Bayesian methodology,

in order to improve the FECRT. However, with these recommendations one must remember the caveats of utilising egg count data as a means of investigating apparent anthelmintic efficacy (as described above), in order to conclude overall on the apparent efficacy status of cattle herds.

Chapter 7

Discussion, future work and an *R Shiny* prototype webpage application

In this thesis, areas of interest and study with regards to the statistical aspects of the FECRT for livestock, such as the identification of appropriate statistical techniques/frameworks (Coles *et al.* 1992; Coles *et al.* 2006; Denwood 2010; Lester and Matthews 2013) for the analysis of FEC data (Presidente 1985; Dobson *et al.* 2009) and robust experimental study designs (Vidyashankar *et al.* 2007; Vidyashankar *et al.* 2012; Lyndal-Murphy *et al.* 2014), have been examined.

In Chapter 2, a review of the relevant literature was presented, which focused on experimental design considerations, various statistical calculations (i.e. different percentage estimates that can be considered) and the statistical frameworks for which interval estimation can be carried out for the FECRT. With respect to statistical frameworks, derivations of confidence intervals using asymptotic approximations for the most commonly used percentage estimates were presented and Bootstrap and Bayesian methodologies were also discussed here.

From this review, it was found that published guidelines, and other communications, promote the use of percentage estimates based on arithmetic group means and associated 95% confidence intervals, which are derived by assuming FEC

data to be of a normal nature. In doing so, we are able to obtain approximate estimates for the *ln*-transformed ratio of means of FEC data their associated variances, which are then used in classifying the apparent efficacy status of treatment groups (Coles *et al.* 1992; Coles *et al.* 2006; Lyndal-Murphy *et al.* 2014). In Chapter 3, using available field study data, i.e. 30 and 15 epg McMaster data and hybrid sets of data involving counts obtained with a 15 and 1 epg sensitivity, it was found that the majority of FEC data were of a non-normal nature, even upon transformation. As part of this work, three transformations were utilised, i.e. $ln(x+1)$ (where x is defined as a FEC), the square-root and $x^{\frac{2}{3}}$ power transformation. If it had been the case that the majority of Day 0 and Day 14 FEC data had been of a normal nature, in light of a particular transformation being applied, then explicit confidence intervals based on asymptotic approximations could have been derived. There is scope to do this as part of future work by considering other transformations, such as other power transformations (Newton and Rudestam 2012), since only three typical transformations (Zar 1996) were utilised as part of this work.

The derivations of confidence intervals that involved the use of both Day 0 and 14 treatment data, explored in Chapter 2, were also found to be dependent on the correlation between the *ln*-transformed versions of treatment data involved and not the correlation between the original treatment data as some communications promote (Lyndal-Murphy *et al.* 2014). It would not be possible to evaluate these confidence intervals if zero-valued FECs were obtained as part of anthelmintic studies. Indeed, this scenario is likely to occur, since the majority of 15 and 30 epg McMaster data, in Chapter 3, were found to be best represented by zero inflated distributions (ZIDs). As a result, the central tendency μ_1. This is defined as the arithmetic mean of the data divided by the proportion of non-structural zero counts, and is estimated as the arithmetic mean of FECs divided by the proportion of non-zero counts. This parameter would be recommended for use when forming percentage estimates, on the basis that this is the maximum likelihood estimator for these types of distributions, as opposed to the use of arithmetic means when faced with zero inflated data. It is worth noting that μ_1 can take account of non-structural zero counts and higher-valued data points present in zero inflated data, which may be less accounted for when locations such as the

arithmetic mean are used. This is due to zero inflation of potential false zeros decreasing these values to that of a value closer to zero. Therefore, we are essentially trying to correct/compensate for the presence of zero inflation by using μ_1 as an estimate of central tendency/location of the discrete count distribution involved.

For the hybrid sets of data involving counts obtained with a 1 epg sensitivity, distributions associated with and including the Negative Binomial distribution were found to be the best fitting, in Chapter 3. Hence, percentage estimates and confidence limits could be estimated using arithmetic group means (i.e. the central tendency estimates associated with these distributions). These findings are re-assuring, since the Negative Binomial distribution and its associated central tendency is recommended for representing parasitological FEC data (Shaw and Dobson 1995; Morgan *et al.* 2005; Denwood *et al.* 2008; Levecke *et al.* 2012).

Overall, the results obtained in Chapter 3 indicated that the diagnostic sensitivity of the counting techniques used to obtain FEC data influence the probability distributions of best fit. This, by extension, influences the central tendency/location parameter to best represent the FEC data, and to subsequently be used as part of constructing percentage estimates. The use of currently recommended forms of confidence intervals, i.e. those derived on the assumption of FEC data being normal, were also found not to be valid. Therefore, we must look to other statistical frameworks for estimating intervals for percentage estimates used as part of the FECRT.

With the probability theory in Chapter 3 underpinned, a simulation study using field study data was conducted in Chapter 4 using Bootstrap methodology to assess the coverage probability of 95% percentile intervals associated with different percentage estimates. The performance of the intervals was considered under various scenarios involving different diagnostic sensitivities, treatment group sizes and classifications of pre-treatment group means. It was at this stage in the work that percentage estimates such as the *Symmetrised Percentage Change* (SPC) and averaging over individual-based egg count percentage reductions/changes of different forms were considered (Cabaret and Berrag 2004; Berry and Ayers 2006). Despite this, very few scenarios consisted of 95% Bootstrapped percentile inter-

vals with adequate coverage probabilities. For those scenarios in which adequate coverage probabilities were obtained however, data were collected with a 1 epg sensitivity and percentage estimates were based on a paired study design being used with a positive treatment group only. It is worth noting though, that only one type of Bootstrapped confidence interval was utilised, namely the percentile interval. As part of future work, a simulation study examining the coverage probabilities involving the *Bias-corrected and Accelerated* (BCA) Bootstrapped intervals (Efron and Tibshirani 1993) could be developed. However, Carpenter and Bithell (2000) tell us that the calculation of estimating the acceleration parameter is tortuous, and therefore time consuming - a cost which is often taken into account when carrying out these types of studies (Burton *et al.* 2006) - particularly for complex parametric problems, such as in our case in dealing with the ratio of means in forming percentage estimates. BCA intervals also require a large number of Bootstrap replications, in fact Efron and Tibshirani (1993) tell us that at least 1000 are needed in order to reduce sampling error. The coverage probabilities of Bootstrapped percentile intervals though was investigated as part of this work due to the fact that percentile Bootstrapping methods have proved popular in the veterinary and parasitological communities when constructing confidence intervals (Cabaret and Berrag 2004; Vidyashankar *et al.* 2007; Traversa *et al.* 2009; Lester and Matthews 2013; Lester *et al.* 2013).

In light of these results, a further simulation study was carried out in Chapter 5, but with Bayesian methodologies (Lee 2004; Rice 2007; Gelman *et al.* 2013) being employed for the analysis of FEC data. A Bayesian approach to analysing data offers benefits such as the usual normality assumption within statistical models being typically removed and simplifications and unrealistic assumptions being avoided when considering data. The more appealing advantage of the Bayesian paradigm is the idea of being able to incorporate external information into an analysis. Though, this concept of being able to introduce external information into an analysis via prior specifications of parameters is one that is controversial, since classical approaches to Statistics, i.e. the frequentist view, involves any analyses being objective and based purely on the observed data. Research has been invested into using Bayesian methods and obtaining credible intervals when investigating apparent anthelmintic performance, but mainly with respect

to equine FEC data (Denwood 2010; Denwood *et al.* 2010); though are being employed to analyse cattle and sheep FEC data in more recent studies (Denwood *et al.* 2008; Dobson *et al.* 2012; Busin *et al.* 2013; Geurden *et al.* 2015; Wang *et al.* 2017).

Using the priors and likelihoods defined in Section 5.2.1, medians of the posteriors obtained for each percentage estimate were utilised as point estimates of the sampling distribution for each percentage estimate and 95% credible intervals, i.e. Highest Posterior Density intervals, were obtained. In fact, priors and likelihoods (such as the truncated normal distribution) were identified and utilised to obtain posteriors for newly considered percentage estimates, such as averaging over individual-based egg count percentage reductions/changes of various forms (Cabaret and Berrag 2004; Berry and Ayers 2006). The accuracy of estimating various percentage estimates was examined in our simulation study (via the Root Mean Squared Error) for scenarios involving different diagnostic sensitivities, treatment group sizes and classification of pre-treatment group means, as before. In the majority of scenarios considered, in order to obtain percentage estimates that were most accurate, one would only need to adopt a paired study design involving a positive treatment group (as highlighted in Table 5.14), with even some of the newly considered percentage estimates being considered the most accurately estimated. This type of experimental design has proved popular in the veterinary parasitology community and has been widely adopted due to the convenience of not having to include a negative control group (Kochapakdee *et al.* 1995; Lyndal-Murphy *et al.* 2010; Levecke *et al.* 2012; Vidyashankar *et al.* 2012; Lester *et al.* 2013; Stratford *et al.* 2014; Geurden *et al.* 2015; George *et al.* 2017), which may not always be possible, depending on the number of animals on farms.

There are several opportunities to further develop the work using Bayesian methodologies work. For instance, other likelihoods and/or the refinement of likelihoods utilised, could be considered in light of the diagnostic sensitivities that featured as part of our study. The Negative Binomial (NB) distribution however, was utilised as a likelihood when considering the hybrid sets of data involving 1 epg counts and the ZIPIG distribution was utilised as a likelihood for 15 and 30 epg McMaster data, on the basis that these were representative of the best fitting distributions

for the majority of the respective data presented in Chapter 3. Additionally, a prior sensitivity analysis could be carried out with respect to parameters of location, i.e. on parameters μ and μ_1 for the NB and ZIPIG likelihoods, respectively. As part of our simulation study, a prior sensitivity analysis was carried out on the parameter σ, mainly due to FECs being over-dispersed/aggregated and this can impact on anthelmintic studies (Wilson *et al.* 1996; Shaw and Dobson 1995; Morgan *et al.* 2005; Levecke *et al.* 2012). Finally, further methods for sampling posterior distributions and different proposal distributions could be explored (Gelman *et al.* 2013) as well as considering different burn-in periods. The latter being of particular interest, since well mixed chains appeared to be present after a burnin-period of 2000 iterations as part of convergence diagnostics, i.e. Appendix B.2, as opposed to the considered burnin period of 5000 iterations. However, a burnin period of 5000 iterations was then considered a conservative number of iterations to feature as part of an initial discard and was still able to satisfy the convergence diagnostics explored.

One limitation of our Bayesian simulation study that is worth reflecting on, was that the smallest number of animals being considered was 15 (i.e. a total of 30 animals in a given scenario). Published guidelines, such as Coles *et al.* (1992) and Coles *et al.* (2006), recommend the use of treatment group sample sizes of between 10-15 animals. In the scenario of having, what could be considered as, small treatment group sample sizes, one must ask if these sizes are able to provide sufficient power for these types of studies. Denwood (2010) comments on this, saying that prospective power calculations are rarely, performed prior to undertaking a FECRT study and, with respect to equine data, are usually found to be under-powered. As a result, it was more of interest to consider larger treatment group sample sizes as part of our simulation studies and power calculations could be explored/developed as part of future work using our developed Bayesian methodologies for cattle FEC data. It is worth mentioning that Denwood (2010) also provides power calculation methods, based on equine FECRT data and their developed Bayesian methodologies, as part of the *bayescount* package available in *R/RStudio* (Denwood 2015; CRAN 2018), which is a testament to the importance of power calculations being included and developed for these types of studies.

In Chapter 6, the apparent anthelmintic efficacy status of cattle herds in the

UK was classified using available field study data and our Bayesian methodologies developed in Chapter 5, along with adapted published guideline thresholds on efficacy. It is worth remembering that treatment groups with pre-treatment means less than 100 epg were not considered as part of this analysis. This is because it is considered unlikely that egg counts less than 100 epg would be associated with sufficient worm burdens and/or pasture contamination to justify anthelmintic use (and associated efficacy testing). Anthelmintic use in such cases would not be consistent with best practice recommendations (Coles *et al.* 1992; Coles *et al.* 2006). Indeed, as part of the Defra project, composite screening FECs were performed, and only those herds with a composite count greater than 150 epg were enrolled onto a subsequent FECRT. However, despite these composite screening attempts, repeated individual baseline faecal egg sampling on some of these farms subsequently revealed mean egg counts less than 100 epg at the baseline sampling point before treatment. This highlights the variability in results that can be obtained between composite and individual samples etc. It is worth mentioning that several studies have investigated the most appropriate methodology for conducting composite samples (Morgan *et al.* 2005; Calvete and Uriarte 2013; George *et al.* 2017). This resulted in 53 positive treatment groups available to be considered as part of the analysis, with 23 of them featuring as part of parallel group designs with a negative control group.

For the hybrid sets of data collected with a 1 epg sensitivity: we observed that the majority of treatment groups were classified, using our Bayesian methodology developed for the percentage estimate $100\left(1-\frac{T_{14}}{T_0}\right)\%$, as having an apparent LOE status (39 out of 53 positive treatment groups). Whereas, using FECs that had been collected with less sensitive counting techniques (i.e. 15 epg sensitivity) and our developed Bayesian methodology for the percentage estimate $\frac{\sum_j^{n_{treat}}\left[100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%\right]}{n_{treat}}$, we observed that the majority of treatment groups were classified as being either apparently efficacious or having an apparent SLOE status (25 and 14 out of 53 positive treatment groups classed as having an apparent efficacious and suspected lack of efficacy status, respectively). As a result, the number of treatment groups classed as having a LOE status increased upon improving the diagnostic sensitivity of the counting techniques that were utilised,

again emphasising on the importance of the diagnostic sensitivity of counting techniques when investigating anthelmintic performance in cattle (El-Abdellati *et al.* 2010). As well as this, we concluded that there was at least an estimated significant moderate agreement between all sets of comparisons of classifications produced by our developed Bayesian and Defra project methods, for both types of diagnostic sensitivities being considered (i.e. 1 and 15 epg counts).

Based on those FECs that had been collected with a 15 epg: the majority of ivermectin injectable treatment groups were classed as exhibiting an apparent lack of efficacy status (10 out of 17 injectable treatment groups). For those who received a pour-on formulation, the majority of these groups exhibited either an apparently efficacious or suspected lack of efficacy status (5 out of 9 pour-on treatment groups). For those that received doramectin, all three and fifteen treatment groups that received a pour-on formulation and injectable formulation, respectively, were apparently efficacious or had a suspected lack of efficacy status. For those that received fenbendazole, the majority (7 out of 9 treatment groups) were apparently efficacious.

On the basis of the results, discussion and caveats of using FEC data highlighted in Section 6.4, it would be recommended to:

- utilise a paired study design using a positive treatment group only;
- collect FEC data using a 15 epg sensitivity method;
- conduct with pre-treatment group means/composite samples greater than 150 epg;
- use a minimum treatment group sample size of 15 animals;
- and adopt the approach of obtaining the mean value of individual-based egg count percentage reductions/changes of animals using the *Symmetrised Percentage Change* form (along with its associated 95% credible interval) with our developed Bayesian methodology,

in order to improve the FECRT, to conclude overall on the apparent efficacy status of cattle herds.

One may ask themselves and argue however, *"Why would it be worth going to all the extra effort of carrying out this recommended Bayesian method, when there is a significant moderate agreement between the classifications it obtains with those obtained using currently recommended frequentist approaches?"* In response to this, the fact that our novel Bayesian method (involving individual-based egg count percentage reductions/changes of animals of the SPC form) is able to obtain a significant moderate agreement in the classifications it obtains compared with classifications from methods currently utilised in anthelmintic studies is encouraging to observe. Our novel Bayesian method is statistically sound as well, on the grounds that one does not need to assume normality of FEC data (which is an assumption that does not hold when dealing with cattle FEC data) in order to produce interval estimates, such as those provided by WAAVP guidelines and other communications (Coles *et al.* 1992; Coles *et al.* 2006; Lyndal-Murphy *et al.* 2014). Indeed, Denwood (2010) mentions that the lack of appropriate statistical guidelines for carrying out a FECRT is unacceptable. Though, some communications have made comment on the availability of those involved/interested in anthelmintic studies to be able to carry out advanced calculations within a Bayesian framework (Matthews 2014), such as our recommended novel method. As a result, it was decided to produce a field-based *FECRT Calculator*. It was envisaged that users would be able to carry out analyses of FECRT data using our recommended developed Bayesian methodology that was observed to have been *robust* and exhibited at least a significant moderate agreement in classifying the efficacy status of treatment groups, i.e. evaluating the average of individual-based egg count percentage reductions/changes of the SPC form. This Bayesian calculation would be carried out with minimal effort required by the user, which is one of the main reasons as to why a webpage application was developed as part of this project, which is described in more detail in the following Section. It would be envisaged, as part of future work, to promote and convince those involved/interested in anthelmintic studies to invest in the Bayesian paradigm for carrying out statistical analyses as part of the FECRT and, particularly, to use our recommended developed Bayesian method and webpage application for analysing FEC data from U.K. cattle populations as part of anthelmintic efficacy/resistance studies.

7.1 *Bayesian FECRT Calculator* prototype webpage application

The *R Shiny* prototype webpage application named: *Bayesian FECRT Calculator* was developed as a means of carrying out the analysis of treatment group FECRT data using our recommended developed Bayesian methodology. The prototype webpage application can be found here: `http://outreach.mathstat.strath.ac.uk/apps/FECRT`. Shiny is an *R/RStudio* package that makes it easy for users to build interactive web applications straight from *R/RStudio* and can host stand alone applications on a webpage (Chang 2017; Shiny 2018). Figure 7.1 highlights our developed webpage:

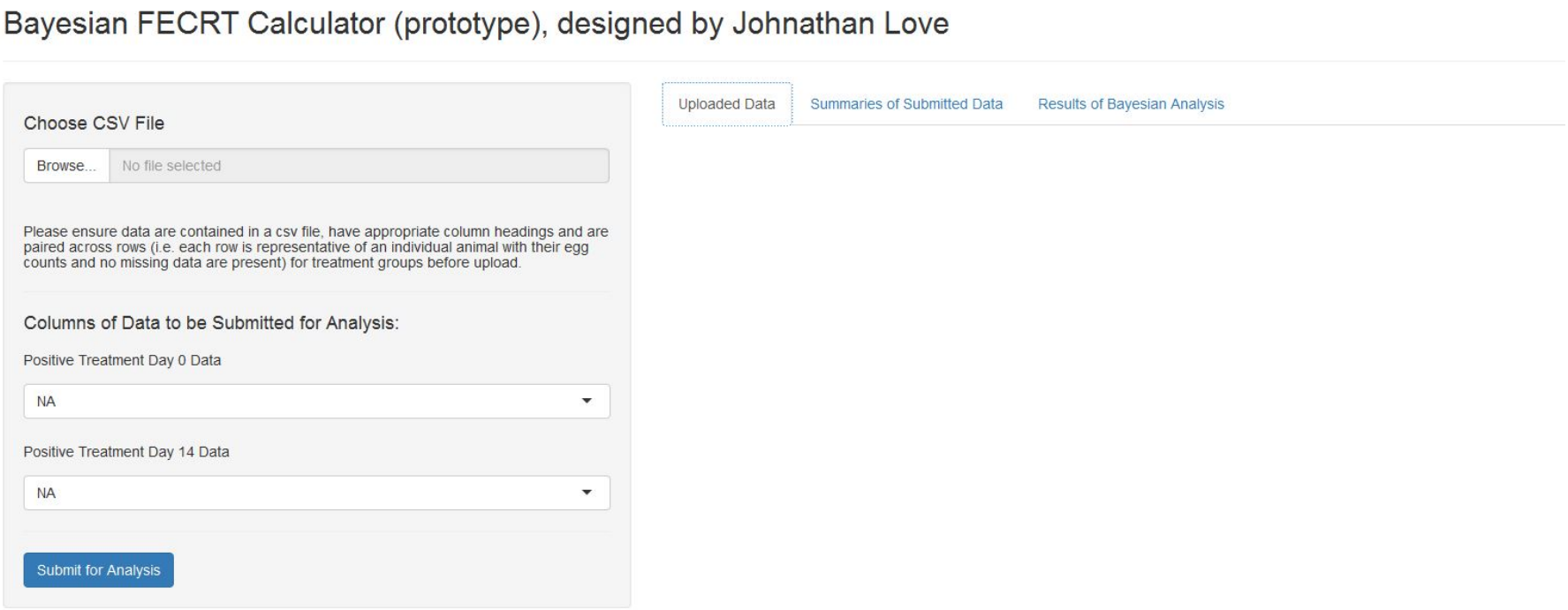

Figure 7.1: *Bayesian FECRT Calculator* prototype webpage application

It is worth noting that the user interface (i.e. the display) of the prototype webpage application may have been updated/altered/further developed, since the time of submission/publishing of this thesis.

