University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2021-11

Big data and machine learning tools to understand mastitis epidemiology and other topics

Naqvi, Syed Ali

Naqvi, S. A. (2021). Big data and machine learning tools to understand mastitis epidemiology and other topics (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. http://hdl.handle.net/1880/114131

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Big data and machine learning tools to understand mastitis epidemiology and other topics

by

Syed Ali Naqvi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

NOVEMBER, 2021

© Syed Ali Naqvi 2021

ABSTRACT

Increased availability of technologies to collect and store individual health data is leading to a growing interest in applying Big Data analytical methodologies to better understand health and disease in both humans and dairy cattle. Data collected through routine observations such as doctor or veterinary visits, milking equipment, or remote sensors can be successfully incorporated to monitor and manage individual and public health, and support operational decision-making on dairy farms. These sources of data also provide an invaluable resource in conducting epidemiological and health research, provided they are appropriately handled during the statistical analysis. In this thesis, 1) data from bacteriological sampling were combined with regularly collected dairy herd improvement (DHI) data to describe udder health in primiparous dairy cattle across Canada; 2) a systematic review and meta-analysis was conducted to synthesize all available research on the effectiveness of pre-calving therapies to improve udder health in primiparous dairy cattle; 3) a model was developed for the detection of clinical mastitis (CM) based on routinely collected data from automated milking systems (AMS); 4) a simulation study assessed the impact of unmeasured heterogeneity in secondary data collected from multiple dairy farms on the performance of a model trained to detect CM onset; 5) the immune fingerprint of children presenting with symptoms of appendicitis are compared by combining emergency department admissions data with results from a multiplex cytokine assay and 6) dietary risk factors for immunological flare-ups in patients with Crohn's disease are explored by combining patient-reported dietary records with results of a multiplex cytokine assay. Chapter 2 demonstrated that the udder health in Canadian primiparous dairy cows was an issue that needed attention, and chapter 3 demonstrated that pre-calving treatments of different types can be effective at improving udder health in early lactation. Both chapters highlighted the need for

routinely collected data to be combined with targeted data collection (monitoring of non-milking dairy cows, culture-based treatment selection) to facilitate targeted management for different parts of a dairy herd. In chapter 5, a deep recurrent neural network (RNN) model was used to detect the onset of CM using regularly collected data from AMS, and chapter 6 demonstrated that predictive performance of deep RNNs is robust to the unmeasured heterogeneity in data collected from multiple farms. Chapter 6 describes how immune response differs between children with abdominal pain symptomatic of appendicitis and provides evidence that data from a multiplex immunoassay conducted on admission may be used to effectively predict disease outcomes. In chapter 7, a similar multiplex immunoassay is used to explore associations between inflammation and diet using food records from patients with Crohn's disease and demonstrates some of the statistical challenges encountered when working with multiple outcomes and large numbers of explanatory variables.

PREFACE

This thesis consists of 6 manuscripts – 4 have been published, 2 are currently under review. The following manuscripts are included in this thesis:

CHAPTER 2: Naqvi, S.A., J. De Buck, S. Dufour, and H.W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101:3233–3247. https://doi.org/10.3168/jds.2017-13579.

CHAPTER 3: Naqvi, S.A., D.B. Nobrega, P.E. Ronksley, and H.W. Barkema. 2018. Invited review: Effectiveness of precalving treatment on postcalving udder health in nulliparous dairy heifers: A systematic review and meta-analysis. J. Dairy Sci. 101:4707–4728. https://doi.org/10.3168/jds.2017-14301.

CHAPTER 4: Naqvi, S.A., M.T.M. King, R.D. Matson, T.J. DeVries, R. Deardon, and H.W. Barkema. 2021. Mastitis detection with recurrent neural networks in farms using automated milking systems. Submitted for peer review in Comput. Electron. Agric.

CHAPTER 5: Naqvi, S.A., M.T.M. King, T.J. DeVries, H.W. Barkema, and R. Deardon. 2021. Data considerations for developing deep learning models for dairy applications. Submitted for peer review in Comput. Electron. Agric.

CHAPTER 6: Naqvi, S.A., G.C. Thompson, A.R. Joffe, J. Blackwood, D.-A. Martin, M. Brindle, H.W. Barkema, and C.N. Jenne. 2019. Cytokines and chemokines in pediatric appendicitis: a multiplex analysis of inflammatory protein mediators. Mediators Inflamm. 2019. https://doi.org/10.1155/2019/2359681.

CHAPTER 7: Naqvi, S.A., L.M. Taylor, R. Panaccione, S. Ghosh, H.W. Barkema, N. Hotte, N. Shommu, S. Kaur, R.A. Reimer, K.L. Madsen, and M. Raman. 2021. Dietary patterns, food groups and nutrients in Crohn's disease: associations with gut and systemic inflammation. Sci. Rep. 11:1–9. https://doi.org/10.1038/s41598-020-80924-8.

Statement of work done

S. Ali Naqvi prepared and analyzed the data (Chapter 2). S. Ali Naqvi designed the study and search strategy and conducted all statistical analyses (Chapter 3). Diego Nobrega assisted with the eligibility screening step of the systematic review (Chapter 3). S. Ali Naqvi designed the study, prepared, and analyzed the data (Chapters 4 and 5). Meagan King and Robert Matson coordinated and executed the data collection for the larger study; Trevor DeVries oversaw the design and data collection for the larger study (Chapters 4 and 5). S. Ali Naqvi prepared and analyzed the data (Chapter 6). Graham Thompson and Craig Jenne oversaw the design and data collection for the larger studies from which secondary data were obtained (Chapter 6). S. Ali Naqvi prepared and analyzed the data, with Lorian Taylor being responsible for interpretation and context-specific writing in the manuscript (Chapter 7). Lorian Taylor and Matreyi Raman planned and coordinated the data collection for the larger study from which the secondary data were obtained (Chapter 7). Herman Barkema oversaw the design and execution (Chapter 2); Herman Barkema and Paul Ronksley oversaw the design and execution (Chapter 3); Herman Barkema and Rob Deardon oversaw the design and execution (Chapters 4 and 5); Herman Barkema, Craig Jenne and Graham Thompson oversaw the design and execution (Chapter 6); and Herman Barkema, Subrata Ghosh and Maitreyi Raman oversaw the design and execution (Chapter 7). All authors gave approval of the final versions to be published and agreed to be

accountable for all aspects of manuscripts. Permission was obtained from co-authors and the publishing journals to reprint the manuscripts as part of this thesis.

The following manuscripts that I co-authored while I was a PhD student at the University of Calgary were not included in this thesis:

Beninger, C., S.A. Naqvi, S. Naushad, K. Orsel, C. Luby, H. Derakhshani, E. Khafipour, and J. De Buck. 2018. Associations between digital dermatitis lesion grades in dairy cattle and the quantities of four *Treponema* species. Vet. Res. 49:1–15. doi: 10.1186/s13567-018-0605-z.

Corbett, C.S., S.A. Naqvi, C.A. Bauman, J. De Buck, K. Orsel, F. Uehlinger, D.F. Kelton, and H.W. Barkema. 2018. Prevalence of *Mycobacterium avium* ssp. *paratuberculosis* infections in Canadian dairy herds. J. Dairy Sci. 101:11218–11228. doi: 10.3168/jds.2018-14854.

Corbett, C.S., S.A. Naqvi, J. De Buck, U. Kanevets, J.P. Kastelic, and H.W. Barkema. 2018. Environmental sample characteristics and herd size associated with decreased herd-level prevalence of *Mycobacterium avium* ssp. *paratuberculosis*. J. Dairy Sci. 101:8092–8099. doi: 10.3168/jds.2018-14661.

Liu, G., L. Ding, B. Han, S. Piepers, S.A. Naqvi, H.W. Barkema, T. Ali, S. De Vliegher, S. Xu, and J. Gao. 2018. Characteristics of *Escherichia coli* isolated from bovine mastitis exposed to subminimum inhibitory concentrations of cefalotin or ceftazidime. Biomed Res. Int. 2018. doi: 10.1155/2018/4301628.

Narayana, S.G., F. Miglior, S.A. Naqvi, F. Malchiodi, P. Martin, and H.W. Barkema. 2018. Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models. J. Dairy Sci. 101:11120–11131. doi: 10.3168/jds.2018-15126.

Narayana, S.G., F. Schenkel, F. Miglior, T. Chud, E.A. Abdalla, S.A. Naqvi, F. Malchiodi, and H.W. Barkema. 2021. Genetic analysis of pathogen-specific intramammary infections in dairy cows. J. Dairy Sci. 104:1982–1992. doi: 10.3168/jds.2020-19062.