To begin using the prototype webpage application, users must first upload a data set with their positive treatment groups' FECs and ensure that:

- data are contained in a comma separated values (CSV) file,
- have appropriate column headings and
- data are paired across rows in the CSV file for treatment groups (i.e. each row is representative of an individual animal with their egg counts and no missing data are present).

Figure 7.2 highlights the webpage after a data set been uploaded (note that users will be able to view the data set on completion of the upload).

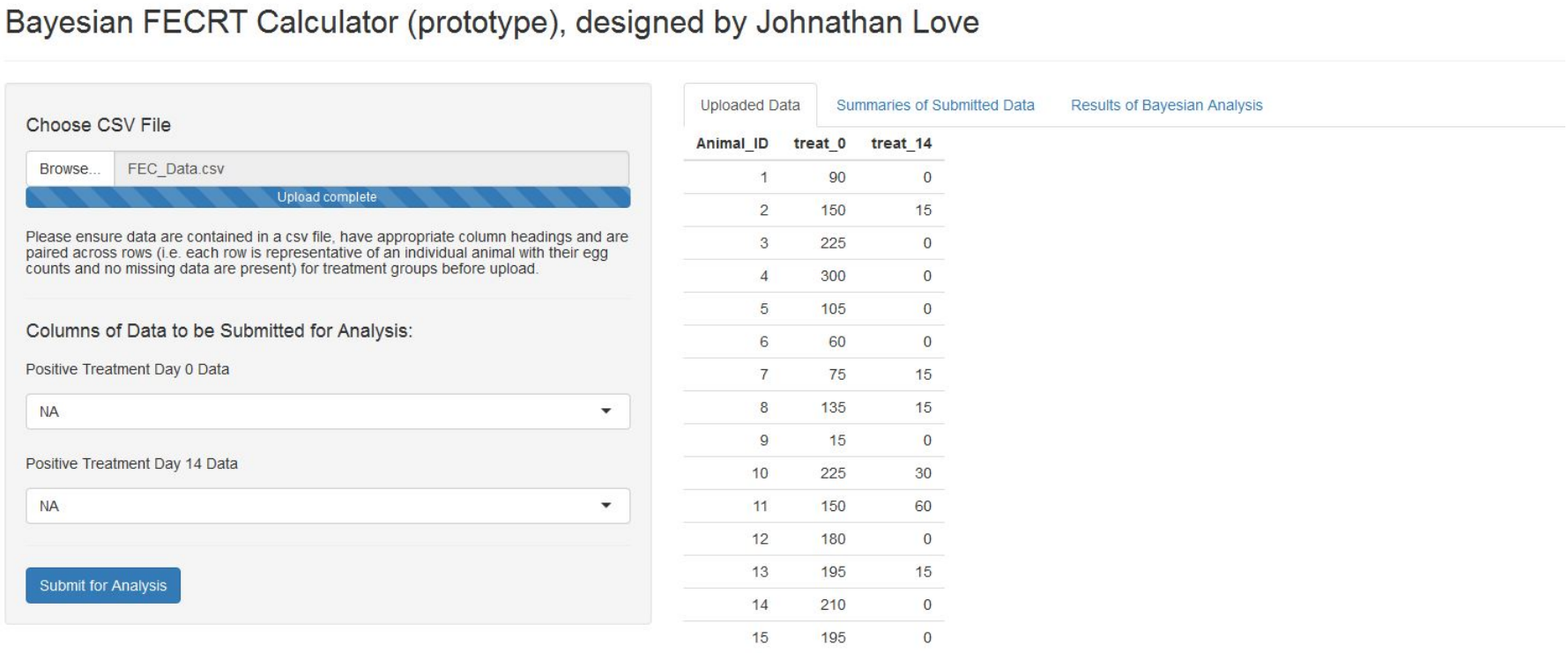

Figure 7.2: Uploading data into the prototype webpage application

Afterwards, users would communicate to the application what columns of data they want to analyse, via the drop down windows featured under the heading: *Columns of Data to be Submitted for Analysis*. So in our example, under the heading *Positive Treatment Day 0 Data* we would select from the drop down window, the column heading treat_0 in our uploaded data set (Figure 7.2) to tell the application that this is the column of data to be analysed as our positive treatment Day 0 data, and so forth. Figure 7.3 shows the prototype webpage application after selecting columns of data treat_0 and treat_14, in our featured data set, from the drop down windows under the headings *Positive Treatment Day 0 Data* and *Positive Treatment Day 14 Data*, respectively.

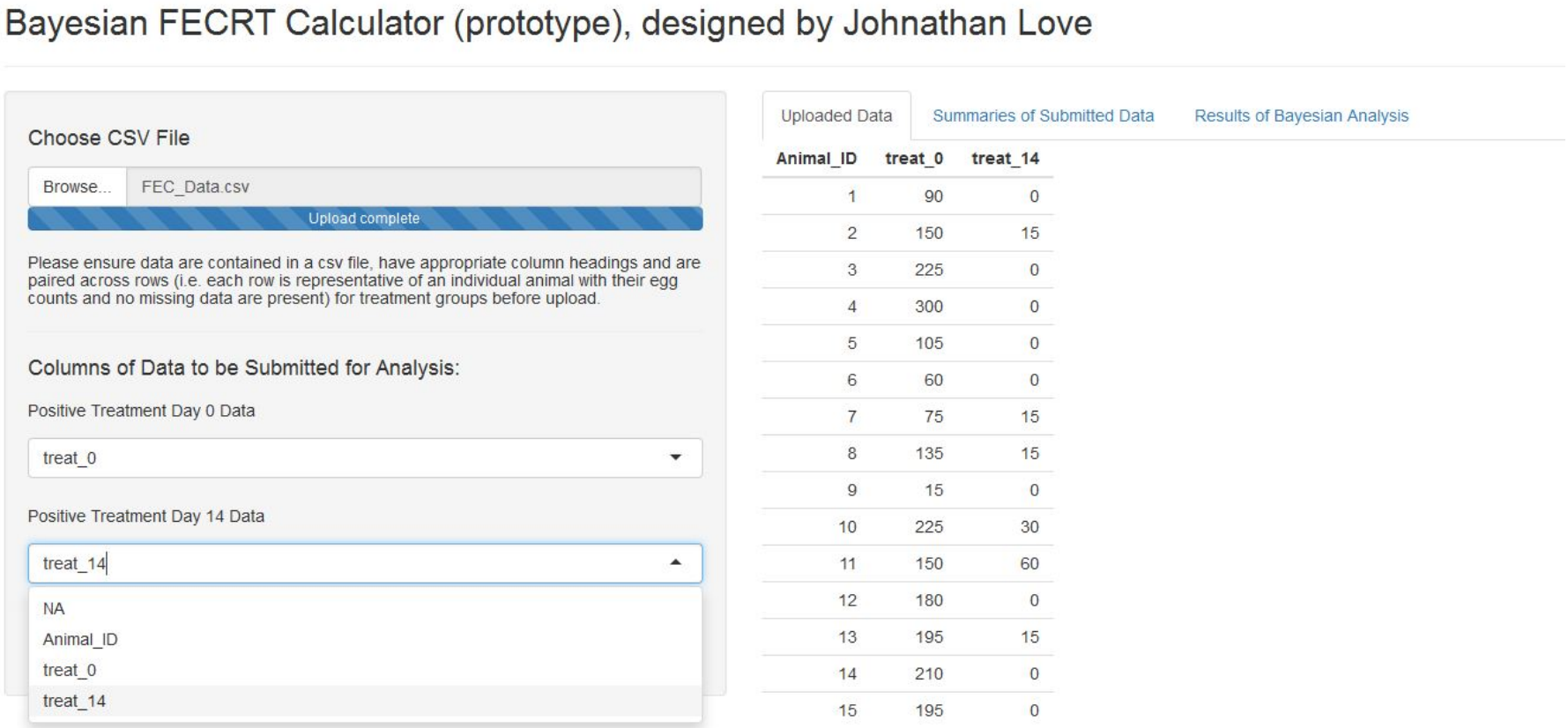

Figure 7.3: Selecting columns of data in drop down windows in prototype webpage application

Once users are satisfied with their selection of columns, then they need only click the *Submit for Analysis* button. Afterwards, users would be expected to click on the *Summaries of Submitted Data* tab to view histograms and summary statistics of the columns of data that have been considered for analysis. It is worth noting however, that after clicking the *Submit for Analysis* button, if users click on the *Summaries of Submitted Data* tab, they may observe a *Running Analysis* counter in the bottom right hand side of their screens - this indicates that the data are in the process of being analysed. Figure 7.4 displays the output that would typically be expected to be viewed from the *Summaries of Submitted Data* tab. It is worth noting that summary statistics and a histogram relating to individual-based egg count percentage reductions/changes of the SPC form are also produced.

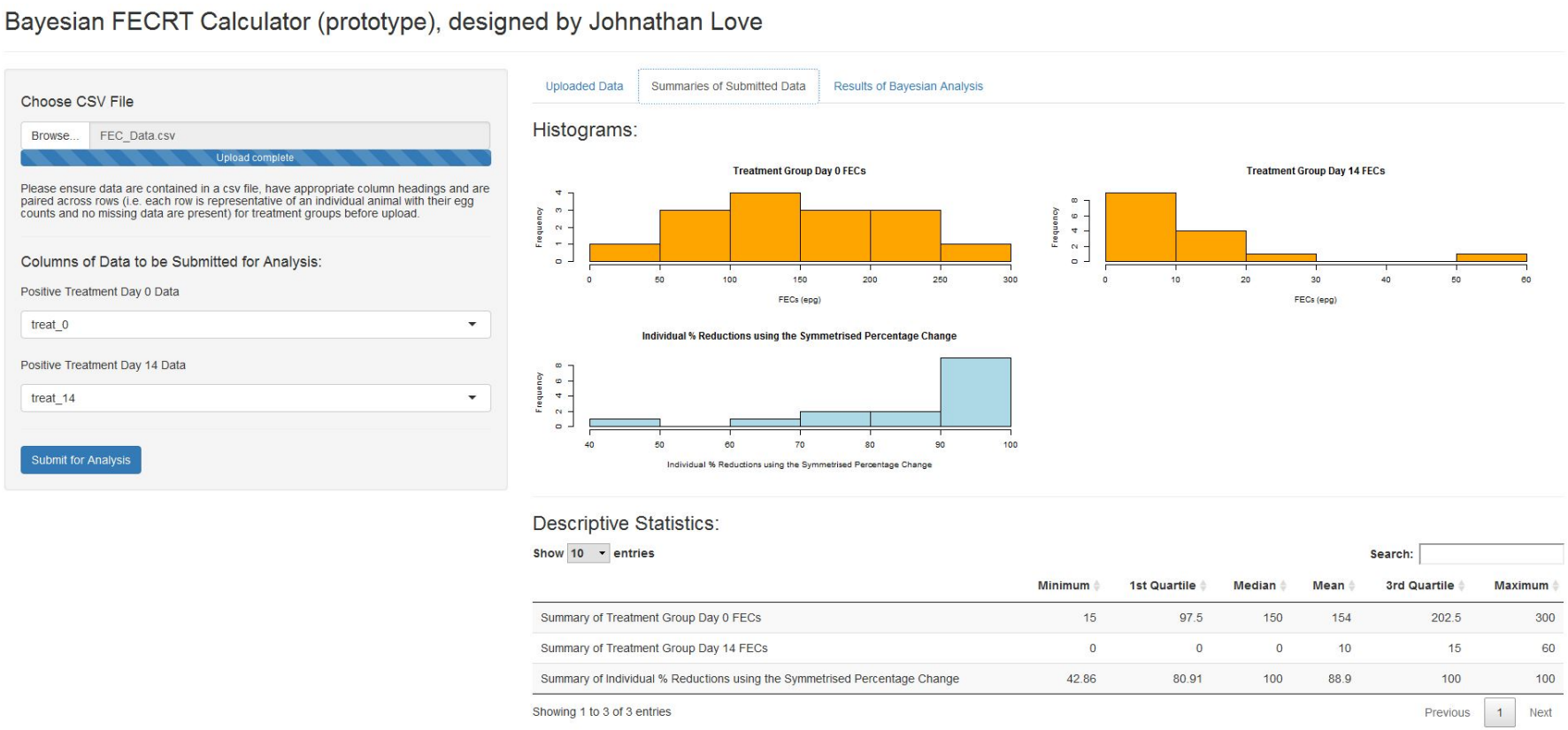

Figure 7.4: *Summaries of Submitted Data* tab output in prototype webpage application

After obtaining summaries of the submitted data, users can click on the *Results of Bayesian Analysis* tab to view the estimated percentage estimate and credible interval that have been obtained based on our developed Bayesian methodology presented in Chapter 5. Figure 7.5, highlights the output expected to be obtained in this tab. Users will note that the level of credibility for the associated interval is 95% and this is the value pre-defined (i.e. the default value) for producing credible intervals in the webpage application.

The result presented in the table is based on utilising a truncated normal distribution likelihood for individual-based egg count percentage reduction/change data of the SPC form, as described in Section 5.2.1. Therefore, the average of individual-based egg count percentage reductions/changes based on the SPC form with an associated 95% credible interval is evaluated and envisaged to be reported.

As well as this, a description of the interpretation associated with a credibility interval is featured and a disclaimer of using the reported result and the caveats of utilising egg count data to investigate anthelmintic performance is also displayed.

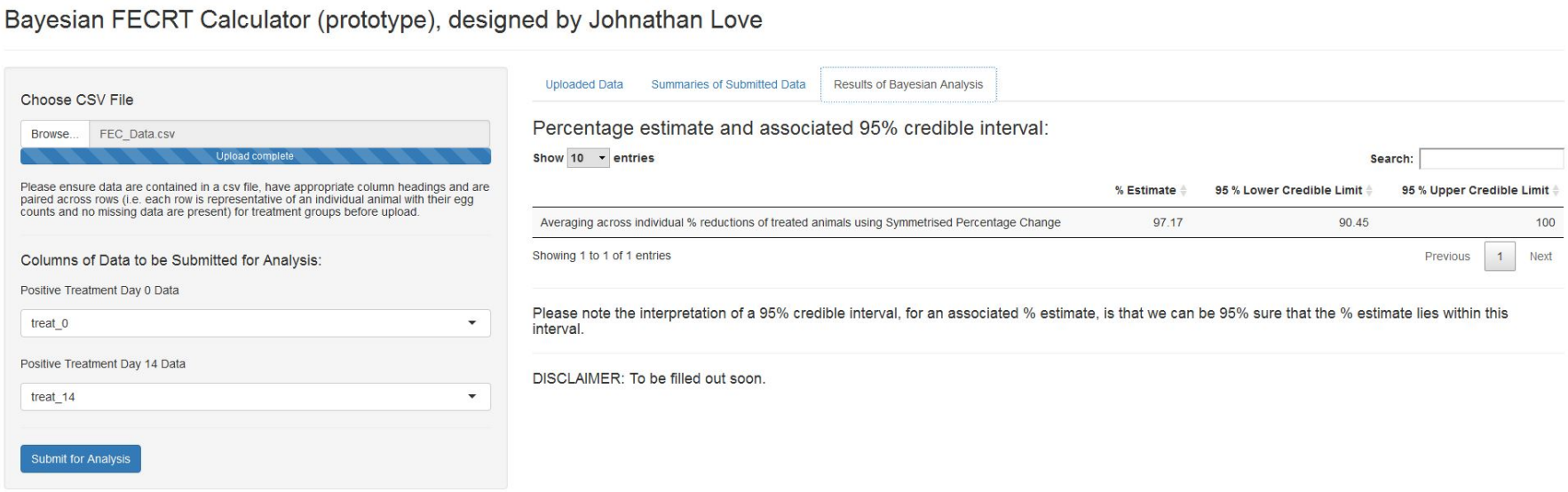

Figure 7.5: *Results of Bayesian Analysis* tab output in prototype webpage application

Future development and work would be needed to develop this prototype webpage application into a more user-friendly smartphone or tablet application into which data could be entered directly from the field or laboratory.

Overall summary:

The work presented in this thesis has given insight into statistical techniques, frameworks, experimental designs and practical considerations for improving the FECRT. Published guidelines have traditionally recommended the use of a parallel group study design (i.e. involving a positive treatment and negative control group), and subsequently evaluate percentage estimates based on the arithmetic means of FECs and their associated 95% confidence intervals (derived by assuming FEC data to be normal). In this thesis, however, it was shown that the majority of available cattle FEC data violated the assumption of normality and as a result, Bayesian methodologies were developed for obtaining percentage estimates and associated 95% credible intervals. Within this statistical framework and based on our results, practical recommendations are made, such as the diagnostic sensitivities of counting techniques and treatment group sample sizes to be used in the field. The use of a robust percentage estimate is also recommended, that is, averaging over individual-based egg count reductions/changes of the *Symmetrised Percentage Change* form. This percentage estimate only requires a paired study design with a positive treatment group, and can be evaluated, along with an associated 95% credible interval, via our *Bayesian FECRT Calculator* prototype webpage application. For the field of veterinary parasitology, it is envisaged that the findings, conclusions, discussions and outputs from this work, will aid and improve further research in anthelmintic studies involving livestock species, in the foreseeable future.

Appendix A

R/RStudio code used for simulating data

A.1 Code for simulating Negative Binomial data

```
require("gamlss")||install.packages("gamlss")
require("mvtnorm")||install.packages("mvtnorm")

#Reading in the data:
control <- read.table("file_name.csv", header=TRUE,sep=",")
treat <- read.table("file_name.csv", header=TRUE,sep=",")

#Obtaining the parameter estimates from using the real data to be used:
mu_control_base<-mean(control$"Baseline_Counts")
sd_control_base<-sd(control$"Baseline_Counts")

mu_control_eos<-mean(control$"EOS_Counts")
sd_control_eos<-sd(control$"EOS_Counts")

mu_treat_base<-mean(treat$"Baseline_Counts")
sd_treat_base<-sd(treat$"Baseline_Counts")
```

```
mu_treat_eos<-mean(treat$"EOS_Counts")
sd_treat_eos<-sd(treat$"EOS_Counts")

#Dealing with the correlation of the control group:
corr_control<-0.4
S_control<-matrix(c(1,corr_control,corr_control,1),2,2)

#Dealing with the correlation of the treatment group:
corr_treat<-0.3
S_treat<-matrix(c(1,corr_treat,corr_treat,1),2,2)

#Obtaining Simulated Data:
counter <- 0
dataList <- vector("list", length = 1000)

while(counter < 1000){

AB_control <- rmvnorm(mean=c(0,0),sig=S_control,n=15)
U_control <- pnorm(AB_control)
simul_cont_base <- qNBII(U_control[,1],mu=mu_control_base,sigma=sd_control_base)
simul_cont_eos <- qNBII(U_control[,2],mu=mu_control_eos,sigma=sd_control_eos)

AB_treat <- rmvnorm(mean=c(0,0),sig=S_treat,n=15)
U_treat <- pnorm(AB_treat)
simul_treat_base <- qNBII(U_treat[,1],mu=mu_treat_base,sigma=sd_treat_base)
simul_treat_eos <- qNBII(U_treat[,2],mu=mu_treat_eos,sigma=sd_treat_eos)


if(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))>0 &
cor(simul_cont_base,simul_cont_eos,method=c("pearson"))<0.5 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))>0 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))<0.5|
```

```
is.na(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))=="TRUE") &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))>0 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))<0.5|
is.na(cor(simul_treat_base,simul_treat_eos,method=c("pearson"))=="TRUE")
& cor(simul_cont_base,simul_cont_eos,method=c("pearson"))>0 &
cor(simul_cont_base,simul_cont_eos,method=c("pearson"))<0.5|
is.na(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))=="TRUE")
& is.na(cor(simul_treat_base,simul_treat_eos,method=c("pearson"))=="TRUE"))
{
    counter<-counter+1

    c0<-rep("c0",15)
    sim_cont0<-rep(paste("Simulation",counter),15)
    dt_c0<-as.data.frame(cbind(simul_cont_base,c0,sim_cont0))
    names(dt_c0)<-c("Data","Treatment Status","Simulation")

    c14<-rep("c14",15)
    sim_cont14<-rep(paste("Simulation",counter),15)
    dt_c14<-as.data.frame(cbind(simul_cont_eos,c14,sim_cont14))
    names(dt_c14)<-c("Data","Treatment Status","Simulation")



    t0<-rep("t0",15)
    sim_treat0<-rep(paste("Simulation",counter),15)
    dt_t0<-as.data.frame(cbind(simul_treat_base,t0,sim_treat0))
    names(dt_t0)<-c("Data","Treatment Status","Simulation")

    t14<-rep("t14",15)
    sim_treat14<-rep(paste("Simulation",counter),15)
    dt_t14<-as.data.frame(cbind(simul_treat_eos,t14,sim_treat14))
    names(dt_t14)<-c("Data","Treatment Status","Simulation")

    dataList[[counter]] <- rbind(dt_c0,dt_c14,dt_t0,dt_t14)
```

```
    print(counter)

  }

}

# bring data together into single data frame
final_data_frame <- Reduce("rbind", dataList)
write.table(final_data_frame,file="Simulated_Data.csv",append=FALSE,
sep=",",col.names=TRUE,row.names=FALSE)
```

A.2 Code for simulating zero inflated data

```
require("gamlss")||install.packages("gamlss")
require("mvtnorm")||install.packages("mvtnorm")

#Location parameter for the zero inflated distribution:
modified_mean<-function(d){
  mean(d)/(1-(length(subset(d,d==0))/length(d)))
}

#Reading in the data:
control <- read.table("file_name.csv", header=TRUE,sep=",")
treat <- read.table("file_name.csv", header=TRUE,sep=",")

#Obtaining the parameter estimates from using the real data to be used:
mu_control_base<-modified_mean(control$"Baseline_Counts")
sd_control_base<-sd(control$"Baseline_Counts")
nu_control_base<-length(subset(control$"Baseline_Counts",
```

```
control$"Baseline_Counts"==0))/length(control$"Baseline_Counts")

mu_control_eos<-modified_mean(control$"EOS_Counts")
sd_control_eos<-sd(control$"EOS_Counts")
nu_control_eos<-length(subset(control$"EOS_Counts",
control$"EOS_Counts"==0))/length(control$"EOS_Counts")

mu_treat_base<-modified_mean(treat$"Baseline_Counts")
sd_treat_base<-sd(treat$"Baseline_Counts")
nu_treat_base<-length(subset(treat$"Baseline_Counts",
treat$"Baseline_Counts"==0))/length(treat$"Baseline_Counts")

mu_treat_eos<-modified_mean(treat$"EOS_Counts")
sd_treat_eos<-sd(treat$"EOS_Counts")
nu_treat_eos<-length(subset(treat$"EOS_Counts",
treat$"EOS_Counts"==0))/length(treat$"EOS_Counts")

if(nu_control_base==0){nu_control_base<-1e-16}
else{nu_control_base<-nu_control_base}
if(nu_treat_base==0){nu_treat_base<-1e-16}
else{nu_treat_base<-nu_treat_base}
if(nu_control_eos==0){nu_control_eos<-1e-16}
else{nu_control_eos<-nu_control_eos}
if(nu_treat_eos==0){nu_treat_eos<-1e-16}
else{nu_treat_eos<-nu_treat_eos}

if(nu_control_base==1){nu_control_base<-0.9999999999999999}
else{nu_control_base<-nu_control_base}
if(nu_treat_base==1){nu_treat_base<-0.9999999999999999}
else{nu_treat_base<-nu_treat_base}
if(nu_control_eos==1){nu_control_eos<-0.9999999999999999}
else{nu_control_eos<-nu_control_eos}
if(nu_treat_eos==1){nu_treat_eos<-0.9999999999999999}
```

```
else{nu_treat_eos<-nu_treat_eos}

#Dealing with the correlation of the control group:
corr_control<-0.4
S_control<-matrix(c(1,corr_control,corr_control,1),2,2)

#Dealing with the correlation of the treatment group:
corr_treat<-0.3
S_treat<-matrix(c(1,corr_treat,corr_treat,1),2,2)

#Obtaining Simulated Data:
counter <- 0
dataList <- vector("list", length = 1000)



while(counter < 1000){

AB_control <- rmvnorm(mean=c(0,0),sig=S_control,n=15)
U_control <- pnorm(AB_control)
simul_cont_base <- qZIPIG(U_control[,1],mu=mu_control_base,
sigma=sd_control_base,nu=nu_control_base)
simul_cont_eos <- qZIPIG(U_control[,2],mu=mu_control_eos,
sigma=sd_control_eos,nu=nu_control_eos)

AB_treat <- rmvnorm(mean=c(0,0),sig=S_treat,n=15)
U_treat <- pnorm(AB_treat)
simul_treat_base <- qZIPIG(U_treat[,1],mu=mu_treat_base,
sigma=sd_treat_base,nu=nu_treat_base)
simul_treat_eos <- qZIPIG(U_treat[,2],mu=mu_treat_eos,
sigma=sd_treat_eos,nu=nu_treat_eos)
```

```
if(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))>0 &
cor(simul_cont_base,simul_cont_eos,method=c("pearson"))<0.5 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))>0 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))<0.5|
is.na(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))=="TRUE") &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))>0 &
cor(simul_treat_base,simul_treat_eos,method=c("pearson"))<0.5|
is.na(cor(simul_treat_base,simul_treat_eos,method=c("pearson"))=="TRUE") &
cor(simul_cont_base,simul_cont_eos,method=c("pearson"))>0 &
cor(simul_cont_base,simul_cont_eos,method=c("pearson"))<0.5|
is.na(cor(simul_cont_base,simul_cont_eos,method=c("pearson"))=="TRUE") &
is.na(cor(simul_treat_base,simul_treat_eos,method=c("pearson"))=="TRUE"))
  {
    counter<-counter+1

    c0<-rep("c0",15)
    sim_cont0<-rep(paste("Simulation",counter),15)
    dt_c0<-as.data.frame(cbind(simul_cont_base,c0,sim_cont0))
    names(dt_c0)<-c("Data","Treatment Status","Simulation")

    c14<-rep("c14",15)
    sim_cont14<-rep(paste("Simulation",counter),15)
    dt_c14<-as.data.frame(cbind(simul_cont_eos,c14,sim_cont14))
    names(dt_c14)<-c("Data","Treatment Status","Simulation")


    t0<-rep("t0",15)
    sim_treat0<-rep(paste("Simulation",counter),15)
    dt_t0<-as.data.frame(cbind(simul_treat_base,t0,sim_treat0))
    names(dt_t0)<-c("Data","Treatment Status","Simulation")

    t14<-rep("t14",15)
    sim_treat14<-rep(paste("Simulation",counter),15)
```

```
    dt_t14<-as.data.frame(cbind(simul_treat_eos,t14,sim_treat14))
    names(dt_t14)<-c("Data","Treatment Status","Simulation")

    dataList[[counter]] <- rbind(dt_c0,dt_c14,dt_t0,dt_t14)

    print(counter)

  }

}
# bring data together into single data frame
final_data_frame <- Reduce("rbind", dataList)
write.table(final_data_frame,file="Simulated_Data.csv",append=FALSE,
sep=",",col.names=TRUE,row.names=FALSE)
```

Appendix B

Information relating to Bayesian simulation study

B.1 Generating samples from posterior distributions

B.1.1 Gibbs sampler

Recall Equation 2.18 from Section 2.4.3:

$$\pi(\theta|x) = \frac{f(x|\theta)p(\theta)}{f(x)},$$

where $\pi(\theta|x)$ is the posterior distribution of the parameters $\theta = \{\theta_1, \ldots, \theta_m\}$ given the data $x = \{x_1, \ldots, x_n\}$, $f(x|\theta)$ is the probability of observing the data x under different parameter values θ (this is known as the likelihood), $p(\theta)$ is the prior distribution of the parameters and $f(x)$ is a normalisation constant so that the posterior distribution is a valid probability density function. In the case of discrete random variables, the probability density functions are replaced with probability mass functions.

Suppose we are interested in the posterior distribution of $\theta = \{\theta_1, \ldots, \theta_m\}$. Then we can construct a Markov Chain using the following *Gibbs Sampling* algorithm (Gelman *et al.* 2013):

The Gibbs Sampler:

- *Begin with arbitrary starting values for the parameters, denoted as* $\theta^0 = \{\theta_1^0, \ldots, \theta_m^0\}$.
- *At iteration* t *say, of the Markov Chain let the set of parameter values be denoted by* θ^t. *From iteration* t *to* $t+1$, *we update each of the parameters in turn, using their corresponding posterior conditional distributions. Mathematically, this algorithm can be summarised as the following:*

$$\textit{Simulate: } \theta_1^{t+1} \sim \pi(\theta_1|\theta_2^t, ..., \theta_m^t).$$
$$\textit{Update: } \theta_2^{t+1} \sim \pi(\theta_2|\theta_1^{t+1}, \theta_3^t, ..., \theta_m^t) \textit{ and so on...}$$
$$\textit{Finally: } \theta_m^{t+1} \sim \pi(\theta_m|\theta_1^{t+1}, ..., \theta_{m-1}^{t+1}).$$

The order in which parameters are updated does not matter, since it can be shown that this algorithm produces a Markov chain with stationary distribution π Gelman *et al.* (2013). In other words, if we run the chain long enough with enough iterations, the simulated values can be regarded as a sample from the posterior distribution.

One useful feature of the Gibbs sampler is that it tends to be efficient MCMC-wise, with respect to the fact that the parameters are always updated within each iteration. In fact one can also use *blocked* Gibbs sampling, which involves two or more groups of variables and samples from their joint distributions, conditioned on all other variables, as opposed to sampling from each one individually. On that note, if the posterior distribution is highly dimensional then the Gibbs sampler can result in being computationally slow (irrespective of blocking being present or not) compared to other samplers, but of course this depends on the number of parameters contributing to the posterior distribution.

One obvious disadvantage of utilising this sampler however, is that the posterior conditional distributions need to be derived for the parameters. Indeed, if conjugate priors are utilised in computations, then the derivations of the posterior conditional distributions can result in simpler algebra being carried out, however the use of a conjugate prior may not always be utilised. If the posterior distribution is of a non-standard form, more complex algorithms need to be implemented to sample from the distribution. One of which is the *Metropolis-Hastings* algorithm, which will be described next.

B.1.2 Metropolis-Hastings (MH) sampler

The Metropolis-Hastings (MH) Sampler:

- *Begin with arbitrary starting values for the parameters, denoted as* $\theta^0 = \{\theta_1^0, \ldots, \theta_m^0\}$.
- *At iteration* t *say, of the Markov Chain let the set of parameter values be denoted by* θ^t. *We update the parameter values using a two-step procedure:*

Step 1: Sample a candidate value $\phi \sim q(\phi|\theta^t)$,
where ϕ *is known as the candidate point*
and q *is known as the proposal distribution.*

Step 2: With probability $\alpha(\theta^t, \phi) = min\left(1, \frac{\pi(\phi|x)q(\theta^t|\phi)}{\pi(\theta^t|x)q(\phi|\theta^t)}\right)$
set $\theta^{t+1} = \phi$(*candidate value is accepted*)
else, set $\theta^{t+1} = \theta^t$(*candidate value is rejected*).

Due to the form of the acceptance probability, we only need to know π up to proportionality. Though, the choice of proposal distribution is essentially arbitrary and often requires some *pilot-tuning*.

B.1.2.1 Flavours of MH schemes

Within the MH scheme, there are a number of special cases depending upon the (essentially arbitrary) choice of proposal distribution and here we will consider a number of the most common type of proposal distributions.

One basic and common choice is to centre the proposal around the current parameter value θ^t. In other words, we propose the candidate value:

$$\phi = \theta^t + z, \text{where } z \sim f.$$

For example, we may consider the proposal distribution (f) to be $\mathcal{N}(0, \sigma^2 I)$, where σ^2 is to be chosen. This choice is known as a *Random Walk* MH scheme.

It is worth noting that the choice of a proposal distribution need not be one which is symmetric. In the instance of a symmetric proposal distribution being utilised however, this is known as a *Metropolis Update*. In the instance of these types of proposals being used, $q(\theta^t|\phi) = q(\phi|\theta^t)$ and so the acceptance probability in the MH algorithm simplifies to:

$$\alpha(\theta^t, \phi) = min\left(1, \frac{\pi(\phi|x)}{\pi(\theta^t|x)}\right)$$

There is also the case of utilising an *Independence Sampler*, in which the proposal distribution is independent of the current state of the chain, i.e. $q(\phi|\theta^t) = q(\phi)$, but this algorithm does not perform as well as previous types of updates mentioned above.

It is also worth considering that in some instances the posterior distribution will be high-dimensional. In this case, we can either update the parameters simultaneously (as in the random walk MH scheme), or we can update each parameter in turn. This is then known as a *Single-Update* MH and is considered below:

A glimpse into the Single-Update MH algorithm:

Suppose the set of parameter values is θ^t.

Step 1: Propose the value ϕ^1 *for* θ_1

(keep all the other parameter values the same).

Accept/Reject the proposed value with standard acceptance probability.

Step 2: Propose the value ϕ^2 *for* θ_2

(Note we now use the updated value θ_1^{t+1}*)*

Accept or Reject this value and so on...

(Recall this is the same idea for the Gibbs sampler).

In order to balance the size of the proposed moves with the chance of accepting them, the proposal variance from the proposal distribution is often *tuned* to obtain an acceptance rate of 20-40%. Recall from Chapter 2, that the acceptance rate is essentially the percentage of times that the chain obtains a unique value. Most MH algorithms require some pilot-tuning in order to achieve satisfactory performance (Gelman *et al.* 2013). It is worth noting that in the case of a Metropolis Update being used, even though the acceptance probability is not a function of the proposal variance σ^2; the candidate value ϕ is dependent on σ^2.

If one were to propose extremely small jumps, one is virtually bound to accept them, but it will take a long time to move around the posterior distribution. Alternatively, when proposing extremely large jumps one does have the potential to move further but will generally have smaller acceptance rates. In other words, there is a trade-off between the acceptance rate of a chain and the speed at which the posterior distribution is traversed. This movement around the parameter space is often referred to as *mixing* and hence why trace plots are often obtained to observe this. In fact, some MH algorithms are programmed to be *adaptive*,

meaning the proposal variance is modified within the running of the algorithm, in order to improve the acceptance rates of the parameters at a given iteration of the chain, which is a concept also discussed by Gelman *et al.* (2013).

One of the main advantages of these schemes is the fact that posterior conditional distributions need not be derived to perform them, particulary in the instance of distributions being of a non-standard form. If one were also to utilise the concept of a symmetric proposal distribution, in the case of a Random Walk MH scheme, then this also makes for easier computation with respect to the acceptance probabilities, since the posterior distributions are all that would be required for these to be evaluated. We could also say that if one were to adopt the concept of updating each parameter in turn, as in the case of a Single-Update MH scheme and with the Gibbs Sampler: the scheme would be efficient MCMC-wise. With these concepts and advantages in mind, it is a variation of the MH algorithm that has been utilised in our simulation study, as described in Chapter 5.

B.2 Examples of convergence diagnostics using developed Metropolis-Hastings (MH) algorithms

There are many methods in which we can assess that the distribution of parameters has reached convergence to the stationary distribution and that our chains are well mixed across the parameter spaces (Gelman *et al.* 2013). For instance, this can be assessed through examining relevant trace plots, acceptance rates (essentially the percentage of unique values the chain has), autocorrelations of the parameters and one could also consider deriving the Brooks-Gelman-Rubin Statistic. It is worth noting however, that even though convergence is guaranteed mathematically, there is no way of being able to *prove* if the distribution of parameters has converged to the stationary distribution - these assessments can only provide an *indication* of lack of convergence and how efficient we are being when sampling estimates from the stationary distribution.

In our examples to follow, for each of the four simulated data sets being consid-

ered, two chains were generated with over-dispersed initial starting values for each parameter involved. This is an approach recommended by Gelman *et al.* (2013). Essentially, we are running several chains and comparing the output from each individual chain and this can provide additional re-assurance since, in theory, the chains should converge to the stationary distribution, i.e. the posterior distribution, and that no major modes have been missed in any one simulation.

As a means of assessing apparent convergence for the MH algorithms implemented, here we consider chains and the parameters for four sets of data: the first simulated Day 14 positive treatment group and $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ sets of data (from farm E32 with treatment group sample sizes of 50), where the relevant Day 0 and Day 14 data were both simulated from a $NBII$ distribution (which was informed by 15EPG_McM_SCFT study data) and a $ZIPIG$ distribution (which was informed by 15EPG_McM field study data).

Overall, there are four examples of pairs of chains being considered here and the assessment of convergence of the MH algorithms will be examined for each.

B.2.1 Trace plots and acceptance rates

When considering the simulated positive treatment group Day 14 data, based on 15EPG_McM_SCFT study data which were used to inform a $NBII$ distribution to simulate from, one chain was ran with initial starting values $\mu = 300$ and $\sigma = 450$ (referred to as *Chain 1*) and the other chain ran from initial starting values $\mu = 1$ and $\sigma = 5$ (referred to as *Chain 2*). Figure B.1 highlights the trace plots for these two chains and the parameters considered.