Naushad, S., S.A. Naqvi, D. Nobrega, C. Luby, J.P. Kastelic, H.W. Barkema, and J. De Buck. 2019. Comprehensive virulence gene profiling of bovine non-*aureus* staphylococci based on whole-genome sequencing data. mSystems 4:e00098-18. doi: 10.1128/mSystems.00098-18.

Naushad, S., D.B. Nobrega, S.A. Naqvi, H.W. Barkema, and J. De Buck. 2020. Genomic analysis of bovine *Staphylococcus aureus* isolates from milk to elucidate diversity and determine the distributions of antimicrobial and virulence genes and their association with mastitis. mSystems 5:e00063-20. doi: 10.1128/mSystems.00063-20.

Nobrega, D.B., J. De Buck, S.A. Naqvi, G. Liu, S. Naushad, V. Saini, and H.W. Barkema. 2017. Comparison of treatment records and inventory of empty drug containers to quantify antimicrobial usage in dairy herds. J. Dairy Sci. 100:9736–9745. doi:10.3168/jds.2017-13116.

Nobrega, D.B., S. Naushad, S.A. Naqvi, L.A.Z. Condas, V. Saini, J.P. Kastelic, C. Luby, J. De Buck, and H.W. Barkema. 2018. Prevalence and genetic basis of antimicrobial resistance in

non-*aureus* Staphylococci isolated from Canadian dairy herds. Front. Microbiol. 9:256. doi:10.3389/fmicb.2018.00256.

Nobrega, D.B., S.A. Naqvi, S. Dufour, R. Deardon, J.P. Kastelic, J. De Buck, and H.W. Barkema. 2020. Critically important antimicrobials are generally not needed to treat nonsevere clinical mastitis in lactating dairy cows: Results from a network meta-analysis. J. Dairy Sci. 103:10585–10603. doi:10.3168/JDS.2020-18365.

Uehlinger, F.D., S.A. Naqvi, S.J. Greenwood, J.T. McClure, G. Conboy, R. O'Handley, and H.W. Barkema. 2017. Comparison of five diagnostic tests for *Giardia duodenalis* in fecal samples from young dogs. Vet. Parasitol. 244:91–96. doi:10.1016/j.vetpar.2017.07.030.

ACKNOWLEDGEMENTS

There are many people without whom the completion of this thesis would not have been possible. First and foremost, I would like to thank my supervisors Drs. Herman Barkema and Rob Deardon. Thank you for supporting me as I flip-flopped between all the potential directions this thesis could go in, for helping me identify my strengths and weaknesses as a researcher, for helping me figure out how to thrive in different roles and for helping me make connections that will benefit me for the rest of my life.

I would also like to acknowledge the contributions of my outstanding supervisory committee members: Drs. Simon Dufour and Tyler Williamson. In addition to providing guidance and expertise for my research, I appreciated all the career and professional advice you have given me throughout my training as a researcher.

It would not have been possible to get through the more tedious parts of graduate school, without being able to laugh and commiserate with my fellow graduate students. Thank you for all your support, and for being someone I could talk to Gang, Diego, Laura, Caroline, Jesse, Emily, Casey, Jenn, Kayley, Saranya, Ellen, Larissa, Linda, and Paul.

I am eternally grateful to the support my parents and my brother have provided me before, throughout, and after graduate school. You always encouraged me to move forward, and take advantage of opportunities, while being there to support me when I failed.

Caroline, Taz, Batman and Fatty – none of this would have been possible without our little family to cheer me up, motivate me, wake me up when it was food time, and remind me to leave the house at least a few times a day. Thank you for being there for me, and I hope I have been able to return at least some of your support while in graduate school.

Lastly, I would like to acknowledge everyone that funded my research and made the required large scale data collection possible. I have been the fortunate recipient of a Queen Elizabeth II Graduate Scholarship (University of Calgary), two NSERC CREATE Milk Quality Graduate Student Scholarships (Canadian Bovine Mastitis and Milk Quality Research Network), and a Workforce Development Initiative Scholarship (Canadian Dairy Commission). Chapters 2 to 5 were also supported by the Industrial Research Chair in Infectious Diseases of Dairy Cattle, funded by Canada's Natural Sciences and Engineering Research Council (NSERC) Industrial Research Chair Program (Ottawa, ON, Canada), with industry contributions from Alberta Milk (Edmonton, AB, Canada), the Dairy Farmers of Canada (Ottawa, ON, Canada), Westgen Endowment Fund (Milner, BC, Canada), the BC Dairy Association (Burnaby, BC, Canada), Canadian Dairy Network (Guelph, ON, Canada), Lactanet (Guelph, ON, Canada), SaskMilk (Regina, SK, Canada), Dairy Farmers of Manitoba (Winnipeg, MB, Canada) and MSD Animal Health (Boxmeer, The Netherlands). Data used in chapters 4 and 5 were derived from projects supported by: the Dairy Research Cluster II Initiative (Dairy Farmers of Canada [Ottawa, ON, Canada], Agriculture and Agri-Food Canada [Ottawa, ON, Canada], the Canadian Dairy Network [Guelph, ON, Canada], and the Canadian Dairy Commission [Ottawa, ON, Canada]); the Canada First Research Excellence Fund (Ottawa, ON, Canada); an Natural Sciences and Engineering Research Council of Canada (NSERC; Ottawa, ON, Canada) Collaborative Research and Development Grant with Liquid Feeds International (Innerkip, ON, Canada), and Quality Liquid Feeds (Dodgeville, WI); the Ontario Agri-Food Innovation Alliance Research Program of the University of Guelph and the Ontario Ministry of Agriculture, Food, and Rural Affairs (Guelph, ON, Canada); the Dairy Research Cluster 3 (Dairy Farmers of Canada [Ottawa, ON, Canada] and Agriculture and Agri-Food Canada [Ottawa, ON, Canada]) under the Canadian

Agricultural Partnership Agriscience Program; and Lactanet (Sainte-Anne-de-Bellevue, QC, Canada). The research in chapter 2 would not have been possible without the participation of dairy producers, animal health technicians, and the CBMRN coordinators that collected the data; the research in chapters 4 and 5 would not have been possible without the efforts of Dr. Trevor DeVries, Dr. Meagan King, and Robert Matson in coordinating all the dairy producers, and collecting the data; the research in chapter 6 was built on the data collection efforts of everyone that was part of the Alberta Sepsis Network; and the research in chapter 7 was made possible by the efforts of everyone at Alberta's Collaboration of Excellence for Nutrition in Digestive Diseases (ASCEND).

TABLE OF CONTENTS

ABSTRACT ii

PREFACE iv

ACKNOWLEDGEMENTS ix

TABLE OF CONTENTS xii

LIST OF TABLES xvi

LIST OF FIGURES xix

LIST OF ABBREVIATIONS xxi

CHAPTER 1: General introduction 1

1.1. Big data in health research 3

1.2. Data considerations for bovine mastitis epidemiology: hierarchical data 7

1.3. Multiple outcome measures and meta-analyses 10

1.4. Machine learning models to predict clinical outcomes 12

1.5. Thesis rationale 18

1.6. Research objectives 21

1.7. References 23

CHAPTER 2: Udder health in Canadian dairy heifers during early lactation 39

2.1. Abstract 39

2.2. Introduction 41

2.3. Material and Methods 42

2.3.1. Herds and samples 42

2.3.2. Laboratory analyses 44

2.3.3. Statistical analyses 45

2.4. Results 49

2.4.1. Clinical mastitis 49

2.4.2. Intramammary infection 50

2.4.2. Subclinical mastitis 51

2.5. Discussion 51

2.6. Conclusions 56

2.7. References 57

CHAPTER 3: Effectiveness of pre-calving treatment on post-calving udder health in nulliparous dairy heifers: A systematic review and meta-analysis 74

3.1. Abstract 74

3.2. Introduction 76

3.3. Material and Methods 77

3.3.1. Data sources and literature search 77

3.3.2. Study eligibility and selection 78
3.3.3. Identification of manuscripts for eligibility in the systematic review and meta-analysis 79
3.3.4. Study quality assessment and data extraction 80
3.3.5. Statistical analyses 82
3.4. Results 83
3.4.1. Description of studies 83
3.4.2. Effectiveness of treatment 85
3.4.3. Meta-analysis statistics 87
3.5. Discussion 88
3.6 Conclusions 93
3.7. References 95
3.8. Appendices 124
CHAPTER 4: Mastitis detection with recurrent neural networks in farms using automated milking systems 129
4.1. Abstract 129
4.2. Introduction 131
4.3. Materials and Methods 134
4.3.1. Herds, animals, and data 134
4.3.1.1 Input variables 135
4.3.1.2 Outcome definition 135
4.3.1.3 Data editing and preprocessing 137
4.3.2. Statistical methodology 139
4.3.2.1 Software packages 139
4.3.2.2 Recurrent neural networks 140
4.3.2.3 Model architecture and performance metrics 141
4.3.2.4 Assessing variable importance 144
4.4. Results 145
4.4.1. Determining the optimal disease window 145
4.4.2. Identification of the most important variable subsets 146
4.4.3. External test performance and variable importance of the optimal model 147
4.5. Discussion 149
4.6. Conclusions 157
4.7. References 158
CHAPTER 5: Data considerations for developing deep learning models for dairy applications 174
5.1. Abstract 174
5.2. Introduction 176
5.3. Materials and Methods 178
5.3.1. Animals, herds, and data 178
5.3.2. Data simulation 179
5.3.2.1. Covariates 179
5.3.2.2. Disease outcomes 180
5.3.2.3. Experiments 182