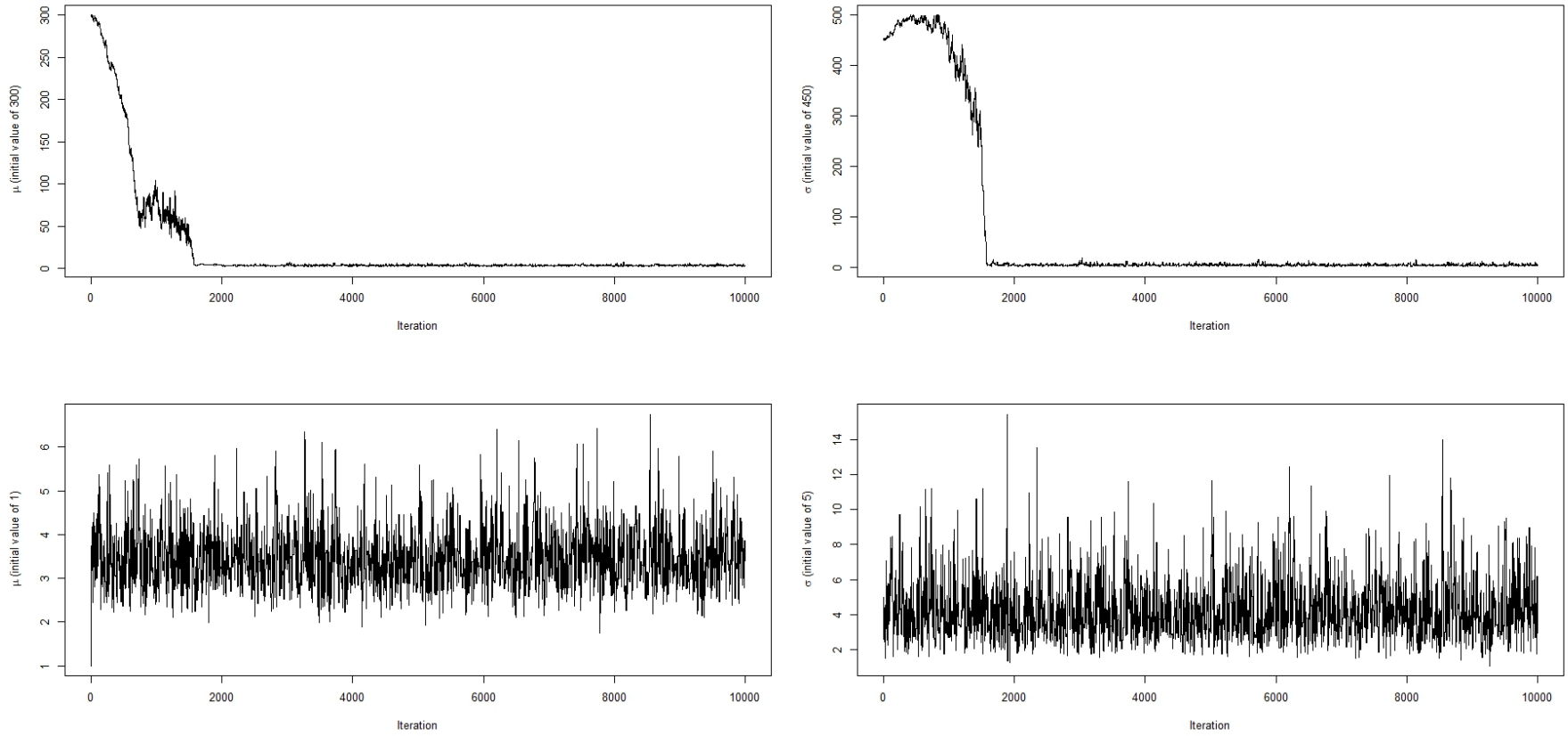

Figure B.1: Trace Plots of parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data)

From Figure B.1, we are able to observe that even though the initial values were over-dispersed for each parameter, the chains for each parameter appear to converge to a common mode and result in well mixed distributions being generated. It is worth noting that in the scenario where initial starting values $\mu = 300$ and $\sigma = 450$ were considered, the chains are not as well mixed up to the $2000th$ iteration. This is due to the fact that to ensure each of the parameters' chains were tuned and explored their respective posterior distributions efficiently (i.e. achieved acceptance rates between 20-40% after the initial burnin period) the MH algorithms were programmed to be adaptive. In fact, based on these plots, a burnin period of 2000 iterations could have been considered, thus a burnin period of 5000 iterations is a conservative approach. After the burnin period of 5000 iterations, the acceptance rates of μ and σ for *Chain 1* were 29.92% and 33.08%, respectively and for *Chain 2*, these were 29.38% and 33.38%, respectively and these all lie within the range of between 20-40% as required.

For the same type of data considered but simulated using 15EPG_McM field study data to inform a $ZIPIG$ distribution to simulate from, one chain was ran with initial starting values $\mu = 450$, $\sigma = 450$ and $\nu = 0.9$ (referred to as *Chain 3*) and the other chain ran from initial starting values $\mu = 1$, $\sigma = 5$ and $\nu = 0.05$ (referred to as *Chain 4*). Figure B.2 highlights the trace plots for these two chains.

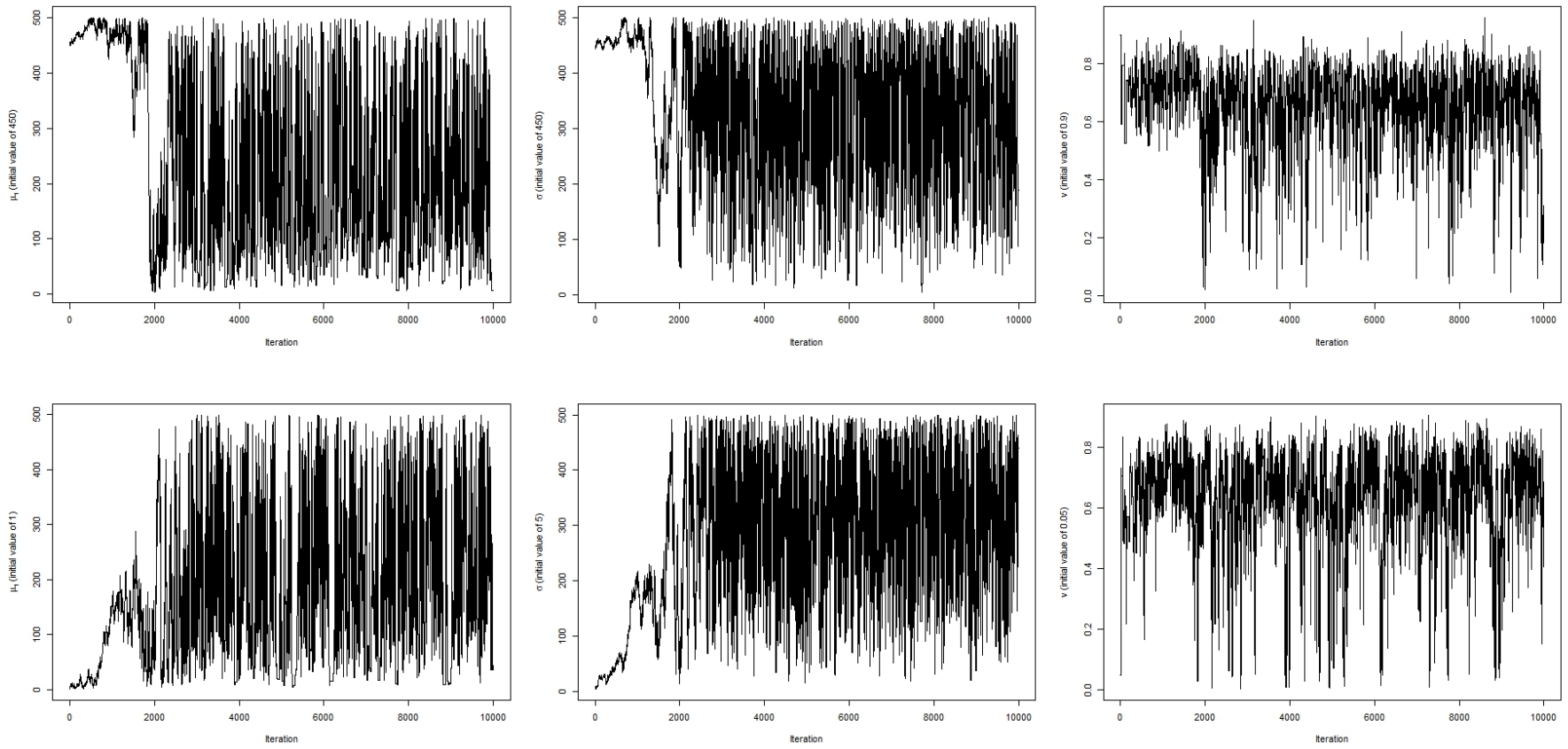

Figure B.2: Trace Plots of parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data)

From Figure B.2, we are able to observe that even though the initial values were over-dispersed for each parameter, i.e. three highly valued starting points and three low valued starting points, the chains for each parameter appear to converge to a common mode and result in well mixed distributions being generated. In fact, with respect to both chains, we observe a lot of the parameter spaces being traversed. Again, we see that both chains are not as well mixed up to the $2000th$ iteration, particularly for the parameters μ and σ. After the burnin period of 5000 iterations however, the acceptance rates of μ, σ and ν for *Chain 3* were 28.42%, 32.78% and 30.8% respectively and for *Chain 4*, these were 28.9%, 32.7% and 30.58% respectively, and these all lie within the range of between 20-40% as required.

For the simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$% data, which consisted of using a $NBII$ distribution to simulate both Day 0 and Day 14 data by using 15EPG_McM_SCFT field study data, one chain was ran with initial starting values $\mu = 100$ and $\sigma = 450$ (referred to as *Chain 5*) and the other chain ran from initial starting values $\mu = 1$ and $\sigma = 5$ (referred to as *Chain 6*). Figure B.3 highlights the trace plots for these two chains.

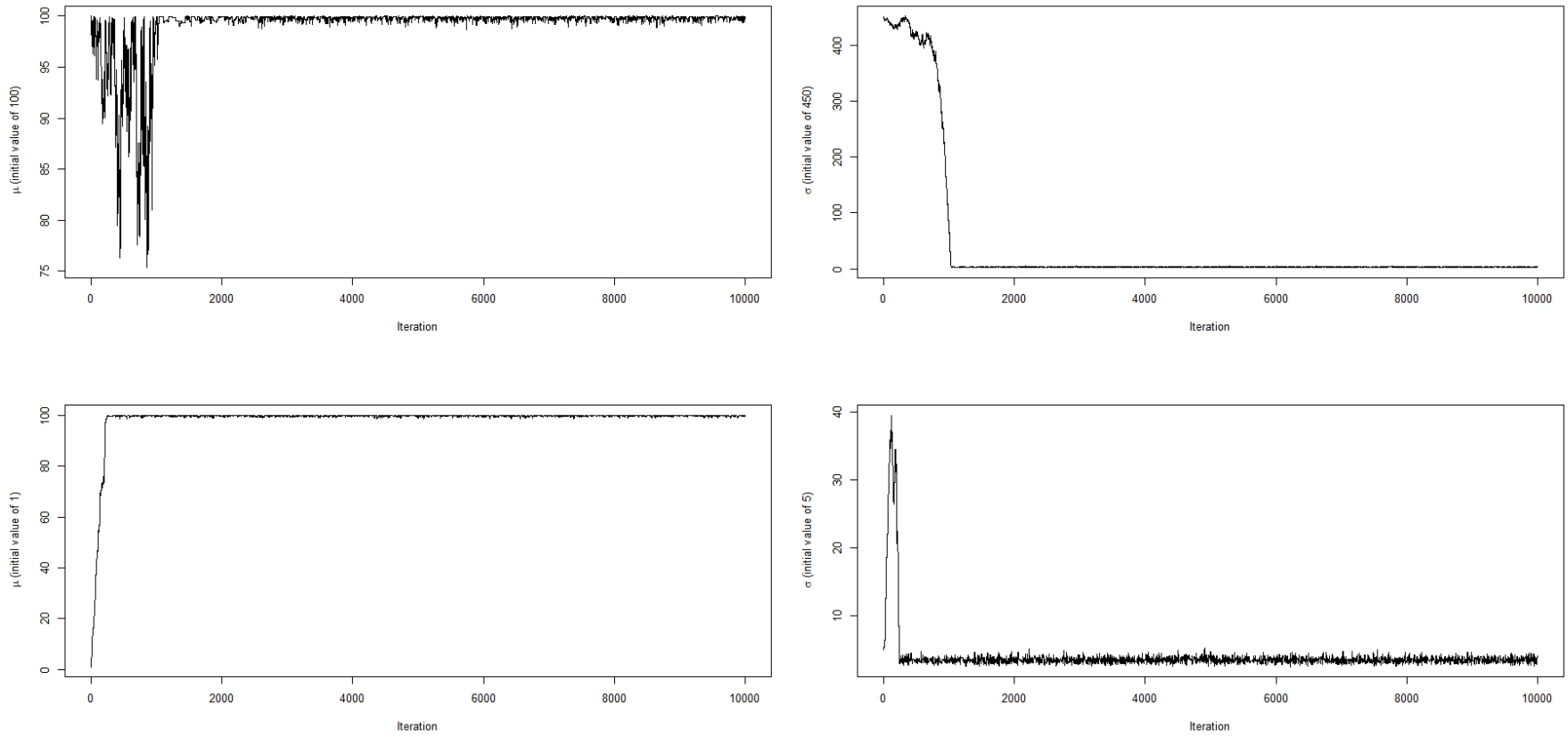

Figure B.3: Trace Plots of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM_SCFT field study data)

From Figure B.3, we are able to observe that the chains for each parameter appear to converge to a common mode and result in well mixed distributions being generated. It is worth noting that for *Chain 5*, we see that a burnin period of 2000 iterations could be considered appropriate; but when considering *Chain 6* a burnin period of 1000 iterations could be considered appropriate. After the burnin period of 5000 iterations however, as programmed into our MH algorithms from a conservative point of view, the acceptance rates of μ and σ for *Chain 5* were 29.9% and 29.16%, respectively and for *Chain 6*, these were 30.58% and 29.1%, respectively, and these all lie within the range of between 20-40% as required.

For the simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$ % data, which consisted of using a $ZIPIG$ distribution to simulate both Day 0 and Day 14 data by using 15EPG_McM field study data, one chain was ran with initial starting values $\mu = 100$ and $\sigma = 450$ (referred to as *Chain 7*) and the other chain ran from initial starting values $\mu = 1$ and $\sigma = 5$ (referred to as *Chain 8*). Figure B.4 highlights the trace plots for these two chains.

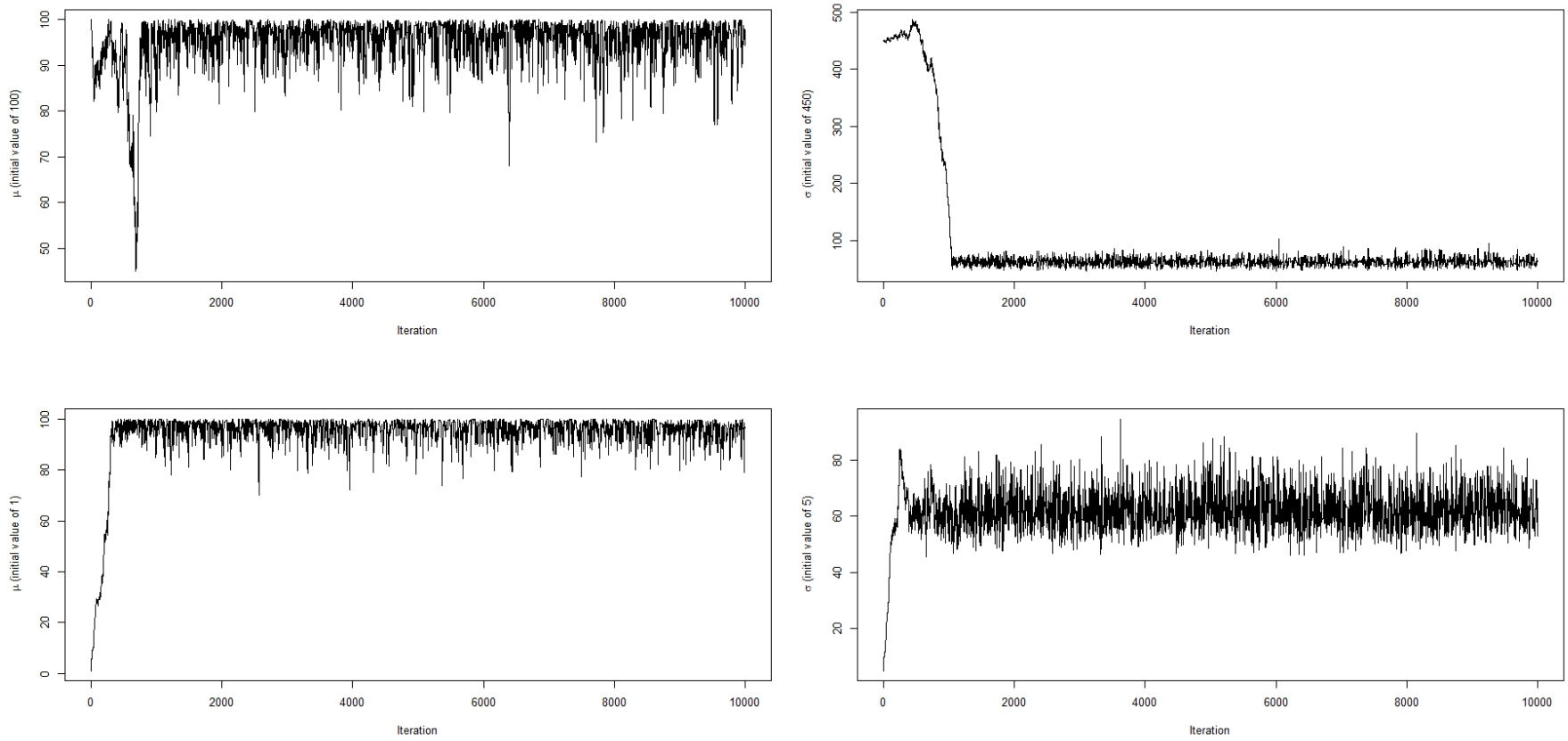

Figure B.4: Trace Plots of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM field study data)

From Figure B.4, we are able to observe that the chains for each parameter appear to converge to a common mode and result in well mixed distributions being generated. It is worth noting that for both chains, a burnin period of 1000 iterations could be considered appropriate. After the burnin period of 5000 iterations however, as programmed into our MH algorithms from a conservative point of view, the acceptance rates of μ and σ for *Chain 7* were 25.64% and 31.4%, respectively and for *Chain 8*, these were 24.82% and 32.5%, respectively and these all lie within the range of between 20-40% as required.

Overall, it appears that, based on inspecting the trace plots and acceptance rates of the chains for the parameters and simulated data considered, we are able to obtain well mixed chains to obtain sample estimates from posterior distributions of the parameters. It would appear that an improvement of the algorithms would be to consider a burnin period of 2000 iterations as opposed to 5000. As a result, 5000 iterations can be considered as a conservative number of iterations to initially discard, and this is still able to leave us with half the total number of iterations to potentially thin out and sample estimates from the posterior distributions and give us acceptance rates of between 20-40% as required.

B.2.2 Autocorrelations

According to Upton and Cook (2011), *autocorrelation* is a measure of the linear relationship between two separate instances of the same random variable. This is distinct from correlation, which refers to the linear relationship between two random variables. As with correlation however, the possible values lie between 1 and -1 inclusive, with unrelated instances having a theoretical autocorrelation of zero.

In the case of MCMC, autocorrelation measures the extent of the linear relationship between parameter estimates at iterations that are a fixed interval (i.e. *lag*) apart in a given chain. The sample autocorrelation for lag l, denoted as r_l, is given (for $l = 1, 2, ..., t-1$) for the ordered sequence of t iterations $\theta^1, \theta^2, ..., \theta^t$ for a parameter θ is defined as:

$$r_l = \frac{\Sigma_{t=1}^{t-l}(\theta^t - \bar{\theta})(\theta^{t+l} - \bar{\theta})}{\Sigma_{t=1}^{t}(\theta^t - \bar{\theta})^2},$$

where $\bar{\theta}$ is the sample mean.

We would expect the lth lag autocorrelation to be smaller as the lag l increases, i.e. estimates from iterations close together will have a higher degree of correlation in comparison to those estimates from iterations that are located further apart. If autocorrelation is still relatively high for higher values of lag l, this would indicate a high degree of correlation between iterations located further apart and slow movement around the parameter space, i.e. slow mixing.

For *Chains 1* and *2*, Figure B.5 displays the autocorrelation, for various lags, for the parameters μ and σ when a burnin of 5000 iterations has been considered with *Chains 1* and *2*. It is clear from this Figure, that the autocorrelation between the estimates reduces as the lag increases for both *Chains 1* and *2*, which indicates that the distributions of parameters are well mixed. Figure B.6 displays similar information, but the autocorrelations are based on a burnin of 5000 iterations having occurred and then thinned-out so every $5th$ estimate contributed towards the sample. From this Figure, we are able to observe that the autocorrelations between estimates for both chains have further reduced, in comparison to those displayed in Figure B.5.