5.3.3. Model development 183
5.3.3.1. Software packages 183
5.3.3.2. Recurrent neural networks 183
5.3.3.3. Model architecture and fitting 183
5.3.3.4. Model comparison and performance metrics 185
5.4. Results 186
5.4.1. Covariate simulation 186
5.4.2. Experiment 1 – effects of unmeasured farm-level variability 186
5.4.3. Experiment 2 – effects of unmeasured cow-level variability 187
5.4.4. Experiment 3 – effects of outcome misclassification 188
5.4.5. Experiment 4 – effects of missing observations in a predictor variable 188
5.4.6. Baseline scenarios 189
5.5. Discussion 189
5.6. Conclusions 195
5.7. References 196
CHAPTER 6: Cytokines and chemokines in pediatric appendicitis: A multiplex analysis of inflammatory protein mediators 213
6.1. Abstract 213
6.2. Introduction 215
6.3. Materials and Methods 217
6.3.1. Study Design 217
6.3.2. Study setting and population 217
6.3.3. Care of the patient 218
6.3.4 Inflammatory protein mediator profiling 219
6.3.5 Measures 220
6.3.6 Statistical analyses 220
6.3.6.1 Regression models 221
6.3.6.2 Decision tree 222
6.4. Results 222
6.4.1. Appendicitis vs non-appendicitis abdominal pain vs non-appendicitis sepsis 223
6.4.2. Appendicitis severity assessment 223
6.4.3. Expansive inflammatory-related protein mediator profiling (heat mapping) 223
6.4.4 Protein mediator decision tree analysis 224
6.4.5 Assessment of current common appendicitis evaluations 225
6.5. Discussion 225
6.6. Limitations 230
6.7. Conclusions 231
6.8. References 232
6.9. Appendices 250
CHAPTER 7: Dietary patterns, food groups and nutrients in Crohn's disease: associations with gut and systemic inflammation 260
7.1. Abstract 260
7.2. Introduction 261

7.3. Materials and Methods **262**
7.3.1. Study design 262
7.3.2. Inclusion/exclusion criteria 263
7.3.3. Data collection 263
7.3.4. Statistical analyses 264
7.3.5. Ethical approval 266
7.4. Results **266**
7.4.1. Demographics 266
7.4.2. Dietary pattern and macro- and micro-nutrients 266
7.4.3. Fecal calprotectin 267
7.4.4. C-reactive protein 267
7.4.5. Serum cytokines 267
7.5. Discussion **267**
7.6. References **272**
CHAPTER 8: Summarizing discussion **285**
8.1. Concluding remarks **285**
8.1.1. Mastitis management in primiparous dairy cattle 286
8.1.2. Clinical mastitis detection using recurrent neural networks 287
8.1.3. Characterizing immune response using multiplex immunoassays 289
8.1.4. Comparing the capabilities of different analytical approaches 290
8.2 Future directions **292**
8.3 References **294**
APPENDIX: Copyright Permissions **297**

LIST OF TABLES

Table 2-1. Distribution of pathogens in clinical mastitis samples from 75 dairy farms in 4 Canadian regions during the first 30 DIM for primiparous (n=197) and multiparous cattle (n=485). 62

Table 2-2. Incidence rate of clinical mastitis (IRCM; cases per 100 cow-months at risk) in the first 30 DIM of heifers in Canadian dairy herds. 63

Table 2-3. Mean incidence rate of clinical mastitis (cases per 100 cow¬-days at risk) in the first 30 DIM of heifers and multiparous cows in dairy herds from 4 regions across Canada, stratified by pathogen. 64

Table 2-4. Mean incidence rate of clinical mastitis (per 100 cow-months at risk) in heifers and multiparous cows in the first 30 DIM in Canada stratified by herd bulk milk somatic cell count (BMSCC) and pathogen category. 65

Table 2-5. Distribution of mastitis pathogens in milk from clinically normal, randomly selected primiparous (n=2,265) and multiparous (n=6,066) lactating cows from 91 dairy farms in 4 Canadian regions in the first 30 DIM. 66

Table 2-6. Quarter-level prevalence (%) of intramammary infection in heifers and multiparous cows in 4 regions across Canada in the first 30 DIM stratified by pathogen group. 67

Table 2-7. Overall and pathogen-specific prevalence (%) of quarter-level intramammary infection in first-lactation heifers in the first 30 DIM stratified by region and bulk milk SCC (BMSCC) from 91 Canadian dairy herds. 68

Table 2-8. Quarter-level prevalence of subclinical mastitis (SCC > 200,000 cells/mL) in herds across 4 regions in Canada. 69

Table 3-1. Study characteristics of the 62 included studies. 108

Table 3-2. Risk of bias as assessed by SYRCLE guidelines (Hooijmans et al. 2013) from full-text manuscripts included in systematic review (n=40). 113

Table 3-3. Outcome definitions in full-text manuscripts included in the systematic review (n=40). 115

Table 3-4. Meta-regression of records included in a systematic review pre-calving therapies for improving udder health in primiparous dairy heifers on the outcome used to assess udder health. 117

Table 3-5. Meta-regression of records included in a systematic review pre-calving treatment for improving udder health in primiparous dairy heifers on the type of treatment assessed in the study. 118

Table 3-6. Meta-regression of records included in a systematic review pre-calving treatment for improving udder health in primiparous dairy heifers on the type of treatment assessed in the study, stratified by type of pathogen. 119

Table 4-1. All variables, their type and source included as inputs for a recurrent neural network to predict the occurrence of clinical mastitis in dairy cattle. For variables that were not measured on all farms, the number of farms where they were measured is indicated. 166

Table 4-2. Mean (range) performance metrics for the detection of CM using all variables recorded by AMS and their daily variances over a 10-fold cross-validation comparing different windows of time before and after CM diagnosis. 167

Table 4-3. Mean (range) performance metrics and number of epochs required for model training over a 10-fold cross-validation using different subsets of variables for the detection of CM within 3 d prior to diagnosis..... 168

Table 4-4. Mean (range) performance metrics for the detection of CM using all variables recorded by AMS and their daily variances over a 10-fold cross-validation with a 3d sick window prior to CM diagnosis. 169

Table 5-1. Sampling distributions for the noise added to each variable when sampling with replacement from observed milking records from 37 herds, and 4,600 cow-lactations. 204

Table 5-2. Sampling distributions used for the predictor variables in equation (1) and the functions used to compute their contribution to the simulated log odds of mastitis diagnosis. . 205

Table 5-3. Varying parameters for the simulation experiments to determine the effects of the parameters on model performance. SD = standard deviation; *j* = farm identifier; *k* = cow identifier. 206

Table 5-4. Mean (IQR) of predictor variables from the original data, and the simulated data created by sampling and perturbing the original data. Means were first calculated for each herd or cow-lactation over the entire observation period, and the reported mean (IQR) represents the unweighted aggregation over all herds or cow-lactations between December 2018 and August 2019..... 207

Table 5-5. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of farm-level random effects, and two different splits of training and valida tion data. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. σ = standard deviation of farm-level random effects where the random effects follow a skewed normal distribution $\mathcal{N}0, \sigma, -1.5$. 208

Table 5-6. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of cow-level random effects, and two different splits of training and validation data. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. σ = standard deviation of cow-level random effects where the random effects follow a skewed normal distribution $\mathcal{N}0, \sigma, -1.5$. 209

Table 5-7. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of misclassification, where randomly selected mastitis cases were recoded to healthy in either the training/validation set or the test set. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. FNR = false-negative rate (proportion of clinical mastitis cases incorrectly recorded as healthy). 210

Table 5-8. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets when a highly predictive variable (SCC) is excluded from either the training/validation or test data by replacing all values

with 0. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. 211

Table 5-9. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in 3 different baseline scenarios: 1) completely random association; 2) realistic, moderate noise and 3) realistic, high noise. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. 212

Table 6-1. Characteristics of children in control, appendicitis and sepsis cohorts. 242

Table 6-2. Inflammatory protein mediators and their test characteristics in children with suspected appendicitis. Area under receiver operator characteristic curve and other test characteristics calculations are based on classification of patients with non-appendicitis abdominal pain and appendicitis (non-appendicitis sepsis group is excluded). 243