In fact, these conclusions hold for the other autocorrelations of parameter estimates in the various other chains that have been considered in this Section. In a similar fashion, Figures B.7 and B.8 represent the autocorrelations for the parameters involved with *Chains 3* and *4*, Figures B.9 and B.10 represent the autocorrelations for the parameters involved with *Chains 5* and *6*, Figures B.11 and B.12 represent the autocorrelations for the parameters involved with *Chains 7* and *8*. From all these Figures we are able to conclude that the thinning out process has been effective in reducing the amount of autocorrelation in our sample of estimates for each simulated data set considered, and that the distributions of parameters are well mixed.

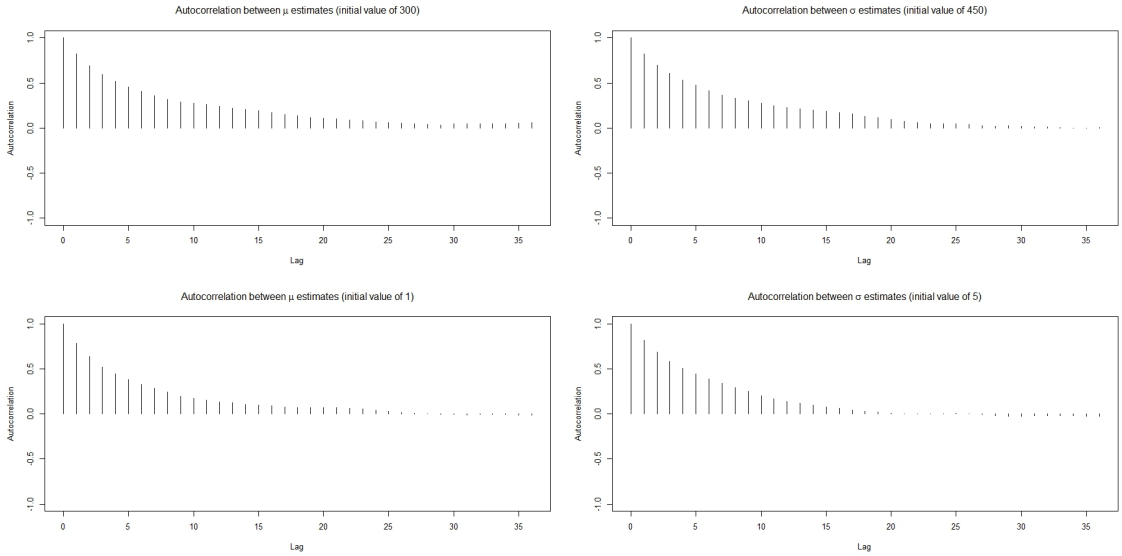

Figure B.5: Autocorrelation between parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) after burnin period

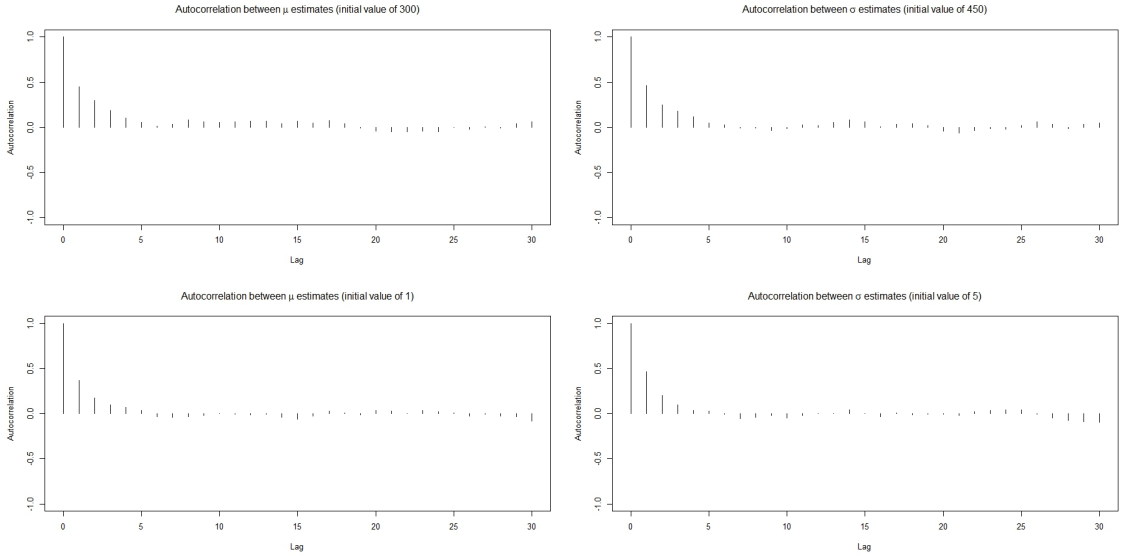

Figure B.6: Autocorrelation between parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data) after burnin period and thinning out

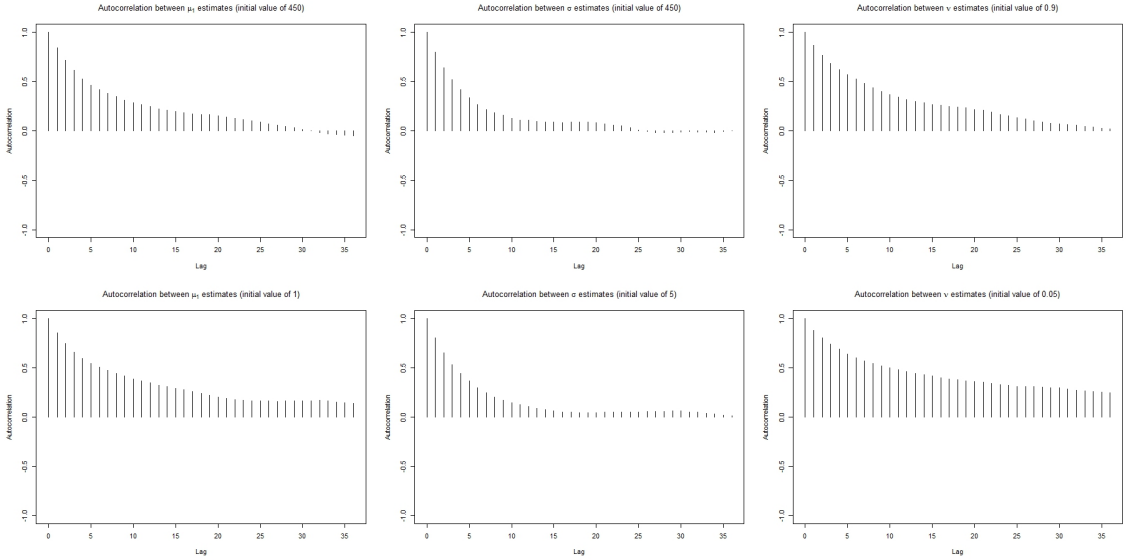

Figure B.7: Autocorrelation between parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) after burnin period

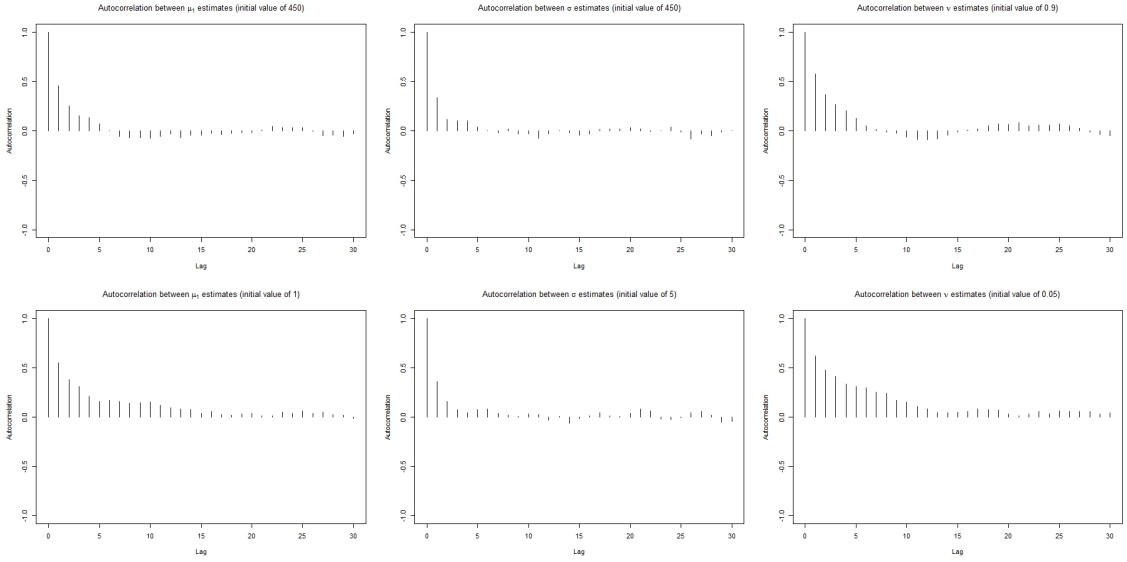

Figure B.8: Autocorrelation between parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data) after burnin period and thinning out

B.2.3 Gelman-Rubin diagnostic

The Gelman-Rubin Statistic, also known as the *Potential Scale Reduction Factor* (PSRF), gives insight into the mixing of chains using *between*- and *within*-chain

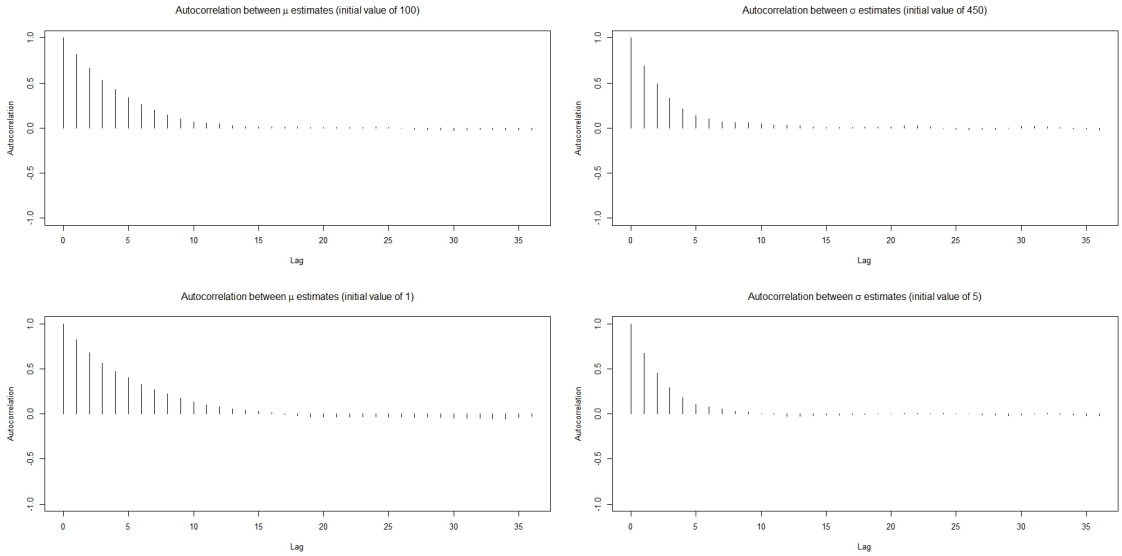

Figure B.9: Autocorrelation between parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$% data (based on 15EPG_McM_SCFT field study data) after burnin period

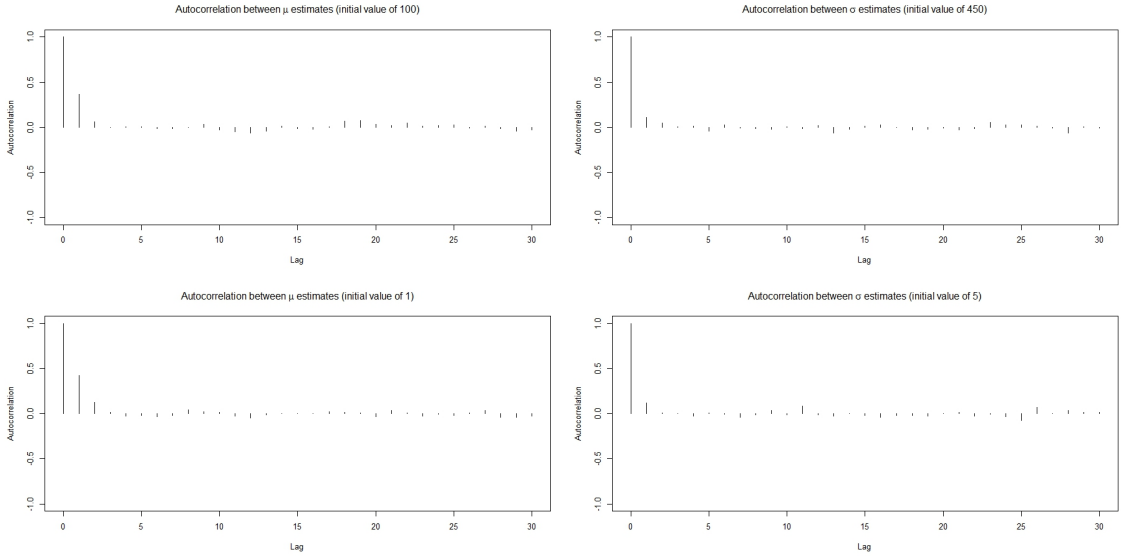

Figure B.10: Autocorrelation between parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$% data (based on 15EPG_McM_SCFT field study data) after burnin period and thinning out

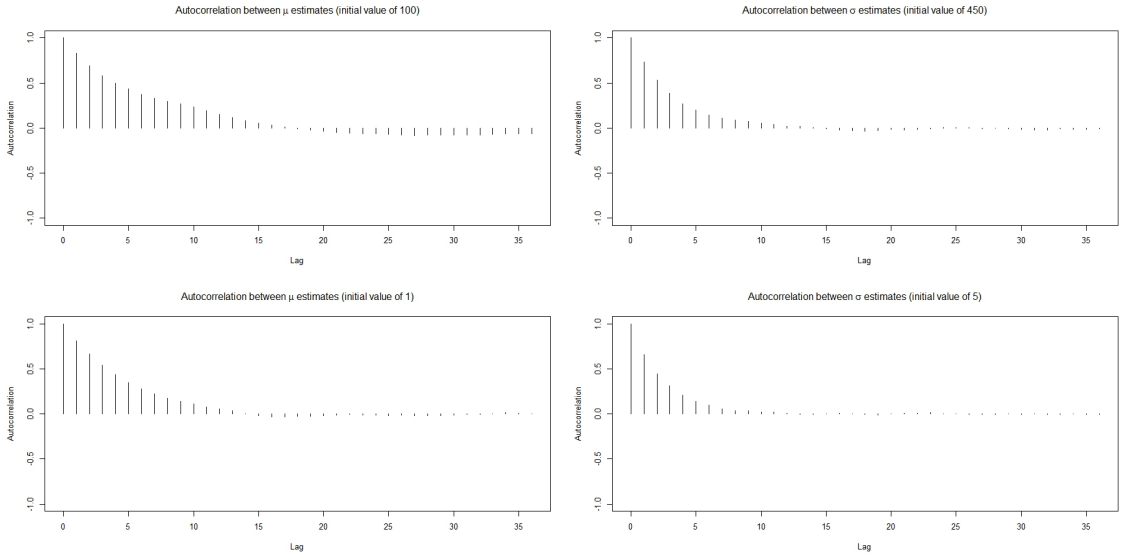

Figure B.11: Autocorrelation between parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$% data (based on 15EPG_McM field study data) after burnin period

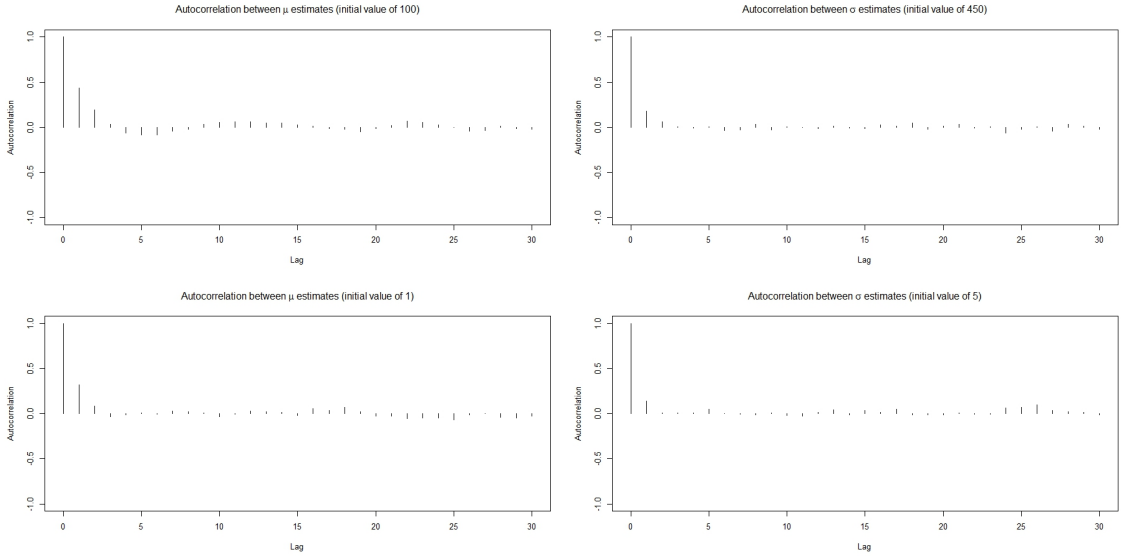

Figure B.12: Autocorrelation between parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)$% data (based on 15EPG_McM field study data) after burnin period and thinning out

variances. According to Gelman *et al.* (2013), we consider this diagnostic after the burnin period has been considered, and as such, the following outputs will be based on the considered chains (i.e. *Chains 1-8*) with the initial burnin period of 5000 iterations having already been considered.

If we let p be the number of chains considered (in each instance for our purposes this value would be 2) and q be the length of each chain (i.e. the number of iterations to be considered which would be 5000 in our case after the considered burnin period of 5000 iterations). For each parameter estimate θ, label the iterations θ_{ij}, $(i = 1, ..., p; j = 1, ..., q)$. Then we can compute B and W, the between- and within-chain variances:

$$B = \frac{p}{q-1}\Sigma_{j=1}^{q}(\bar{\theta}_j - \bar{\theta})^2,$$
$$\text{where } \bar{\theta}_j = \frac{\Sigma_{i=1}^{p}\theta_{ij}}{p} \text{ and } \bar{\theta} = \frac{\Sigma_{j=1}^{q}\bar{\theta}_j}{q}.$$
$$W = \frac{\Sigma_{j=1}^{q}s_j^2}{q} \text{ where } s_j^2 = \frac{\Sigma_{i=1}^{p}(\theta_{ij} - \bar{\theta}_j)^2}{p-1}.$$

The between-chain variance, B, contains a factor p because it is based on the variance of the within-chain means, $\bar{\theta}_j$, each of which is an average of p values θ_{ij}.

We can therefore estimate $var(\theta|x)$, the marginal posterior variance of the estimate, by a weighted average of B and W, namely

$$\hat{var(\theta|x)} = \frac{p-1}{p}W + \frac{1}{p}B.$$

As a result, we monitor convergence of the chain by estimating the factor by which the scale of the current distribution for θ might be reduced if the iterations were continued in the limit $p \to \infty$. The PSRF is then estimated by

$$\hat{R} = \sqrt{\frac{var(\hat{\theta}|x)}{W}},$$

and this declines to the value of one as $p \to \infty$. If the PSRF is high, then this is reason to believe that proceeding with further iterations may improve inference about the distribution of the associated estimate. According to Gelman *et al.* (2013), if $\hat{R}$ is not near the value of one (the word *near* meaning anything less than the value of 1.1 as recommended by Gelman *et al.* (2013)) for each estimate considered, then it is advisable to continue to run the chains for further iterations. As well as this, Brooks and Gelman (1997) have also proposed a multivariate version of the PSRF and the requirements mentioned above are still desired when considering this multivariate case.