Table 6-3. Operating characteristics (95% CI) of patient categorization based on the decision tree..... 244

Table 7-1. Patient demographics and health characteristics at baseline and 3-month follow-up 279

Table 7-2. Description of dietary pattern variables in the mixed-stepwise selection process for the regression models..... 280

Table 7-3. Description of macro- and micro-nutrients and single dietary components included in a mixed-stepwise selection process for regression models. 281

Table 7-4. Final multivariate model results for fecal calprotectin, C-reactive protein, and pro- and anti-inflammatory cytokines. 282

LIST OF FIGURES

Figure 2-1. Incidence rate of clinical mastitis (IRCM) in the first 14 days of lactation in primiparous (18,157 quarter-months) and multiparous (35,936 quarter-months) dairy cows in Canada. Confidence intervals based on marginal models using generalised estimating equations. 70

Figure 2-2. Within-herd comparison of incidence rate of clinical mastitis (IRCM) in the first 30 DIM between primiparous and multiparous cows in 75 Canadian dairy herds. 71

Figure 2-3. Within-herd comparison of the prevalence of intramammary infection (IMI) in the first 30 DIM between primiparous and multiparous cows in 75 Canadian dairy herds ($R^2 = 0.20$). 72

Figure 2-4. Within-herd comparison of the prevalence of subclinical mastitis (SCM) in the first 30 DIM between primiparous and multiparous cows in 91 Canadian dairy herds ($R^2 = 0.45$). ... 73

Figure 3-1. Flow diagram of the manuscript selection process. 120

Figure 3-2. Forest plot of effect size estimates from all (n = 48) studies included in the meta-analysis assessing effectiveness of pre-calving therapy at improving post-calving udder health in dairy heifers. Size of boxes reflect relative size of effect estimate; diamonds represent pooled estimates within strata. 121

Figure 3-3. Egger's regression for small-study effects in a meta-analysis of studies assessing effectiveness of pre-calving therapies at improving udder health in dairy heifers plotted as the natural log of the risk ratio from a given study against the estimate of variance from the same study. Vertical bars indicate 95% confidence intervals at the intercept (variance = 0; largest possible studies). 122

Figure 3-4. Funnel plot to assess asymmetry and publication bias in meta-analysis of studies assessing effectiveness of pre-calving therapies at improving udder health in dairy heifers. Relative risk = 1 indicates equal risk of udder disease in treated and untreated animals (no effect of treatment). Darkest triangle indicates area of 90% confidence, light grey triangle indicates area of 95% confidence and white triangle indicates area of 99% confidence. 123

Figure 4-1. Mean ± range sensitivity (proportion of animals with CM, or animals that would be diagnosed with CM identified by the model) relative to the day of diagnosis for the detection of CM in dairy cows using a RNN. Different plots (a, b, c, d) represent the 4 windows of time where the animal was considered to be CM-positive for model training. 170

Figure 4-2. Mean sensitivity (proportion of CM animals, or animals that would be diagnosed with CM identified by the model) relative to the day of diagnosis for the detection of CM in dairy cows using a RNN. Different colours represent models trained using various subsets of recorded variables. 171

Figure 4-3. Mean sensitivity for the detection of cows with severe CM, as defined by an increase over the baseline in average milk temperature during the sick period. The different lines represent the different datasets; relative size of points reflect the number of animals identified as having severe CM. 172

Figure 4-4. Top 20 variables as defined by the absolute value of the gradient, averaged over different subsets of observations for a RNN detecting the onset of CM. 173

Figure 6-1. Selection of study candidates and final illness categories. 245

Figure 6-2. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category. Outliers (values > 95th percentile) are not presented but were included in statistical comparisons using multivariate normal regression followed by least-square means post-hoc pairwise contrasts using the Bonferroni correction. 246

Figure 6-3. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category and appendicitis severity. Outliers (values > 95th percentile) are not presented in the plots but were included in statistical comparisons conducted using multivariate normal regression followed by least-square means post-hoc pairwise contrasts using the Bonferroni correction. 247

Figure 6-4. Heatmap showing relative change over mean concentration in patients without appendicitis of 54 selected cytokines grouped by patient category, appendicitis severity and class of cytokine based on 196 pediatric patients. 248

Figure 6-5. Decision tree for classifying patients into different categories or severities of appendicitis based on concentrations of 4 informative cytokines. Thresholds presented are blood concentrations of the cytokines Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for all others. Numbers below category refer to how many individuals in each terminal node have been classified into each category: green = no appendicitis; orange = appendicitis; red = severe appendicitis; blue = non-appendicitis sepsis. 249

Figure 7-1. Fecal calprotectin ≤ 100 μg/mg is significantly related to leafy greens intake. 283

Figure 7-2. C-reactive protein ≤ 5 mg/L is significantly related to ratio of omega-6 to omega-3 polyunsaturated fat. 284

LIST OF ABBREVIATIONS

AM	antimicrobial
AMS	automated milking system
BMSCC	bulk milk somatic cell count
BTSCC	bulk tank somatic cell count
BW	body weight
CAD	Canadian dollar
CBMRN	Canadian Bovine Mastitis Research Network
CFU	colony-forming unit
CI	confidence interval
CM	clinical mastitis
CNS	coagulase-negative staphylococci
CRP	c-reactive protein
d	day
DHI	dairy herd improvement
DIM	days in milk
EC	electrical conductivity
FCP	fecal calprotectin
FE	fixed effects
GEE	generalized estimating equation
GL(M)M	generalized linear (mixed) model
ICC	intraclass correlation coefficient
IFN	interferon
IL	interleukin
IMI	intramammary infection
IQR	interquartile range
IRCM	incidence rate of clinical mastitis
LSTM	long short-term memory
LR	likelihood ratio
MCP	monocyte chemoattractant protein
ML	machine learning
MY	milk yield
NAS	non-*aureus* staphylococci
NMC	National Mastitis Council
OR	odds ratio
PCT	procalcitonin
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RE	random effects
RNN	recurrent neural network

RR	relative risk
SAA	serum amyloid A
SCC	somatic cell count
SCM	subclinical mastitis
SE	standard error
SYRCLE	Systematic Review Centre for Laboratory Animal Experimentation
TD	teat dip
TNF	tumor necrosis factor
TS	teat sealant
VAC	vaccine
yr	year

CHAPTER 1: General introduction

Recent developments in computational and data storage technologies, combined with increasing availability of high-throughput and remote sensors have created opportunities to apply big data analytics to all fields of scientific research. There are numerous definitions for big data, but there are 3 important defining factors: 1) information that is characterized by high volume, high velocity of collection and variety in sources of data; 2) specialized technology and methods are needed to appropriately analyze the data; and 3) new types of scientific or economic insights can be generated by applying the appropriate methods to these data (De Mauro et al., 2016).

Not only have technological advances facilitated efficient data collection and storage, but the increased availability of high-performance computational resources also allows for the application of computationally intensive statistical and machine learning methods even on personal computers. With such large volumes of data, it becomes even more important to ensure that appropriate methods are used for inferential analysis. Datasets with increasing numbers of samples greatly increase statistical power, such that associations or differences that are not practically different will still show statistical significance. On the other hand, when examining increasingly wide datasets with large numbers of covariates, a higher false-discovery rate can be expected, and must be accounted for when making multiple comparisons using the same dataset. Proponents of big data analytics have claimed (with criticism) that two of the benefits are that since every data point can theoretically be captured, statistical sampling techniques are not necessary, and that appropriate statistical model specification is less important since with enough samples; the observed values alone are sufficient to model the data (Harford, 2014). When analysis is focused on developing models with good predictive performance, big data proponents

also claim that understanding the underlying causal structures is unnecessary as long as models consistently perform as expected (Harford, 2014). Although it is possible to develop well-performing predictive models without a thorough understanding of the relationship between variables, it is then difficult to understand why these models fail, when they do. Appropriate and effective inference from big data therefore should still adhere to the same level of rigor in selecting analysis methods as is typically expected in scientific research.

Dairy research is a field that has always used large datasets from varied sources, but recent technological advances and increased availability of animal-monitoring sensors have led to the considerable development in the field of precision dairy farming to facilitate operational decision-making (Barkema et al., 2015; Lokhorst et al., 2019). The study of human health and epidemiology has also benefited from the increased availability of large datasets, digitization of health records, and sensors that allow for constant monitoring of patients (Mooney et al., 2015). These fields of study also present an intersection of two, possibly conflicting goals in big data analytics: understanding and defining relationships between variables in the observed data and developing high-performance predictive models. Developing predictive models can be a useful tool for physicians or producers to identify and diagnose sick individuals, but having a complete understanding of the underlying mechanisms of disease development is just as important for developing early preventive methods and interventions. In all application of big data analytics, robust inference and reliable predictions require that appropriate methods are chosen by careful consideration of the structure and limitations in the data (the variety, velocity, and volume), as well as an understanding of what kind of insights and value the data could provide.