Figure B.13 displays the relevant *R/RStudio* output and plots, which have been adapted for display purposes, for the Gelman-Rubin statistics obtained for *Chains 1* and *2* with a burnin period of 5000 iterations. When observing the *R/RStudio* output, the point estimate that is referred to as "[1,]" is the median PSRF for the parameter μ (value 1) and the point estimate that is referred to as "[2,]" is the median PSRF for the parameter σ (also with value 1). The respective upper confidence limits for these median point estimates are also given in the output, both of value approximately 1, since the PSRF is estimated with uncertainty because our chain lengths are finite. The plots available in this Figure display the estimates and the upper confidence limits of the PSRF over the iterations considered. The plot on the left hand side displays this information for parameter μ and the plot on the right hand side displays this information for parameter σ. The multivariate version of the PSRF is also given as standard output, and this has a value of 1.

Overall, based on the output and plots of this Figure, we observe that the median point estimates and upper confidence limits for the PSRF (for both parameters) and the multivariate version of the PSRF are all less than the value of 1.1, meaning that that *Chains 1* and *2* appear to be well mixed and there appears to be no indication of a lack of convergence.

Figure B.14 displays the relevant *R/RStudio* output and plots, which have been adapted for display purposes, for the Gelman-Rubin statistics obtained for *Chains 3* and *4* with a burnin period of 5000 iterations. The main difference between this Figure and Figure B.13, is that there is an additional median point estimate "[3,]" and an additional plot, both of which are to give inference on the estimated PSRF for parameter ν. Again, we observe that the median point estimates and upper confidence limits for the PSRF (for all parameters) and the multivariate version of the PSRF are all less than the value of 1.1, meaning that there appears to be no indication of a lack of convergence for *Chains 3* and *4*.

In fact, this is indeed the case when investigating if there is an indication of a lack of convergence for the pairs of *Chains 5* and *6* and *Chains 7* and *8*, for which the relevant *R/RStudio* output and plots are highlighted in Figures B.15 and B.16.

```
Potential scale reduction factors:

     Point est. Upper C.I.
[1,]          1          1
[2,]          1          1

Multivariate psrf

1
```

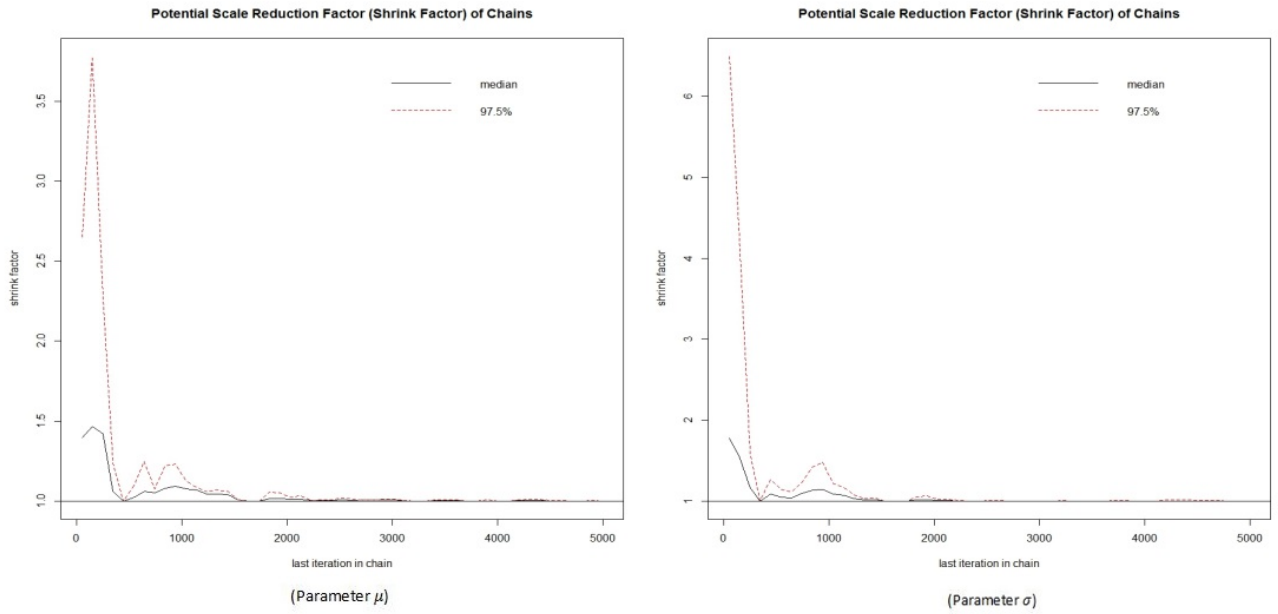

Figure B.13: Potential Scale Reduction Factors of parameters μ and σ using simulated Day 14 positive treatment group data (based on 15EPG_McM_SCFT field study data)

```
Potential scale reduction factors:

     Point est. Upper C.I.
[1,]       1.01       1.03
[2,]       1.00       1.01
[3,]       1.00       1.02

Multivariate psrf

1.01
```

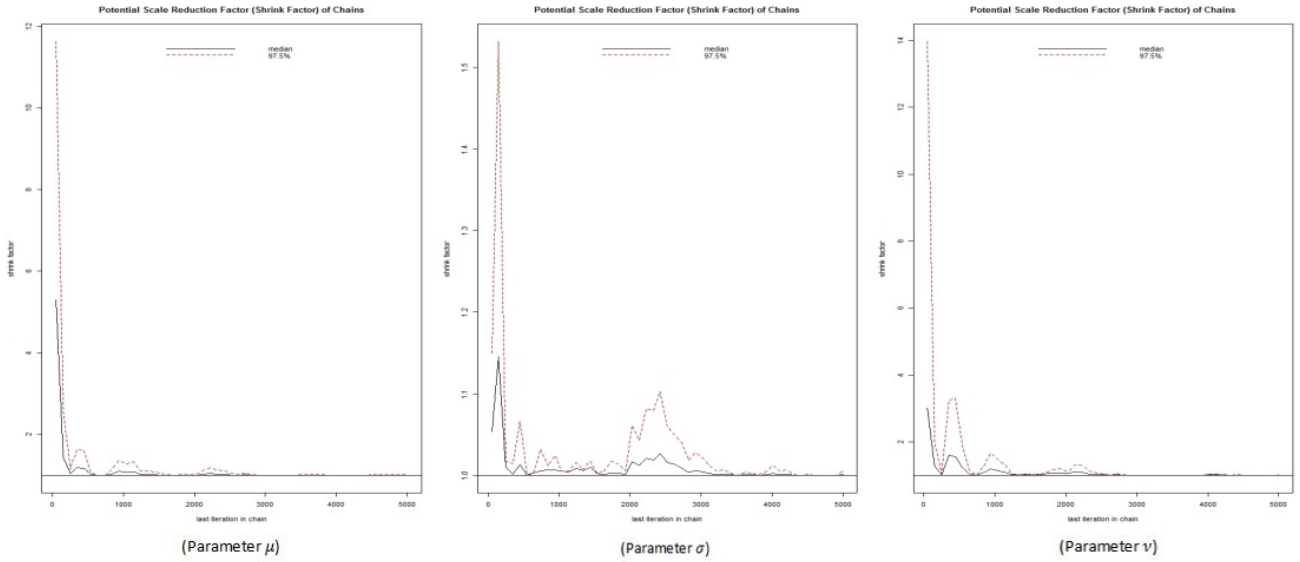

Figure B.14: Potential Scale Reduction Factors of parameters μ, σ and ν using simulated Day 14 positive treatment group data (based on 15EPG_McM field study data)

```
Potential scale reduction factors:

     Point est. Upper C.I.
[1,]          1          1
[2,]          1          1

Multivariate psrf

1
```

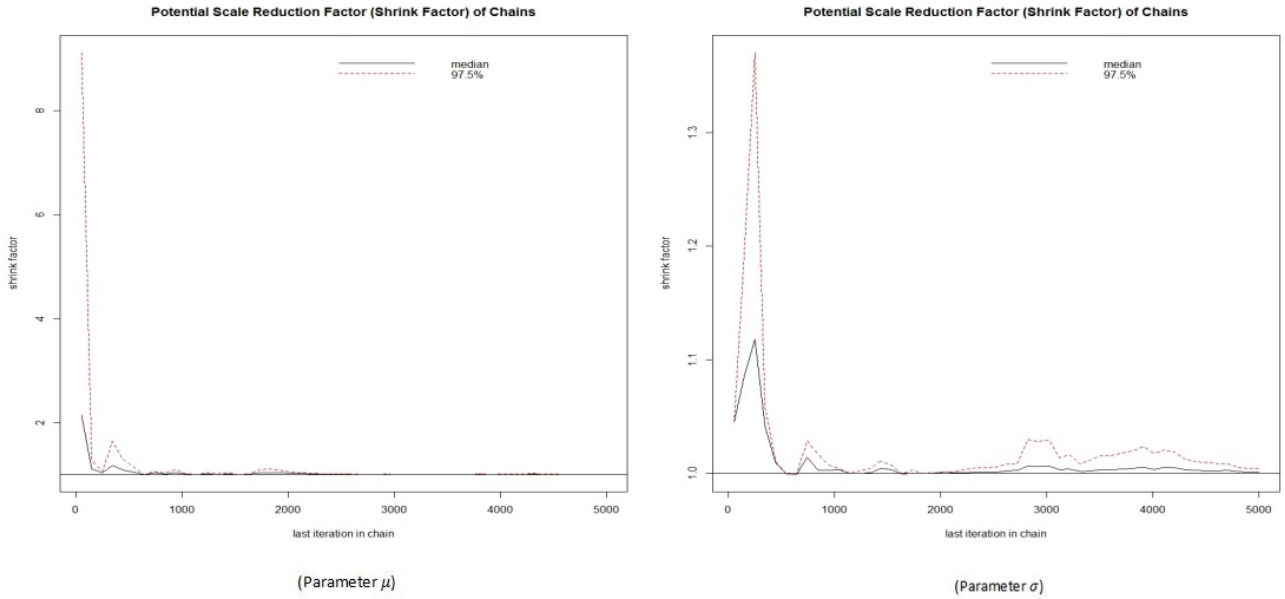

Figure B.15: Potential Scale Reduction Factors of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM_SCFT field study data)

```
Potential scale reduction factors:

     Point est. Upper C.I.
[1,]          1       1.00
[2,]          1       1.01

Multivariate psrf

1
```

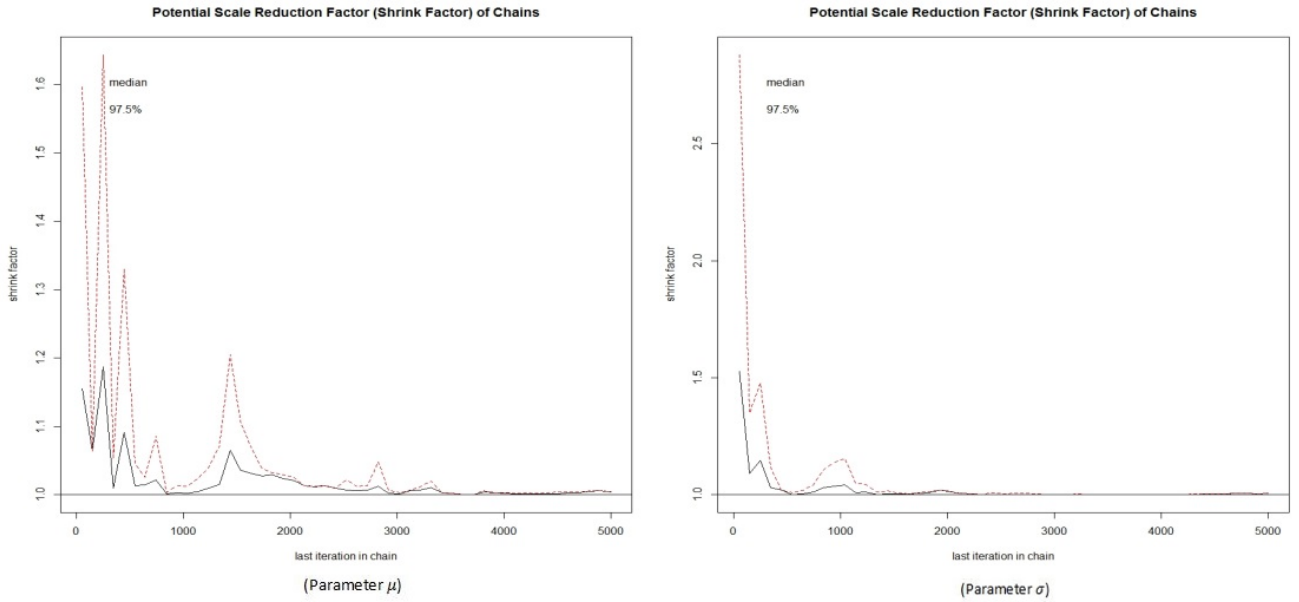

Figure B.16: Potential Scale Reduction Factors of parameters μ and σ using simulated $100\left(\frac{T_{0,j}-T_{14,j}}{T_{0,j}+T_{14,j}}\right)\%$ data (based on 15EPG_McM field study data)

B.3 Additional Concepts: Mean squared errors and the truncated Normal distribution

B.3.1 Mean squared errors

Recall from Section 2.4.2, that the bias of an estimator, $\hat{\theta}$ say, is defined as

$$Bias = E[\hat{\theta}] - \theta,$$

where $E[\hat{\theta}]$ is the average value of the sampling distribution for the estimate $\hat{\theta}$ in a set of simulations (Collins *et al.* 2001; Burton *et al.* 2006; Upton and Cook 2011).

While bias can quantify the average difference to be expected between an estimator and a true parameter, an estimator based on a finite sample can additionally be expected to differ from the true parameter due to the randomness in the sample being considered.

The *Mean Squared Error* (MSE) of an estimator, given in (B.1), can be used to try and reflect both types of difference (Collins *et al.* 2001) and is a measure of overall accuracy for an estimator (Burton *et al.* 2006):

$$\begin{aligned} MSE &= E[(\hat{\theta} - \theta)^2] \\ &= Var[\hat{\theta}] + (Bias)^2, \end{aligned} \tag{B.1}$$

where $Var[\hat{\theta}]$ is the variance of the sampling distribution for the estimate $\hat{\theta}$.

As a result, we see that the MSE is able to incorporate information about the bias and the variance of estimators and this is often referred to as the *Bias-Variance decomposition*. It is worth noting that in the case of an unbiased estimator being considered, the MSE is essentially the variance of the the distribution for the estimate $\hat{\theta}$. Comparisons of estimators, even those that are biased, are often

based on the MSE (Upton and Cook 2011), since lower values of the MSE closer to zero are an indication of more accurate estimates being produced.

In addition, the *Root Mean Squared Error* (RMSE) of an estimator is the square-root of the above MSE value and is given in (B.2):

$$RMSE = \sqrt{E[(\hat{\theta} - \theta)^2]}. \tag{B.2}$$

By taking the square root of the MSE, this transforms the MSE back onto the same scale as the parameter under consideration and we are able to report smaller values in comparison to the MSE values (Collins *et al.* 2001; Burton *et al.* 2006). Similarly, in the scenario where one finds themselves having to compare estimators, lower values of the RMSE closer to zero being evaluated are an indication of more accurate estimates being produced.

B.3.2 The truncated Normal distribution

Recall from Section 4.5.1, for a random variable Y to follow a normal distribution, then it is said to have probability density function

$$f(y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}$$

and we say that $Y \sim \mathcal{N}(\mu, \sigma^2)$.

It is worth noting that the *standard normal distribution* has mean 0 and variance 1. A random variable with this distribution is often denoted by Z and we write $Z \sim \mathcal{N}(0, 1)$. Its probability density function is usually denoted by ϕ and is defined as

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(z)^2}{2}}$$

for $-\infty < z < \infty$.

If $Y \sim \mathcal{N}(\mu, \sigma^2)$, then Z defined by the standardizing transformation

$$Z = \frac{Y - \mu}{\sigma},$$

has a standard normal distribution. The cumulative distribution of Z is usually denoted by Φ and tables of values of $\Phi(z)$ are commonly available (Upton and Cook 2011). These tables usually give $\Phi(z)$ only for $z > 0$, since values for negative values of z can be found using

$$\Phi(z) = 1 - \Phi(-z).$$

The tables can be used to find cumulative probabilities for $Y \sim \mathcal{N}(\mu, \sigma^2)$ via the standardizing transformation given above since, for example,

$$P(Y < y) = \Phi\left(\frac{y - \mu}{\sigma}\right).$$

In the scenario where $Y \sim \mathcal{N}(\mu, \sigma^2)$ and is defined on the interval $[a, b]$ where $-\infty \leq a < b \leq \infty$, then Y conditional on $a \leq Y \leq b$ has a *truncated* normal distribution given by:

$$g(y, \mu, \sigma) = \frac{\phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}.$$

B.4 Measures of agreement: The Kappa (κ) and Weighted Kappa (κ_w) Statistics

B.4.1 The Kappa statistic (κ)

Agreement between categorical assessments is usually considered as a problem of comparing the ability of different raters (obervers) to classify subjects into one of several groups. The approach outlined here also applies to the comparison of two alternative categorisation schemes.

For our purposes, we are interested in assessing the agreement of classifications of apparent efficacy of treatment groups, either EFF, SLOE or LOE according to our adapted classification criteria presented in Section 6.2. For this, percentage estimates and associated intervals based on utilising our developed Bayesian methodologies and those methodologies considered as part of the Defra project shall be utilised. As such, a measure of agreement rather than association is required, therefore a χ^2 test for association is not applicable - this is not a hypothesis testing problem (and also the data are paired and the classifications are ordinal).

The simplest approach to assessing agreement is to see how many exact agreements are observed. For example, Table B.1 displays the hypothetical situation of two raters, 1 and 2, rating a total of $a + b + c + d$ observations between two types of classifications, 1 and 2.

	Classification 1 by Rater 1	Classification 2 by Rater 1	(Row) Total
Classification 1 by Rater 2	a	b	a+b
Classification 2 by Rater 2	c	d	c+d
(Column) Total	a+c	b+d	a+b+c+d

Table B.1: 2x2 Contingency Table of hypothetical classifications between 2 Raters

In the instance of Table B.1 being considered, the number of exact agreements is simply $a + d$, i.e. the sum of the diagonal entries in the Table. Therefore there

would be $100\left(\frac{a+d}{a+b+c+d}\right)$ % agreement between the two raters.

There are two issues with this simple calculation however. Firstly, the calculation takes no account of where in the Table the agreement was and secondly, some agreement between the two raters would be expected by chance, even if they were guessing. A more reasonable approach would be to consider the agreement in excess of the amount of agreement that would be expected by chance and so we look to evaluate the κ statistic, i.e. the *chance corrected proportional agreement.*

The κ statistic (Cohen 1960; Cohen 1968; Fleiss *et al.* 1969) is calculated from the observed and expected frequencies on the diagonal of a square table of frequencies (Note: the expected frequency in a cell of a frequency table is the product of the total of the relevant column and the total of the relevant row, divided by the overall total of observations). If there are n observations in g categories, then the observed proportional agreement is

$$p_o = \Sigma_{i=1}^{g} \frac{f_{ii}}{n},$$

where f_{ii} is the number of agreements for category i.

The expected proportion of agreement by chance is given by

$$p_e = \Sigma_{i=1}^{g} \frac{r_i c_i}{n^2},$$

where r_i and c_i are the row and column totals for the ith category.

The index of agreement, i.e. κ, is therefore given by

$$\kappa = \frac{p_o - p_e}{1 - p_e}.$$

The κ statistic has a maximum of one when agreement is perfect, a value of zero indicates that there is no agreement better than chance and negative values (up

to the value of negative one) show worse than chance agreement (Cohen 1960).