1.1. Big data in health research

This section aims to highlight recent trends in the use of big data for research in human health and animal health on dairy farms, while also considering some of the practical issues and challenges that must be considered when using this data.

Much of the data that can be effectively used for health research are not collected specifically for the purpose of research, rather they are collected as routine operational support in clinical and production settings, or for individual health monitoring and management. These may be broadly classified into two categories: 1) regularly collected data about health and disease treatment such as electronic health records, or on-farm production and treatment records; and 2) data from sensors designed specifically to monitor physiological and health-related measures such as pedometers, wearable EKG monitors, and glucose monitoring systems for humans, or activity and rumination monitors, rumen thermometers, and RFID-based location sensors in dairy cattle.

Electronic health records in the human health setting can include information about patients' demographics, visit and admission dates, diagnoses and physician's notes, patient-reported outcomes, medical prescriptions, and diagnostics. Records that were initially by-products of the health care system, have now become an important tool for research, and the ease with which datasets from different geographic locations and healthcare systems can be linked to facilitate extremely large epidemiological studies (Ehrenstein et al., 2017). As an example, a recent study on the epidemiology of hypertension and diabetes conducted using electronic medical records from one health system in Rio de Janeiro was able to include data from 2.1 million distinct individuals, allowing for extremely precise estimates of age- and sex-stratified prevalence of the two diseases (Pinto and Santos, 2020). Another study assessing rates of hip and knee

replacements in patients with rheumatoid arthritis included over 27,000 individuals by using data from the UK's single healthcare system (Hawley et al., 2020). Similarly, large sample sizes are seen in the field of dairy research, where farms can collect milk production and constituent data for each animal multiple times per day, along with detailed records of health events and administered medications. Outside of the research context, these routinely collected herd records allow producers to monitor productivity and improve management by developing an understanding of how their herd changes over time, and compares with others (Warner et al., 2020b). An example of a research application of these records was done in Austria where legislation requires records of antimicrobial use. Over a 1-year period, researchers were able to analyze records from 248 herds to accurately describe the reason for use and dose of antimicrobials (Firth et al., 2017). In addition to on-farm recording, dairy producers may also participate in dairy herd improvement (DHI) recording where technicians visit the herd monthly to record information and take milk samples from all animals in a herd. These DHI records have long been used in epidemiological studies (Hueston et al., 1987; Curtis et al., 1988; Hueston, 1988), and often combine information from regularly collected DHI data with additional study-specific sampling (Reyher et al., 2011; Dufour et al., 2012; Warner et al., 2020a; Pralle et al., 2021) again demonstrating the effectiveness of linking large and diverse datasets.

Data from sensors, whether on humans or dairy cattle can provide continuous, real-time data to facilitate effective monitoring of health, and enable timely decision-making for interventions. The earliest use of remote sensors for health monitoring in dairy cattle was for estrus detection (Senger, 1994), but recent research has also focused on their potential to detect numerous other diseases such as mastitis, lameness, and post-partum diseases (Neethirajan, 2020). Specifically, on-cow sensors have shown potential for the detection of diseases monitor feeding, rumination,

lying, standing and walking time, temperature, rumen pH, and the cows location (Eckelkamp, 2019). The other type of sensor commonly seen on dairy farms measures characteristics of the milk, and these measurements tend to be included with production records. Outside of the research context, the findings from these sensors can also be integrated, and interpreted to facilitate real-time monitoring of animals on dairy farms through the use of interactive dashboards (Snijders et al., 2019). In addition to detecting disease onset in dairy cattle (Steensels et al., 2016), the same sensors have also been used to assess the effect of health problems on cow behaviour and production (Steensels et al., 2017). Health research using on-person sensors on the other hand is less common, likely due to concerns about privacy and most wearable devices tend to be for personal health monitoring and are not validated for medical applications (Brodie et al., 2018). Although there have been considerable advances in development of hardware for these sensors (Majumder et al., 2017), their integration with commonly used medical and health management software to provide real-time diagnostics and measurements to aid practitioners in decision-making is a more nascent area of research (Yin and Jha, 2017).

Despite the potential of large and varied datasets, there are some challenges that need to be considered even before the statistical analysis stage of research, that relate directly to the 3 V's of big data: volume, variety, and velocity. In the human health context, where data can be generated by individuals at a large scale, these individuals represent an important stakeholder, and the amount of data available for everyone makes it more important than ever to ensure that properly informed consent is acquired. With increasing numbers of patients for whom data are available, acquiring informed consent will scale with the number of study participants presenting a challenge since, informed consent was not originally intended for data of this type (Vayena and Blasimme, 2018). In addition, as patients and study participants are more thoroughly informed,

they may be motivated to monetize their data, presenting a paradigm shift where research objectives may begin to be defined by participants rather than the researchers (Vayena and Blasimme, 2018). Another issue arising from the volume of data generated concerns how the data will be stored, how the databases will be managed, and how access will be controlled since database administrators would not likely be the ones designing and conducting the research (Wang et al., 2019b). The last issue arising from the volume of data generated concerns systematic error, as it would also be magnified in larger datasets, requiring that clinical knowledge be more rigorously applied in designing studies, and that hypotheses and objectives be clearly laid out prior to starting the research (Ehrenstein et al., 2017). The variety of sources used to create large datasets allows for the inclusion of many more variables of different types, but when working with secondary data measurement error can have a considerable impact on the quality of data and the inferences that can be made (Mooney et al., 2015). In the dairy setting, there are often multiple sensors that measure the same variable such as activity or electrical conductivity of milk, but before combining data from these sensors and treating them equally, it is necessary to ensure that they have been independently validated beforehand, and to account for any differences that exist. When considering health monitoring software on mobile devices, another source of variability to consider is bias that may be included on the part of the software programmers in the way measurements are interpreted and the results presented to users (Brodie et al., 2018). The last challenge of using big data for research and applying it in practical settings relates to the velocity with which these data are generated. High velocity data can allow for better monitoring of health of individuals and production of dairy cattle, which can lead to more timely decision-making for interventions. When using these high velocity data to decide on interventions, it is, however, important to remember that the evidence provided by these data is

limited by the variables that have been measured. In the dairy setting for example, health and productivity of the animals may not directly correlate with their welfare, and it is necessary to consider the welfare of the animals beyond the measurements recorded in the datasets (Buller et al., 2020). This is less of a concern in the human setting however, since interventions would never be given without properly informing the patient and subsequently monitoring their feedback.

The following sections will consider challenges in the statistical analysis and inference stages of big data analytics.

1.2. Data considerations for bovine mastitis epidemiology: hierarchical data

Mastitis is considered the most important disease of dairy cattle (Ruegg, 2003, 2017), and is the most expensive diseases on dairy farms due to costs associated with management and the resultant losses in production (Halasa et al., 2007; Aghamohammadi et al., 2018). Research on bovine mastitis in dairy herds has focused on disease etiology, risk factors to devise control strategies, therapeutic effectiveness, impacts on production and future health and economic impacts (Halasa et al., 2007; De Vliegher et al., 2012a; Jamali et al., 2018; Nobrega et al., 2020). Studies on mastitis epidemiology in dairy cattle often combine data from various sources such as: DHI data and bacteriological culture results (Reyher et al., 2011), production records and data from activity monitors (King et al., 2018), and treatment records and garbage can inventories (Nobrega et al., 2017). Regardless of the sources of data, the type of explanatory variables or the specific outcome of interest, validity of inference and conclusions requires that the hierarchical nature of the data be appropriately accounted for in analysis.

The hierarchical structure of data used for dairy epidemiological studies results in clustering of observations such that the usual assumption of statistical independence between observations is not maintained (Dohoo et al., 2014a). This clustering can be caused by animals under observations sharing a common environment (animals from the same farm, animals in the same pens), having multiple units of analysis within a single animal (four quarters within an individual cow), or taking multiple measurements over time from the same region, farm, or animal. Because observations within a cluster (cows within a farm, quarters within a cow) are more similar to each other than they are to observations from other clusters, it is necessary that appropriate statistical inference disentangles cluster-specific effects from the effects of explanatory variables which are of primary interest. The failure to do so will typically result in the variance for model coefficients being underestimated because the estimation procedure is based on a distributional assumption of independence of observations. However, mismatches in the hierarchical levels of explanatory variables and outcomes can also inflate the variance of statistical estimates if within-cluster variability of the outcome is incorrectly attributed to an observed cluster-level explanatory variable (Ukoumunne et al., 2002).

Epidemiological studies of bovine mastitis commonly use mixed effects models to account for within-cluster observations at different levels (De Vliegher et al., 2004; Piepers et al., 2011; Levison et al., 2016). In addition to the explanatory variables of interest, mixed effects models incorporate cluster-specific random effects, that are applied to all members and represent the unmeasured variability attributed to each cluster. Random effects can also capture cluster-specific variability associated with particular explanatory variables, increasing their flexibility and the underlying complexity they are able to capture. The fixed part of the mixed effects model represents the average of the cluster-specific estimates of the associations with the explanatory

variables, where the coefficients and variance estimates are then adjusted to account for the random effects terms (Dohoo et al., 2014b). These models are typically fit using iteratively reweighted least-squares regression, restricted maximum likelihood, maximum likelihood, or Bayesian estimation. Due to the added complexity when the outcome being modeled is not continuous, the likelihood function needs to be approximated using methods such as numerical quadrature when using a maximum likelihood-based estimator, which can make fitting these models very computationally expensive. A limitation of most software implementations of mixed models is that random effects are generally assumed to be normally distributed, and a limitation specific for non-continuous outcomes is that it may be very difficult to specify a particular correlation structure to describe how observations within a cluster relate to one another. Generalized estimating equations are an alternative to mixed models, which allow for the specification of a between-observation correlation structure, which is subsequently used to fit a model that is averaged over all clusters (Dohoo et al., 2014c). Although the correlation structures for generalized estimating equations are typically limited to only a single level of clustering, as long as the highest level of clustering is described, the resulting estimates of regression coefficients and standard errors will be unbiased (Masaoud and Stryhn, 2010). Time-to-event models are also to evaluate outcomes such as culling, time to mastitis onset, time to cure after treatment, and time to mastitis recurrence. Hierarchically structured time-to-event data can be modelled using Cox proportional hazards models with frailty terms (De Vliegher et al., 2005; Van den Borne et al., 2011; Ampe et al., 2012), or accelerated failure time models with random effects. Nested frailty Cox models can also be fit as mixed effects models where the outcome is Poisson distributed, and the clustering is captured by random effects terms (Schukken et al., 2010; Elghafghuf et al., 2014).

Many observational and experimental studies of mastitis epidemiology have already incorporated data from multiple farms, and as datasets continue to become more easily linked this trend will only continue. To continue extracting robust and valid insights from these data researchers must ensure that the underlying hierarchical structures are considered by using the methods described in this section.

1.3. Multiple outcome measures and meta-analyses

Health research has benefitted not only from increased availability of data storage, retrieval and linking methodologies, but also from technologies used to collect these data. Technologies that enable simultaneous identification and measurement of multiple types of biologically relevant compounds and micro-organisms, along with increasing ease with which study subjects can monitor and record their own outcomes, have made it possible for researchers to define increasingly complex and multifactorial health outcomes.

Multiplex assays of cytokines have been used in studies to assess immune response to many different diseases in different species. Differential expression of cytokines has been associated with thyroid disease (Linkov et al., 2008), COVID-19 severity (Del Valle et al., 2020), cats with septic shock (Troia et al., 2020), and chronic lameness in dairy cattle (Herzberg et al., 2020). Although selection of cytokines for multiplex assays typically needs to be species-specific (Lesueur et al., 2019), immunoassays for disease specific antibodies and toxins have also been used for potential detection of *Staphylococcus aureus* mastitis (Abril et al., 2020) and Johne's disease (Li et al., 2017) in dairy cattle. When so many different outcomes are measured and analyzed simultaneously, the issue of multiple comparisons arises where the likelihood of

making a type I error increases beyond the accepted threshold since multiple tests are being conducted using the same data. This occurs in other big data contexts as well, where increasingly wide datasets (number of predictors > number of samples) are common, and it is important to avoid these spurious associations. One approach is to use univariate analyses for each outcome separately, and then apply a post-hoc correction to each P-value to control for the inflation of the type I error rate (Troia et al., 2020), although some studies fail to account for the multiple comparisons entirely (Del Valle et al., 2020; Herzberg et al., 2020). Another approach is to model the outcomes using a multivariate distribution after assessing the underlying assumptions, such that all the comparisons of interest are made within a single statistical model that is able to account for between-outcome correlations (Linkov et al., 2008). Although developed for gene expression data, a mixed effects model approach that separates within- and between-subject variability for each outcome can easily be applied to other settings with multiple outcomes (Liquet et al., 2012).

One area where the impact of studies reporting on multiple outcome measures has grown is in meta-analyses of study findings. As the number of journals indexed online continues to grow (Basu, 2010; Collazo-Reyes, 2013), it is likely that the number of studies included in systematic reviews and meta-analyses will also continue to grow. With increasing numbers of studies, and multiple outcomes within studies, meta-analyses will require that the researcher accurately identify different sources of heterogeneity in study findings and incorporate this knowledge into their inference. Understanding the sources of heterogeneity is an important aspect of conducting meta-analyses, and methods such as stratified meta-analysis and meta-regression can be used to assess whether reported study characteristics are contributing to the underlying heterogeneity (Thompson and Sharp, 1999). If even after the use of these methods, there is considerable

heterogeneity, the meta-analysis should be conducted using a random effects model, which quantifies the degree to which unmeasured study characteristics contribute to the residual heterogeneity in the results. The random effects meta-analysis model is an application of the mixed-effects models described in section 1.2, where the study is the hierarchical level at which clustering is expected to occur. The estimates for study-level variance and residual variance are then used to compute the tau statistic, which is comparable to the intra-class correlation coefficient in epidemiological studies. In the absence of any unmeasured, study-specific heterogeneity, a random-effects meta-analysis model reduces to a fixed-effects model, leading to the recommendation that random-effects models should always be used by default (Hunter and Schmidt, 2000). These random-effects models can also be fit in a Bayesian framework, allowing for the further incorporation of study-specific uncertainty in outcome recording, providing a flexible way to model the same data while making fewer assumptions (Higgins et al., 2009). Recent interest in developing methods to update meta-analysis results by automating the inclusion of new publications (Tsafnat et al., 2014; Reddy et al., 2020) means that it is even more important that robust and appropriate statistical methods are chosen for analysis from the start.

1.4. Machine learning models to predict clinical outcomes

The last area of health research that has benefitted greatly from the advancements in data storage and retrieval, as well as computational technologies has been in the development of clinical prediction models. Increased availability of large datasets has resulted in more data being used to train and validate models, while increased access to computational resources have made it possible to train more complex and resource-intensive models than was previously possible.

There has been growing interest in developing models to predict disease onset, and response to treatment in both human health, as well as in the dairy industry.

The specific type of data used for model development is an important consideration when determining what the goal of the prediction model would be. As an example, electronic health records or clinical assessments collected infrequently and at irregular intervals are best suited for models that facilitate decision-making regarding prognosis, treatment and support, whereas high-frequency sensor data, or diagnostic images are better suited for detection and diagnosis of disease (Shatte et al., 2019). When compared against common clinical risk scoring, machine learning methods using electronic health records were better able to predict strokes (Wang et al., 2020; Lip et al., 2021) and allowed for earlier detection of type II diabetes (Kopitar et al., 2020). When used with machine learning models, data from sensors can aid in diagnosis of many diseases (Yin and Jha, 2017), but they also provide a useful tool for patient monitoring to better understand mental health (Mohr et al., 2017) and facilitate health management of individuals with dementia (Enshaeifar et al., 2018). With new computational technologies facilitating the handling of more complex data types, machine learning models have been able to successfully detect COVID-19 from X-rays (Kassania et al., 2021; Rasheed et al., 2021), identify cardiac abnormalities using computed tomography (Singh et al., 2018) and predict numerous neurological outcomes based on MRI scans (Sabuncu and Konukoglu, 2015). In the dairy industry, machine learning models have primarily been developed for disease detection, although there has also been interest in applications relating to milk production and milk quality (Slob et al., 2021). Within disease detection, particular attention has been paid to developing models for the timely detection of mastitis, with predictor variables consisting primarily of milk characteristics measured during the milking process (Hogeveen et al., 2010; Slob et al., 2021).

Prediction models using images have also been developed to detect mastitis (Xudong et al., 2020) and lameness (Wu et al., 2020), as well as for animal identification (Jiang et al., 2019).