An approximate standard error for κ presented by Cohen (1960) and Cohen (1968) is

$$se(\kappa) = \sqrt{\frac{p_o(1-p_o)}{n(1-p_e)^2}}, \tag{B.3}$$

so that a 95% confidence interval for the population value of κ is

$$\kappa \pm 1.96(se(\kappa)).$$

However, it was shown by Fleiss *et al.* (1969) that this standard error seems to overestimate the true standard error and variance, resulting in conservative estimates. As a result, Fleiss *et al.* (1969) presented valid formulae for an approximate, large sample standard error and variance of the κ statistic. We refer the reader to Fleiss *et al.* (1969) for further information. In fact, it is the formulae presented by Fleiss *et al.* (1969) that are used in the calculation of standard errors and confidence intervals in the *psych* package in *R/RStudio* (Revelle 2017).

It is worth noting however that, according to NCSS , the standard error (B.3) is often used because of its simplicity, for planning purposes and is often close to the approximation presented by Fleiss *et al.* (1969).

In general, the confidence interval for κ is not all that useful because, unless the sample size is small, the confidence interval will be narrow and thus will not allow for much variation in interpretation.

It is worth mentioning that there are no absolute definitions for interpreting values between zero and one but a general guideline provided by Landis and Koch (1977) is shown below in Table B.2.

The reduction of the data to a single number inevitably yields an answer that is not terribly meaningful without examination of the table of frequencies. There is

Value of κ	Strength of Agreement
$<$0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very Good

Table B.2: Interpreting agreement using κ

no substitute for inspecting the table of frequencies because many different tables will yield similar values of κ.

B.4.2 Weighted Kappa (κ_w)

A weakness of the κ Statistic is that it takes no account of the degree of disagreement - all disagreements are treated equally. Where the categories are ordered, as is often the case, it may be preferable to give different *weights* to disagreements according to the magnitude of the discrepancy. In this case, observations near to the diagonal, representing a difference of only one category, are considered less serious than those where the discrepancy is two or three categories.

The idea can be built into the calculation of κ to obtain a statistic called a *weighted* κ (Cohen 1968; Fleiss *et al.* 1969). This is obtained by giving weights to the frequencies in each cell of the table according to their distance from the diagonal that indicates agreement. For the cell in row i and column j, with observed frequency f_{ij} a weight can be calculated as

$$w_{ij} = 1 - \frac{|i-j|}{g-1}.$$

Thus, cells on the diagonal are given a weight of one, while those where the difference is by one category are given a weight of $1 - \frac{1}{g-1}$, etc.

The weighted observed and expected proportional agreement are obtained as

$$p_{o(w)} = \frac{\Sigma_{i=1}^{g}\Sigma_{j=1}^{g} w_{ij} f_{ij}}{n}$$

and

$$p_{e(w)} = \frac{\Sigma_{i=1}^{g}\Sigma_{j=1}^{g} w_{ij} r_i c_j}{n}.$$

As a result, the weighted κ is given by

$$\kappa_w = \frac{p_{o(w)} - p_{e(w)}}{1 - p_{e(w)}}.$$

Again, we refer the reader to Fleiss *et al.* (1969) for further information on the approximate, large sample standard error and variance for the weighted κ statistic.

As with other methods of observing small, squared frequency tables, there are difficulties associated with the use and interpretation of κ. The main issue being that the value of κ depends on the proportion of subjects in each category and the consequence of this property of κ is that it is misleading to compare values of κ from different studies where the prevalence of the categories differ. For larger tables the same is true, but it is even more complicated to judge comparability. Despite this shortcoming, the κ statistic and the weighted κ statistic, are the best approaches to the analysis of these types of problems, but it is important to display the raw data, if possible.

Bibliography

Abbott, K., M. Taylor and L. Stubbings (2009). *SCOPS Sustainable Worm Control Strategies for Sheep, 3rd Edition*. Sustainable Control of Parasites in Sheep.

AHDB (2018). About Us. `http://www.ahdb.org.uk/about/default.aspx`.

AHDB Beef and Lamb (2018). About Us. `http://beefandlamb.ahdb.org.uk/about/`.

AHDB Dairy (2018). About Us. `http://dairy.ahdb.org.uk/about-ahdb-dairy/what-is-ahdb-dairy/`.

AHVLA (2012). Lungworm in cattle. Animal Health and Veterinary Laboratories Agency Information Handout.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, 267–281.

APHA (2018). About Us. `https://www.gov.uk/government/organisations/animal-and-plant-health-agency`. Animal and Plant Health agency.

Arbenz, P. (2013). Bayesian copulae distributions, with application to operational risk management—some comments. *Methodology and Computing in Appliad Probability* **15**, 105–108.

Barrere, V., L. Falzon, K. Shakya, P. Menzies, A. Peregrine and R. Prichard (2013). Assessment of benzimidazole resistance in *Haemonchuscontortus* in sheep flocks in ontario, canada: Comparison ofdetection methods for drug resistance. *Veterinary Parasitology* **198**, 159–165.

Bartley, D., F. Jackson, E. Jackson and N. Sargison (2004). Characterisation of two triple resistant field isolates of *Teladorsagia* from scottish lowland sheep farms. *Veterinary Parasitology* **123**, 189–199.

Berry, D. and G. Ayers (2006). Symmetrized percentage change for treatment comparisons. *The American Statistician* **60**, 27–31.

Birnbaum, Z. and F. Tingey (1951). One-sided confidence contours for probability distribution functions. *The Annals of mathematical Sciences* **22/4**, 592–596.

Bishop, S. and M. Stear (2003). Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections. *Veterinary Parasitology* **115**, 147–166.

Bogale, B., M. Chanie, A. Melaku, T. Fentahun and A. Berhanu (2014). Occurrence, intensity and parasite composition of gastrointestinal helminth parasites in walia ibex (capra walie) at semien mountains national park, natural world heritage site, northern ethiopia. *Acta Parasitologica Globalis* **5**, 19–25.

Bohning, D., E. Dietz, P. Schlattmann, L. Mendonca and U. Kirchner (1999). The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **162**, 195–209.

Bolker, B. (2008). *Ecological Models and Data in R*. Princeton University Press.

Borenstein, M., L. Hedges, J. Higgins and H. Rothstein (2009). *Introduction to Meta-Analysis*. Wiley.

Brooks, S. and A. Gelman (1997). General methods for monitoring for convergenceof iterative simulations. *Journal of Computational and Graphical Statistics* **7**, 434–455.

Broyden, C. (1970). The convergence of a class of doubl-rank minimization algorithms. *Journal of teh Institute of Mathematics and its Applications* **6**, 76–90.

Burton, A., D. Altman, P. Royston and R. Holder (2006). The design of simulation studies in medical statistics. *STATISTICS IN MEDICINE* **25**, 4279–4292.

Busin, V., F. Kenyon, N. Laing, M. Denwood, D. McBean, N. Sargison and K. Ellis (2013). Addressing sustainable sheep farming: Application of a targeted selective treatment approach for anthelmintic use on a commercial farm. *Small Ruminant Research* **110**, 100–103.

Byrom, W. (1990). *Simulation Models for Investigating East Coast Fever and other Parasitic Diseaes*. Ph. D. thesis, University of Strathclyde.

Cabaret, J. and B. Berrag (2004). Faecal egg count reduction test for assessing anthelmintic efficacy: average versus individually based estimations. *Veterinary Parasitology* **121**, 105–113.

Callinan, A. L., F. Morley, J. Arundel and D. White (1982). A model of the life cycle of sheep nematodes and the epidemiology of nematodiasis in sheep. *Agricultural Systems* **9**, 199–225.

Calvete, C. and J. Uriarte (2013). Improving the detection of anthelmintic resistance: Evaluation of faecal egg count reduction test procedures suitable for farm routines. *Veterinary Parasitology* **196**, 438–452.

Campbell, W., M. Fisher, E. Stapley, G. Albers-Schonberg and T. Jacob (1983). Ivermectin: A potent new antiparasitic agent. *SCIENCE* **221**, 823–828.

Canty, A. and B. Ripley (2015, February). R package boot. requires R version 3.0.0 or above.

Carpenter, J. and J. Bithell (2000). Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in Medicine* **19**, 1141–1164.

Chang, W. (2017, December). Web application framework for r. R Package shiny, requires R version 3.0.2 or above.

Charlier, J., M. Voort, F. Kenyon, P. Skuce and J. Vercruysse (2014). Chasing helminths and their economic impact on farmed ruminants. *TRENDS in Parasitology* **30**, 361–367.

Cherubini, U., E. Luciano and W. Vecchiato (2004). *Copula methods in finance*. Wiley.

Christie, M. and F. Jackson (1982). Specific identification of strongyle eggs insmall samples of sheep faeces. *Equine Veterinary Journal* **32**, 113–117.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37–46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**, 213–220.

Coles, G., C. Bauer, F. Borgsteede, S. Geerts, T. Klei, M. Taylor and P. Waller (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. *Veterinary Parasitology* **44**, 35–44.

Coles, G., F. Jackson, W. Pomroy, R. Prichard, G. von Samson-Himmelstjerna, A. Silvestre, M. Taylor and J. Vercruysse (2006). The detection of anthelmintic resistance in nematodes of veterinary importance. *Veterinary Parasitology* **136**, 167–185.

Collins, L., J. Shafer and C. Kam (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* **6**, 330–351.

Conover, W. (1971). *Practical Nonparametric Statistics*. New York: John Wiley and Sons.

Consul, P. and F. Famoye (1989). The truncated generalized poisson distribution and its estimation. *Communications in Statistics - Theory and Methods* **18**.

Corwin, R. (1997). Economics of gastrointestinal parasitism of cattle. *Veterinary Parasitology* **72**, 451–460.

Cowles, M. and C. David (1982). On the origins of the .05 level of statistical significance. *American Psychologist* **37(5)**, 553–558.

COWS (2013). *Liver Fluke, Sustainable parasite control strategies for cattle*. Control of Worms Sustainably.

COWS (2014). Cattle Parasite Control Guide, A comprehensive list of products for the control of internal and external parasites of cattle. Information Guide. Control of Worms Sustainably.

COWS (2015a). *Control of lungworm in cattle*. Control of Worms Sustainably.

COWS (2015b). *Control of parasitic gastroenteritis in cattle*. Control of Worms Sustainably.

COWS (2018). About Us. `http://www.cattleparasites.org.uk/about.html`. Control of Worms Sustainably.

Cox, D. and D. Hinkley (1974). *Theoretical Statistics*. Chapman & Hall.

CRAN (2018). The R Project for Statistical Computing. `http://www.r-project.org/`.

Cringoli, G. (2006). Flotac, a novel apparatus for a multivalent faecal egg count technique. *Parassitologia* **48**, 381–384.

Cringoli, G., L. Rinaldi, M. Maurelli and J. Utzinger (2010). Flotac: new multivalent techniques for qualitative and quantitative copromicroscopic diagnosis of parasites in animals and humans. *Nature Protocols* **5**, 503–515.

Cringoli, G., L. Rinaldi, V. Veneziano, G. Capelli and A. Scala (2004). The influence of flotation solution, sample dilution and the choice of mcmaster slide area (volume) on the reliability of the mcmaster technique in estimating the faecal egg counts of gastrointestinal strongyles and *Dicrocoelium dendriticum* in sheep. *Veterinary Parasitology* **123**, 121–131.

Crofton, H. (1966). *Nematodes*. 178-202 Great Portland Street, London W1: Hutchison and Co (Publishers) LTD.

Dash, K., E. Hall and I. Barger (1988). The role of arithmetic and geometric worm egg counts in faecal egg count reduction test and in monitoring strategic drenching programs in sheep. *Australian Veterinary Journal* **65**, 66–68.

Defra (2015). Anthelmintic resistance in cattle in England – strategies for early detection and maintenance of efficacy in currently available products. Technical Report.

Defra (2018a). About Us. https://www.gov.uk/government/organisations/department-for-environment-food-rural-affairs/about. Department for Environment, Food and Rural affairs.

Defra (2018b). Anthelmintic resistance in cattle in England - strategies for early detection and maintenance of efficacy in currently available products. - VM0503. http://randd.defra.gov.uk/Default.aspx?Menu=Menu&Module=More&Location=None&Completed=0&ProjectID=17982. Department for Environment, Food and Rural affairs.

Demeler, J., A. V. Zeveren, N. Kleinschmidt, J. Vercruysse, J. Hoglund, R. Koopmann, J. Cabaret, E. Claerebout, M. Areskog and G. von Samson-Himmelstjerna (2009). Monitoring the efficacy of ivermectin and albendazole against gastro intestinal nematodes of cattle in northern europe. *Veterinary Parasitology* **160**, 109–115.

Denwood, M. (2010). *A Quantitative Approach to Improving the Analysis of Faecal Worm Egg Count Data*. Ph. D. thesis, University of Glasgow.

Denwood, M. (2015, March). Power calculations and bayesian analysis of count distributions and fecrt data using mcmc. R Package bayescount, requires R version 3.0 or above.

Denwood, M., S. Reid, S. Love, M. Nielsen, L. Matthews, I. McKendrick and G. Innocent (2010). Comparison of three alternative methods for analysis of equine faecal egg count reduction test data. *Preventive Veterinary Medicine* **93**, 316–323.

Denwood, M., M. Stear, L. Matthews, S. Reid, N. Toft and G. Innocent (2008). The distribution of the pathogenic nematode nematodirus battus in lambs is zero-inflated. *Parasitology* **135**, 1225–1235.

Diaz-Emparanza, I. (2002). Is a small monte carlo analysis a good analysis? checking the size, power and consistency of a simulation-based test*. *Statistical Papers* **43**, 567–577.

Dijk, J., N. D. Sargison, F. Kenyon and P. J. Skuce (2009). Climate change and infectious disease: helminthological challenges to farmed ruminants in temperate regions. *Animal* **4:3**, 377–392.

Dobson, R. and E. Barnes (1995). Interaction between *Ostertagia circumcincta* and *Haemonchus contortus* infection in young lambs. *International journal for Parasitology* **25**, 495–501.

Dobson, R., B. Hosking, C. Jacobson, J. Cotter, R. Besier, P. Stein and S. Reid (2012). Preserving new anthelmintics: A simple method for estimating faecal egg count reduction test (FECRT) confidence limits when efficacy and/or nematode aggregation is high. *Veterinary Parasitology* **186**, 79–92.

Dobson, R., N. Sangster, R. Besier and R. Woodgate (2009). Geometric means provide a biased efficacy result when conducting a faecal egg count reduction test (fecrt). *Veterinary Parasitology* **161**, 162–167.

Dossou-Gbété, S. and D. Mizère (2006). An overview of probability models for statistical modelling of count data. *Monografías del Seminario Matemático García de Galdeano* **33**, 237–244.

Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function*. Society for Industrial and Applied Mathematics SIAM.

Edmonds, M., E. Johnson and J. Edmonds (2010). Anthelmintic resistance of *Ostertagia ostertagi* and *Cooperia oncophora* to macrocyclic lactones in cattle from the western united states. *Veterinary Parasitology* **170**, 224–229.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

El-Abdellati, A., J. Charlier, P. Geldhof, B. Levecke, J. Demeler, G. von Samson-Himmelstjerna, E. Claerebout and J. Vercruysse (2010). The use of a simplified

faecal egg count reduction test for assessing anthelmintic efficacy on belgian and german cattle farms. *Veterinary Parasitology* **169**, 352–357.

EMA (2014). Reflection paper on anthelmintic resistance. `http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/04/WC500165561.pdf`. European Medicines Agency.

Eysker, M. and H. Ploeger (2000). Value of present diagnostic methods for gastrointestinal nematode infections in ruminants. *Parasitology* **120**, S109–S119.

Falzon, L., P. Menzies, K. Shakya, A. Jones-Bitton, J. Vanleeuwen, J. Avula, H. Stewart, J. Jansen, M. Taylor, J. Learmount and A.S.Peregrine (2013). Anthelmintic resistance in sheep flocks in ontario, canada. *Veterinary Parasitology* **193**, 150–162.

Fiel, C., C. Saumell, P. Steffan and E. Rodriguez (2001). Resistance of *Cooperia* to ivermectin treatments in grazing cattle of the humid pampa, argentina. *Veterinary Parasitology* **97**, 211–217.

Fisher, R. (1970). *Statistical methods for research workers* (14th ed.). Darien, Connecticut: Hafner Publishing Company.

Fleiss, J., J. Cohen and B. Everitt (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* **72**, 323–327.

Fletcher, R. (1970). A new approach to variable metric algorithms. *Computer Journal* **13(3)**, 23–26.

Fulford, A. (1994). Dispersion and bias: Can we trust geometric means? *Parasitology Today* **10**.

Gardiner, W. and G. Gettinby (1998). *EXPERIMENTAL DESIGN TECHNIQUES IN STATISTICAL PRACTICE, A Practical Software-based Approach.* HORWOOD PUBLISHING LIMITED.

Gelman, A., J. Charlin, H. Stern, D. Dunson, A. Vehtari and D. Rubin (2013). *Bayesian Data Analysis* (Third ed.). Chapman and Hall/CRC Press.

George, M., K. Paras, S. Howell and R. Kaplan (2017). Utilization of composite fecal samples for detection of anthelmintic resistance in gastrointestinal nematodes of cattle. *Veterinary Parasitology* **240**, 24–29.

Geurden, T., C. Chartier, J. Fanke, A. F. di Regalbono, D. Traversa, G. von Samson-Himmelstjerna, J. Demeler, H. Vanimisetti, D. Bartram and M. Denwood (2015). Anthelmintic resistance to ivermectin and moxidectin in gastrointestinal nematodes of cattle in europe. *International Journal for Parasitology: Drugs and Drug Resistance* **5**, 163–171.

Gilleard, J. and R. Beech (2007). Population genetics of anthelmintic resistance in parasitic nematodes. *Parasitology* **134**, 1133–1147.

Goldfarb, D. (1970). A family of variable metric updated derived by variation means. *Mathematics of Computation* **24(109)**, 23–26.

Gordon, H. and H. Whitlock (1939). A new technique for counting nematode eggs in sheep faeces. *Journal of the Council of Scientific Industrial Research* **12**, 50–52.

Greene, W. (2012). *Econometric Analysis* (International ed.). Edinburgh Gate, Harlow,Essex CM20 2JE, England: Pearson.

Hoberg, E., G. Zimmerman and J. Lichtenfells (1986). First report of *Nematodirus battus* (nematoda: *Trichostrongyloidea*) in north america: Redescription and comparison to other species. *The Helminthological Society of Washington* **53**, 80–88.

Hosmer, D. W., S. Lemeshow and S. May (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data, Second Edition.* John Wiley and Sons, Inc. Appendix 1, The Delta Method.

Hybu Cig Cymru (2018). About Us. `http://hccmpw.org.uk/about_hcc/`. Welsh Red Meat Levy.

Jackson, F. and R. Coop (2000). The development of anthelmintic resistance in sheep nematodes. *Parasitology* **120**, S95–S107.

Johnson, N., A. Kemp and S. Kotz (2005). *Univariate Discrete Distributions, Third Edition*. Wiley.

Kaplan, R. (2002). Anthelmintic resistance in nematodes of horses. *Veterinary Research* **33**, 491–507.

Kaplan, R. (2004). Drug resistance in nematodes of veterinary importance: a status report. *TRENDS in Parasitology* **20 No. 10**, 477–481.

Kochapakdee, S., V. Pandey, W. Pralomkarm, S. Choldumrongkul, W. Ngampongsai and A. Lawpetchara (1995). Anthelmintic resistance in goat in southern thailand. *Veterinary Record* **137**, 124–125.

Kostyshak, S. (2015, February). R package bootstrap. requires R version 2.10.0 or above.

Landis, J. and G. Koch (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

Learmount, J., M. Taylor, G. Smith and C. Morgan (2006). A computer model to simulate control of parasitic gastroenteritis in sheep on uk farms. *Veterinary Parasitology* **142**, 312–329.

Lee, P. (2004). *Bayesian Statistics: An Introduction* (Fourth ed.). John Wiley and Sons.

Lester, H. and J. Matthews (2013). Faecal worm egg count analysis for targeting anthemlimintic treatment in horses: Points to consider. *Equine Veterinary Journal* **46**, 139–145.