Many different machine learning algorithms exist, each with its own benefits and drawbacks, and existing literature shows the diversity of algorithms used for developing predictive models in the health and dairy settings (Hogeveen et al., 2010; Shatte et al., 2019; Triantafyllidis and Tsanas, 2019; Slob et al., 2021). For dairy applications in predicting milk production and udder health, models built using artificial neural networks tend to perform the best (Slob et al., 2021), whereas a recent review of clinical prediction models in the human health settings found no measurable benefit of using other algorithms over logistic regression (Christodoulou et al., 2019). Effective machine learning models should be able to take full advantage of the big data that is now used for their development by: 1) using large sample sizes to avoid overfitting training data and conducting robust validation of models, 2) incorporate different types of data from varied sources as predictor variables, and 3) make full use of the rate at which new data are generated. Most machine learning algorithms used for the development of clinical prediction models take advantage of the first two characteristics described above, but they tend to base their predictions on a single vector of observed variables at a time. Information about previous observations can be captured by computed variables such as the mean or variance over the preceding observations, or the time since the previous observation, but these variables need to be specified by the researcher beforehand. When developing models for the detection of mastitis in dairy cattle, some approaches have been to specify variable-specific temporal patterns beforehand to assess their association with mastitis (Cavero et al., 2008; Kamphuis et al., 2008); use principal component analysis on a fixed sequence length to compare with other sequences of

the same length to identify outliers (Miekley et al., 2013); and to ignore the sequential nature of the data entirely (Ebrahimi et al., 2019; Hajaya et al., 2019).

In addition to their effectiveness for the prediction of clinical outcomes, interest in developing artificial neural network models continues to grow as they consistently outperform other machine learning algorithms for tasks in computer vision (O'Mahony et al., 2019), text analysis (Sumathi and Pugalendhi, 2021) and even creating artwork and music (Jing et al., 2020). Neural network can be thought of as a non-parametric modelling approach, as the user has no input on how the relationship between predictors and outcome variables are defined. Neural networks are an extension of generalized linear regression that are able to identify extremely complex and abstract interactions between predictor variables without the need for explicit specification (Hastie et al., 2009). Prediction in neural networks occurs over multiple "hidden layers", each of which is composed of multiple nodes, each of which in turn compute a weighted sum of all the outputs of the previous layer, starting with the input variables and are described as "feed-forward". The final layer takes a weighted sum of the outputs of the previous layer, and the output of the final layer corresponds to the prediction task such as predicting a value for a regression problem or presenting class probabilities in a classification problem. As each node in the first hidden layer performs a weighted sum over all the input values, it can represent a D-way interaction, where D is the total number of input variables. Each subsequent hidden layer is then an interaction of interactions, increasing in complexity, thus networks with larger numbers of hidden layers and nodes are able to capture increasingly complex relationships among predictor variables and the outcome of interest. Model training proceeds by estimating parameter weights (the vector of weights used to take the weighted sum at each node) to minimize a loss function that measures the error in predictions. For each epoch (pass over the entire dataset), the network

first predicts the outcome based on the current parameter estimates (randomly initialized for the first epoch). The predicted value is then used to calculate the value for the loss function, and the partial derivative of the loss function with respect to each parameter in the model. The values of the partial derivatives are then used to update the parameter estimates, and the next epoch begins. Training can continue for a pre-defined number of epochs, or continue until the model's performance stops improving by a specified amount.

As each node within each hidden layer has a vector of weights that need to be estimated, complex neural networks can easily have hundreds of thousands or millions of parameters whose values need to be optimized. Although increased availability of graphical processing units (GPUs) with large amounts of memory have made training these models computationally feasible, the large number of parameters to be estimated does present some disadvantages for the model. As neural network training is entirely data driven (the user does not specify the variable relationships), neural networks are extremely prone to overfitting training data. When collecting data, neural networks need very large numbers of samples, that adequately represent the population they are sampled from, and a separate test set to be able to accurately estimate model generalizability. Regularization also needs to be used during the training process to restrict the magnitude of parameter estimates or to control how often parameter estimates are updated to limit overfitting. Limiting the number of epochs for training is another way to reduce the potential for overfitting, but this should be tied to changes in model performance. The complexity of the parameter relationships and the resulting partial derivatives at each epoch also present a computational challenge for optimization (Murugan and Durairaj, 2017). To avoid sub-optimal model fits resulting from local minima in the loss function's gradient, users can specify

adaptive optimization algorithms that adjust the learning rate (size of parameter updates) as the estimates approach minima.

Although neural networks are non-parametric, users have considerable control over the architecture, parameter regularization and optimization through specification of hyperparameters. Hyperparameter tuning is part of the process for developing any machine learning algorithm, but the large number of hyperparameters that can be controlled in neural network make this task more challenging than for other algorithms. As an example, when performing hyperparameter tuning for a support vector machine classifier, model fitting is controlled primarily by the kernel type and regularization parameter. To choose optimal values for those two parameters, a user might use a grid or random search over candidate values, and compare model fit to select the best combination. In a neural network on the other hand, the user can control the number of hidden layers, the number of nodes per layer, parameter regularization for each layer, the random initial values, and the learning rate for the optimization algorithm. A grid search over even a few candidate values for each of these hyperparameters quickly results in a large number of models that need to be compared, and due to the computational requirements to train each one, it typically is not feasible. Many studies rely on 'ad-hoc' tuning, changing only few hyperparameters, while using software defaults for the rest, but developing efficient methods for hyperparameter tuning in neural networks is an active area of research. Recurrent neural networks (RNNs) are an extension of the feed-forward artificial neural network, that are able to take into account temporal dependencies between predictor variables and the outcome of interest (Goodfellow et al., 2016). Long short-term memory cells can be implemented within RNNs to increase the ability of these networks to capture long-term dependencies between predictors and outcomes while reducing the computational requirements of parametrizing these relationships

(Hochreiter and Schmidhuber, 1997). Since RNNs are able to capture complex, non-linear variable interactions, there has been substantial interest in applying these to predict mortality, significant health events, and prescription medication usage based on electronic medical records (Choi et al., 2016b; a; Sha and Wang, 2017; Wang et al., 2019a). In dairy research, the only identified application of recurrent neural networks has been for the prediction of calving intervals (Keceli et al., 2020). Mastitis diagnosis is often preceded by changes in behaviour, milk composition, and production (Sepúlveda-Varas et al., 2016; King et al., 2017), so using models which can accurately capture these long-term patterns would likely lead to improved performance. One challenge of sequence modeling however, is the need to differentiate between early disease detection and false-positives, leading to the recommendation that outcome-independent reference point be used for model evaluation rather than evaluating the performance at only a specific point in time (Sherman et al., 2017).

1.5. Thesis rationale

As technology adoption continues to grow as a way to facilitate individual health monitoring and management, as well as to provide operational support to healthcare providers and dairy producers, the volume and variety of data generated will also grow. As these large datasets provide ample opportunities for scientific research, it is important that appropriate methods are chosen to maximize the value of inferences made using these secondary data, while also considering their limitations.

Mastitis in primiparous dairy cattle (heifers) has received less attention in research than in older cattle, since they have not been exposed to the milking process, and are often assumed to be less susceptible than older animals (De Vliegher et al., 2012). The rates of mastitis in dairy

heifers compared to older animals has not been described in Canada, although an earlier study from the Netherlands found that early lactation incidence of clinical mastitis in heifers can be 4 times higher than in older cattle (Barkema et al., 1998). Recommendations for overall mastitis control in dairy herds have existed for over 50 years (Neave et al., 1969), but recommendations specific for the control of heifer mastitis were only developed in 2014 (National Mastitis Council, 2014), despite the fact that risk factors for and consequences of mastitis in heifers are very different from in older animals (Gröhn et al., 2004; Piepers et al., 2009; Supré et al., 2011). Describing the occurrence of heifer mastitis in Canada and evaluating the effectiveness of heifer-specific management strategies therefore requires the use of data from multiple herds and varied sources, and the appropriate handling of the hierarchical nature of the data and explanatory variables at the lactation, cow, and herd-levels.

Models for the detection of mastitis onset have been developed as early as 1997 (De Mol et al., 1997), but the need for high performance models that accurately detect onset has grown due to the increasing adoption of automated milking systems on dairy farms (Barkema et al., 2015; Dairy barns by type, 2020). Dairy producers who transition to automated milking systems need to change their disease management practices, and perceive changes (either positive or negative) in the occurrence of various diseases in their herds (Tse et al., 2017). As producers have fewer direct interactions with their animals in these automated systems, surveillance for diseases in their herds will need to rely more on measurements taken by these systems rather than direct observation of animals. The development accurate mastitis detection models should therefore incorporate data from various sources and herds to maximize generalizability, use predictor variables of different types, while also appropriately accounting for the underlying heterogeneity that accompanies varied data sources.

As described in Section 1.3, multiplex cytokine assays provide a valuable tool for understanding and describing the immune response to many different diseases. In particular, the cytokine "fingerprint" of an individual can be used to understand their response to interventions for the management of autoimmune diseases such as Crohn's and ulcerative colitis (Dann et al., 2014; Powell et al., 2015), or as a biomarker to facilitate early detection of disease, infection and sepsis (Paajanen et al., 2002; Vermeire et al., 2004; Kafetzis et al., 2005; Sack et al., 2006; D'Haens et al., 2012). To gain insights from cytokine assays and combine them with other explanatory variables requires that analytical methods appropriately account for within-subject correlations of outcome measures, multiple within-subject comparisons, and multiple observations per subject.