Lester, H., J. Spanton, C. Stratford, D. Bartley, E. Morgan, J. Hodgkinson, K. Coumbe, T. Mair, B. Swan, G. Lemon, R. Cookson and J. Matthews (2013). Anthelmintic efficacy against cyathostomins in horses in southern england. *Veterinary Parasitology* **197**, 189–196.

Levecke, B., R. Dobson, N. Speybroeck, J. Vercruysse and J. Charlier (2012). Novel insights in the faecal egg count reduction test for monitoring drug efficacy against gastrointestinal nematodes of veterinary importance. *Veterinary Parasitology* **188**, 391–396.

Levecke, B., L. Rinaldi, J. Charlier, M. Maurelli, A. Bosco, J. Vercruysse and G. Cringoli (2012). The bias, accuracy and precision of faecal egg count reduction test results in cattle using mcmaster, cornell-wisconsin and flotac egg counting methods. *Veterinary Parasitology* **188**, 194–199.

Linden, A. and S. Mantyniemi (2011). Using the negative binomial distribution to model overdispersion in ecological data. *Ecology* **92(7)**, 1414–1421.

Livestock and Meat Commission (2018). About Us. `https://www.lmcni.com/about-us/`. Livestock and Meat Commission for Northern Ireland.

Lunn, D., A. Thomas, N. Best and D. Spiegelhalter (2000). Winbugs - a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10(4)**, 325–337. `http://www.mrc-bsu.cam.ac.uk/software/bugs/`,.

Lyndal-Murphy, M., D. Rogers, W. Ehrlich, P. James and P. Pepper (2010). Reduced efficacy of macrocyclic lactone treatments in controlling gastrointestinal nematode infections of weaner dairy calves in subtropical eastern australia. *Veterinary Parasitology* **168**, 146–150.

Lyndal-Murphy, M., A. Swain and P. Pepper (2014). Methods to determine resistance to anthelmintics when continuing larval development occurs. *Veterinary Parasitology* **199**, 191–200.

Lyons, K. (1978). *The Biology of Helminth Parasites*. 41 Bedford Square, London, WC1B 3DQ: Edward Arnold (Publishers) Limited.

M. Jackson (2004). The humble leech's medical magic . `http://news.bbc.co.uk/1/hi/health/3858087.stm`.

M. Plummer (2008). Just Another Gibbs Sampler. `http://mcmc-jags.sourceforge.net/`.

MAFF (1986). *Manual of Veterinary Parasitological Laboratory Techniques*. Ministry of Agriculture, Fisheries and Food. Reference Book 418, 3rd edn. HMSO.

Malevergne, Y. and D. Sornette (2003). Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance* **3**, 231–250.

Marsaglia, G., W. Tsang and J. Wang (2003). Evaluating kolmogorov's distribution. *Journal of Statistical Software* **8/18**.

Mas-Coma, S., M. Valero and M. Bargues (2008). Effects of climate change on animal and zoonotic helminthiases. *Revue scientifique et technique (International Office of Epizootics)* **27**, 443–452.

Matthews, J. (2014). Anthelmintic resistance in equine nematodes. *International Journal for Parasitology: Drugs and Drug Resistance* **4**, 310–315.

McArthur, C., D. Bartley, D. Shaw and J. Matthews (2011). Assessment of ivermectin efficacy against gastrointestinal nematodes in cattle on four scottish farms. *Veterinary Record*, 169–658.

McCoy, M., L. Dawson, A. Carson and H. Edgar (2005). Parasite control in farm animals - present and future. *77th Annual Report Agricultural Research Institute for Northern Ireland*.

McKellar, Q. and F. Jackson (2004). Veterinary anthelmintics: old and new. *TRENDS in Parasitology* **20**, 456–461.

McKenna, P. (2006). Further comparison of faecal egg count reduction test procedures: Sensitivity and specificity. *New Zealand Veterinary Journal* **54**, 365–366.

Mejia, M., B. Igartua, E. Schmidt and J. Cabaret (2003). Multispecies and multiple anthelmintic resistance on cattle nematodes in a farm in argentina: the beginning of high resistance? *Vet.Res.* **34**.

Michaela, P. and R. Furrer (2014, September). Hierarchical modelling of faecal egg counts. R package eggCounts, requires coda package.

Molento, M. (2009). Parasite control in the age of drug resistance and changing agricultural practices. *Veterinary Parasitology* **163**, 229–234.

Mood, A., F. Graybill and D. Boes (1913). *Introduction to the Theory of Statistics* (Third ed.). McGraw-Hill, Inc.

Moredun (2018). Anthelmintic resistance in sheep. `http://www.moredun.org.uk/research/research-@-moredun/parasitic-worms/anthelmintic-resistance`. Moredun Research Institute.

Morgan, E., L. Cavill, G. Curry, R. Wood and E. Mitchell (2005). Effects of aggregation and sample size on composite faecal egg counts in sheep. *Veterinary Parasitology* **131**, 79–87.

Morgan, E. and J. Dijk (2012). Climate and the epidemiology of gastrointestinal nematode infections of sheep in europe. *Veterinary Parasitology* **189**, 8–14.

NCSS. Confidence intervals for kappa. NCSS Statistical Software.

Nelsen, R. (2006). *An Introduction to Copulas* (Second ed.). Springer.

Newton, R. and K. Rudestam (2012). *YOUR STATISTICAL CONSULTANT* (Second ed.). SAGE.

Newyman, J. and E. Pearson (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika* **20A(1/2)**, 9.

NOAH (2018a). About Us. `http://www.noah.co.uk/about/`. National Office for Animal Health.

NOAH (2018b). Compendium. `http://www.noahcompendium.co.uk/Compendium/Overview/-21789.html`. National Office for Animal Health.

Omura, S. (2008). Ivermectin: 25 years and still going strong (review). *International Journal of Antimicrobial Agents* **31**, 91–98.

O'Shaughnessy, J., B. Earley, J. Mee, M. Doherty, P. Crossan, D. Barrett, R. Prendiville, M. Macrelli and T. de Waal (2014). Detection of anthelmintic resistance on two irish beef research farms. *Veterinary Record*, 120–175.

Paton, G. (1983). *Mathematical Models for the Prediction and Control of Ovine parasitic Gastro-Enteritis*. Ph. D. thesis, University of Strathclyde.

Pepper, P., A. Swain and M. Lyndal-Murphy (2003). Using simulation techniques to investigate methods to determine resistance of helminths to anthelmintic treatment. *In: Proceedings MODSIM 2003 International Congress of Modelling and Simulation* **4**, 1580–1585.

Playford, M., A. Smith, S. Love, R. Besier, P. Kluver and J. Bailey (2014). Prevalence and severity of anthelmintic resistance in ovine gastrointestinal nematodes in australia. *Australian Veterinary Journal* **92**, 464–471.

Pook, J., M. Power, N. Sangster, J. Hodgson and D. Hodgson (2002). Evaluation of tests for anthelmintic resistance in cyathostomes. *Veterinary Parasitology* **106**.

Powers, K., I. Wood, J. Eckert, T. Gibson and H. Smith (1982). World association for the advancement of veterinary parasitology (W.A.A.V.P.) guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine and ovine). *Veterinary Parasitology* **10**, 265–284.

Presidente, P. (1985). Methods for detection of resistance to anthelmintics. *In: Anderson, N.,Waller, P.J. (Eds.), Resistance in Nematodes to Anthelmintic Drugs, CSIRO Division of Animal Health, Glebe, NSW, Australia*, 13–28.

Quality Meat Scotland (2018). HomePage. `http://www.qmscotland.co.uk/qms`.

R. Nuwer (2013). Worm therapy: Why parasites may be good for you. `http://www.bbc.com/future/story/20130422-feeling-ill-swallow-a-parasite`.

Razali, N. and Y. Wah (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics* **2**, 21–33.

Relf, V., H. Lester, E. Morgan, J. Hodgkinson and J. Matthews (2014). Anthelmintic efficacy on uk thoroughbred stud farms. *International Journal for Parasitology* **44**, 507–514.

Renard, B. and M. Lang (2007). Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology. *Advances in Water Resources* **30**, 897–912.

Revelle, W. (2017, May). Procedures for psychological, psychometric, and personality research. R package psych and requires R version 2.10 or above.

Rice, J. (2007). *Mathematical Statistics and Data Analysis* (Third/International ed.). Brooks/COle CENGAGE Learning.

Rigby, B., M. Stasinopoulos, G. Heller and V. Voudouris (2014). The distribution toolbox of gamlss. Booklet.

Rigby, R. and D. Stasinopoulos (2005). Generalized additive models for location, scale and shape. *Applied Statistics* **54**, 507–554.

Rinaldi, L., G. Coles, M. Maurelli, V. Musella and G. Cringoli (2011). Calibration and diagnostic accuracy of simple flotation, mcmaster and flotac for parasite egg counts in sheep. *Veterinary Parasitology* **177**, 345–352.

Rinaldi, L., M. Maurelli, V. Musella, A. Santaniello, G. Coles and G. Cringoli (2010). Flotac: An improved method for diagnosis of lungworm infections in sheep. *Veterinary Parasitology* **169**, 395–398.

Royston, P. (1982a). Algorithm as 181: The w test for normality. *Applied Statistics* **31**, 176–180.

Royston, P. (1982b). An extension of shapiro and wilk's w test for normality to large samples. *Applied Statistics* **31**, 115–124.

Royston, P. (1995). A remark on algorithm as 181: The w test for normality. *Applied Statistics* **44**, 547–551.

RUMA (2007). Responsible use of vaccines and vaccination in dairy and beef cattle production. Guidelines.

RUMA (2018). About Us. `http://www.ruma.org.uk/`. Responsible Use of Medicines in Agriculture.

Rumsey, D. J. (2016). *Statistics for Dummies* (Second ed.). Wiley.

Sangster, N. (1999). Anthelmintic resistance: past, present and future. *International Journal for Parasitology* **29**, 115–124.

Sargison, N., F. Jackson, D. Bartley and A. Moir (2005). Failure of moxidectin to control benzimidazole–, levamisole– and ivermectin-resistant *Teladorsagia circumcincta* in a sheep flock. *Veterinary Record* **156**, 105–109.

Sargison, N., D. Wilson and P. Scott (2009). Relative inefficacy of pour-on macrocyclic lactone anthelmintics treatments against *Cooperia* species in highland calves. *Veterinary Record* **164**, 603–604.

Schmidt, G. and L. Roberts (1989). *Foundations of Parasitology* (Fourth ed.). 11830 Westline Industrial Drive, St. Louis, Missouri 63146: Times Mirror/Mosby College Publishing (Division of The C.V. Mosby Company).

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6(2)**, 461–464.

SCOPS (2016). Know your Anthelmintic Groups. `http://www.scops.org.uk/anthelmintics-choosing-product.html`. Sustainable Control of Parasites in Sheep.

SCOPS (2018). Homepage. `http://www.scops.org.uk/`. Sustainable Control of Parasites in Sheep.

Shanno, D. (1970). Conditioning of quasi-newton methods for function minimization. *Math. Comp.* **24(111)**, 647–656.

Shaw, D. and A. Dobson (1995). Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S133.

Shiny (2018). Shiny from R Studio. `https://shiny.rstudio.com/`.

Skuce, P., E. Morgan, J. van Dijk and M. Mitchell (2013). Animal health aspects of adaptation to climate change: beating the heat and parasites in a warming world. *Animal*.

Smothers, C., F. Sun and A. Dayton (1999). Comparison of arithmetic and geometric means as measures of a central tendency in cattle nematode populations. *Veterinary Parasitology* **81**.

Song, P. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics* **27**, 305–320.

Spiegelhalter, D., N. Best, B. Carlin and A. Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* **64(4)**, 583–639.

Sprent, P. (1989). *Applied Nonparametric Statistical Methods*. Chapman and Hall.

Stafford, K. and G. Coles (1999). Nematode conrol practices and anthelmintic resistance in dairy calves in the south west of england. *Veterinary Record* **144**, 659–661.

Stasinopoulos, M., B. Rigby, V. Voudouris, G. Heller and F. de Bastiani (2015). Flexible regression and smoothing the gamlss packages in r. Booklet.

Stasinopoulos, M. and R. Rigby (2007). Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software* **23**.

Stear, M., M. Doligalska and K. Donskow-Schmelter (2007). Alternatives to anthelmintics for the control of nematodes in livestock. *Parasitology* **134**, 139–151.

Stratford, C., H. Lester, K. Pickles, B. McGorum and J. Matthews (2014). An investigation of anthelmintic efficacy against strongyles on equine yards in scotland. *Equine Veterinary Journal* **46**, 17–24.

Stromberg, B. (1997). Environmental factors influencing transmission. *Veterinary Parasitology* **72**, 247–264.

Sutherland, I. and D. Leathwick (2011). Anthelmintic resistance in nematode parasites of cattle: a global issue? *TRENDS in Parasitology* **27**, 176–181.

Taylor, M. (2010a). *COWS Control of Worms Sustainably*. Dairy Co and Eblex, Divisions of the Agriculture and Horticulture Development Board.

Taylor, M. (2010b). The diagnosis of parasitism in sheep. *Journal of Small Ruminant Research* **92**, 120–125.

Taylor, M. (2012). Scops and cows - 'worming it out of uk farmers'. *Veterinary Parasitology* **186**, 65–69.

Taylor, M., R. Coop and R. Wall (2007). *Veterinary Parasitology* (Third ed.). Blackwell Publishing Ediorial Offices: Blackwell Publishing.

Taylor, M., K. Hunt and K. Goodyear (2002). Review anthelmintic resistance detection methods. *Veterinary Parasitology* **103**, 183–194.

Taylor, M., J. Learmount, E. Lunn, C. Morgan and B. Craig (2009). Multiple resistance to anthelmintics in sheep nematodes and comparison of methods used for their detection. *Small Ruminant Research* **86**, 67–70.

Techion Group Ltd, Technology in Action (2018). FECPAK. `http://www.techiongroup.co.nz/`.

Torgerson, P., M. Paul and R. Furrer (2014). Evaluating faecal egg count reduction using a specifically designed package "eggcounts" in r and a user friendly web interface. *International Journal for Parasitology* **44**, 299–303.

Torgerson, P., M. Schnyder and H. Hertzberg (2005). Detection of anthelmintic resistance: a comparison of mathematical techniques. *Veterinary Parasitology* **128**, 291–298.

Traversa, D., G. von Samson-Himmelstjerna, J. Demeler, P. Milillo, S. Schürmann, H. Barnes, D. Otranto, S. Perrucci, A. F. di Regalbono, P. Beraldo, A. Boeckh and R. Cobb (2009). Anthelmintic resistance in cyathostomin populations from horse yards in italy, united kingdom and germany. *Parasites and Vectors* **2**.

Upton, G. and I. Cook (2011). *Dictionary of Statistics* (Revised ed.). Oxford.

VanderPlas, J. (2014). Frequentism and bayesianism: A python-driven primer. *PROC. OF THE 13th PYTHON IN SCIENCE CONF. (SCIPY 2014)*, 1–9.

Vercruysse, J., P. Holdsworth, T. Letonja, D. Barth, G. Conder, K. Hamamoto and K. Okano (2001). International harmonisation of anthelmintic efficacy guidelines. *Veterinary Parasitology* **96**.

Vercruysse, J. and R. Rew (2002). *Macrocyclic Lactones in Antiparasitic Therapy*. CABI Publishing.

Vidyashankar, A., B. Hanlon and R. Kaplan (2012). Statistical and biological considerations in evaluating drug efficacy in equine strongyle parasites using faecal egg count data. *Veterinary Parasitology* **185**, 45–56.

Vidyashankar, A. N., R. M. Kaplan and S. Chan (2007). Statistical approach to measure the efficacy of anthelmintic treatment on horse farms. *Parasitology* **134**, 2027–2039.

Vlassoff, A. and P. McKenna (1994). Nematode parasites of economic importance in sheep in new zealand. *New Zealand Journal of Zoology* **21**, 1–8.

VMD (2018). About Us. `https://www.gov.uk/government/organisations/veterinary-medicines-directorate`. Veterinary Medicines Directorate.

Voort, M., J. Charlier, L. Lauwers, J. Vercruysse, G. Huylenbroeck and J. Meensel (2013). Conceptual framework for analysing farm-specific economic effects of helminth infections in ruminants and control strategies. *Preventive Veterinary Medicine* **109**, 228–235.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.

Waghorn, T., D. Leathwick, A. Rhodes, R. Jackson, W. Pomroy, D. West and J. Moffat (2006). Prevalence of anthelmintic resistance on 62 beef cattle farms in the north island of new zealand. *New Zealand Veterinary Journal* **54**, 278–282.

Wak, J. V. (2001). Refugia-overlooked as perhaps the most potent factor concerning the development of anthelmintic resistance. *Onderstepoort Journal of Veterinary Research* **68**, 55–67.

Waller, P. (1997). Anthelmintic resistance. *Veterinary Parasitology* **72**, 391–412.

Waller, P. (2006). Sustainable nematode parasite control strategies for ruminant livestock by grazing management and biological control. *Animal Feed Science and Technology* **126**, 277–289.

Waller, P., O. Schwan, B. Ljungstrom, A. Rydzik and G. Yeates (2004). Evaluation of biological control of sheep parasites using *Duddingtonia flagrans* under commercial farming conditions on the island of gotland, sweden. *Veterinary Parasitology* **126**, 299–315.

Wang, C., P. Torgerson, J. Hoglund and R. Furrer (2017). Zero-inflated hierarchical models for faecal egg counts to assess anthelmintic efficacy. *Veterinary Parasitology* **235**, 20–28.

Weiss, N. and M. Hasset (1991). *Introductory Statistics* (Third ed.). Addison Wesley.

Whitlock, H. (1948). Some modifications of the mcmaster helminth egg-counting technique and apparatus. *Journal of the Council of Scientific Industrial Research* **21**, 177–180.

Wilson, K., B. Grenfell and D. Shaw (1996). Analysis of aggregated parasite distributions: a comparison of methods. *Funct. Ecol.* **10**, 592–601.

Wilson, P. (2015). The misuse of the vuong test for non-nested models to test for zero-inflation. *Economic Letters* **127**, 51–53.

Wimmer, G. and G. Altmann (1999). *Thesaurus of univariate discrete probability distributions*. Stamm.

Winer, B., D. Brown and K. Michels (1991). *STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN* (Third ed.). McGraw-Hill Inc.

Winkelmann, R. (2000). *Econometric Analysis of Count Data*. Springer.

Wolstenholme, A., I. Fairweather, R. Prichard, G. von Samson-Himmelstjerna and N. Sangster (2004). Drug resistance in veterinary helminths. *TRENDS in Parasitology* **20**, 469–476.

Wolstrup, J., J.Gronvold, S. Henriksen, P. Nansen, M. Larsen, H. Bogh and B. Ilsoe (1994). An attempt to implement the nematodetrapping fungus *Duddingtonia flagrans* in biological control of trichostrongyle infections of first year grazing calves. *Journal of Helminthology* **68**, 175–180.

Wood, I., N. Amaral, K. Bairden, J. Duncan, T. Kassai, J. J. Malone, J. Pankavich, R. Reinecke, O. Slocombe, S. Taylor and J. Vercruysse (1995). World association for the advancement of veterinary parasitology (w.a.a.v.p.) second edition of guidelines for evaluating the efficacy of anthelmintics in ruminants (bovine, ovine, caprine). *Veterinary Parasitology* **58**, 181–213.

Worm Therapy (2018). Homepage. `http://www.wormtherapy.com/`.

Yadav, P. and R. Singh (2011). A review on anthelmintic drugs and their future scope. *International Journal of Pharmacy and Pharmaceutical Sciences* **3**, 17–21.

Zar, J. (1996). *Biostatistical Analysis* (Third ed.). Prentice-Hall, Upper Saddle River, USA.

Zuur, A., E. Ieno, N. Walker, A. Saveliev and G. Smith (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA.

Zuur, A., A. Saveliev and E. Ieno (2012). *Zero Inflated Models and Gneralized Linear Mixed Models in R*. Highland Statistics Ltd.