The overall aims of this thesis were therefore: 1) to explore the scenarios described in this section by using secondary datasets that combine high volume data from varied sources; 2) identify valuable insights that can be generated from these data, while considering the context-specific priorities and limitations; and 3) apply the most appropriate big data methodologies for the analysis of these data. As described in the previous sections, there are many similarities in the large datasets used to understand human health and animal health in the dairy industry, so the selected methods will also be broadly applicable to other scenarios. The methods employed for these analyses will span the range of statistical tools available for epidemiological research including parametric regression models with defined variable-outcome relationships and rigid distributional assumptions, mixed models that are able to relax many of those assumptions and incorporate the effects of latent processes, and neural networks that make no distributional assumptions about the data and are able to learn undefined predictor-outcome relationships relevant for prediction. Lastly, along with gaining an understanding of the underlying data

structures in each of these datasets, insights from these datasets need to be relevant within their field, and address identified knowledge gaps while considering limitations of the secondary data for the specific research question.

1.6. Research objectives

1. Use data from bacteriological culturing of milk samples conducted for a larger study, combined with DHI records to describe the occurrence of clinical mastitis, subclinical mastitis, and intramammary infections in primiparous dairy cows in Canada (Chapter 2);
2. Conduct a systematic review and meta-analysis of studies assessing the effectiveness of different pre-calving therapies to improve post-calving udder health in primiparous dairy cows (Chapter 3);
3. Combine data from automated milking systems, activity monitors, on-farm treatment records, DHI and environmental temperature records to develop a recurrent neural network for the detection of clinical mastitis (Chapter 4);
4. Use simulated data to determine the impact of common issues with data collection on multiple dairy farms on the performance of a recurrent neural network for the detection of clinical mastitis (Chapter 5);
5. Combine data from emergency department admissions with results of a multiplex cytokine assay to compare immune responses of children presenting with symptoms of appendicitis (Chapter 6);

6. Combine data from dietary surveys with results of a multiplex cytokine assay to determine the association of dietary factors with gut inflammation in patients with Crohn's disease (Chapter 7).

1.7. References

Abril, A.G., T.G. Villa, J. Barros-Velázquez, B. Cañas, A. Sánchez-Pérez, P. Calo-Mata, and M. Carrera. 2020. Staphylococcus aureus exotoxins and their detection in the dairy industry and mastitis. Toxins 2020, Vol. 12, Page 537 12:537. doi:10.3390/TOXINS12090537.

Aghamohammadi, M., D. Haine, D.F. Kelton, H.W. Barkema, H. Hogeveen, G.P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5. doi:10.3389/fvets.2018.00100.

Ampe, B., K. Goethals, H. Laevens, and L. Duchateau. 2012. Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model. Prev. Vet. Med. 106:251–257. doi:10.1016/j.prevetmed.2012.04.004.

Barkema, H.W., M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.-P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare.. J. Dairy Sci. 98:7426–45. doi:10.3168/jds.2015-9377.

Barkema, H.W., Y.H. Schukken, T.J.G.M. Lam, M.L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419. doi:10.3168/jds.S0022-0302(98)75591-2.

Basu, A. 2010. Does a country's scientific 'productivity' depend critically on the number of country journals indexed?. Scientometrics 82:507–516. doi:10.1007/S11192-010-0186-8.

Van den Borne, B.H.P., J.C.M. Vernooij, A.M. Lupindu, G. van Schaik, K. Frankena, T.J.G.M. Lam, and M. Nielen. 2011. Relationship between somatic cell count status and subsequent clinical mastitis in Dutch dairy cows. Prev. Vet. Med. 102:265–273.

doi:10.1016/J.PREVETMED.2011.07.013.

Brodie, M.A., E.M. Pliner, A. Ho, K. Li, Z. Chen, S.C. Gandevia, and S.R. Lord. 2018. Big data vs accurate data in health research: Large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. Med. Hypotheses 119:32–36. doi:10.1016/J.MEHY.2018.07.015.

Buller, H., H. Blokhuis, K. Lokhorst, M. Silberberg, and I. Veissier. 2020. Animal welfare management in a digital world. Animals 10:1779. doi:10.3390/ANI10101779.

Cavero, D., K.H. Tölle, C. Henze, C. Buxadé, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114:280–286. doi:10.1016/j.livsci.2007.05.012.

Choi, E., M.T. Bahadori, A. Schuetz, W.F. Stewart, and J. Sun. 2016a. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. Pages 301–318 in Proceedings of Machine Learning for Healthcare. PMLR, Los Angeles, CA, USA.

Choi, E., A. Schuetz, W.F. Stewart, and J. Sun. 2016b. Using recurrent neural network models for early detection of heart failure onset. J. Am. Med. Informatics Assoc. 24:ocw112. doi:10.1093/jamia/ocw112.

Christodoulou, E., J. Ma, G.S. Collins, E.W. Steyerberg, J.Y. Verbakel, and B. Van Calster. 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J. Clin. Epidemiol. 110:12–22. doi:10.1016/J.JCLINEPI.2019.02.004.

Collazo-Reyes, F. 2013. Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. Scientometrics 98:197–209. doi:10.1007/S11192-013-1036-2.

Curtis, C.R., H.N. Erb, and M.E. White. 1988. Descriptive epidemiology of calfhood morbidity and mortality in New York holstein herds. Prev. Vet. Med. 5:293–307. doi:10.1016/0167-5877(88)90015-3.

D'Haens, G., M. Ferrante, S. Vermeire, F. Baert, M. Noman, L. Moortgat, P. Geens, D. Iwens, I. Aerden, G. Van Assche, G. Van Olmen, and P. Rutgeerts. 2012. Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease. Inflamm. Bowel Dis. 18:2218–2224. doi:10.1002/ibd.22917.

Dairy barns by type. 2020. . Accessed. https://www.dairyinfo.gc.ca/eng/dairy-statistics-and-market-information/farm-statistics/dairy-barns-by-type/?id=1502467060775.

Dann, S.M., C. Le, B.K. Choudhury, H. Liu, O. Saldarriaga, E.M. Hanson, Y. Cong, and L. Eckmann. 2014. Attenuation of Intestinal Inflammation in Interleukin-10-Deficient Mice Infected with Citrobacter rodentium. Infect. Immun. 82:1949–1958. doi:10.1128/IAI.00066-14.

Dohoo, I., W. Martin, and H. Stryhn. 2014a. Introduction to clustered data. 2nd ed. S.M. McPike, ed. VER Inc., Charlottetown, PE, Canada.

Dohoo, I.R., W. Martin, and H. Stryhn. 2014b. Mixed models for discrete data. 2nd ed. S.M. McPike, ed. VER Inc., Charlottetown, PE, Canada.

Dohoo, I.R., W. Martin, and H. Stryhn. 2014c. Repeated measures data. 2nd ed. S.M. McPike, ed. VER Inc., Charlottetown, PE, Canada.

Dufour, S., I.R. Dohoo, H.W. Barkema, L. DesCôteaux, T.J. DeVries, K.K. Reyher, J.-P.P. Roy, and D.T. Scholl. 2012. Epidemiology of coagulase-negative staphylococci intramammary infection in dairy cattle and the effect of bacteriological culture misclassification. J. Dairy Sci. 95:3110–3124. doi:10.3168/jds.2011-5164.

Ebrahimi, M., M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K.R. Petrovski. 2019. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Comput. Biol. Med. 114:103456. doi:10.1016/j.compbiomed.2019.103456.

Eckelkamp, E.A. 2019. Invited Review: Current state of wearable precision dairy technologies in disease detection. Appl. Anim. Sci. 35:209–220. doi:10.15232/AAS.2018-01801.

Ehrenstein, V., H. Nielsen, A.B. Pedersen, S.P. Johnsen, and L. Pedersen. 2017. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin. Epidemiol. 9:245. doi:10.2147/CLEP.S129779.

Elghafghuf, A., S. Dufour, K. Reyher, I. Dohoo, and H. Stryhn. 2014. Survival analysis of clinical mastitis data using a nested frailty Cox model fit as a mixed-effects Poisson model. Prev. Vet. Med. 117:456–468. doi:10.1016/j.prevetmed.2014.09.013.

Enshaeifar, S., A. Zoha, A. Markides, S. Skillman, S.T. Acton, T. Elsaleh, M. Hassanpour, A. Ahrabian, M. Kenny, S. Klein, H. Rostill, R. Nilforooshan, and P. Barnaghi. 2018. Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. PLoS One 13:e0195605. doi:10.1371/JOURNAL.PONE.0195605.

Firth, C.L., A. Käsbohrer, C. Schleicher, K. Fuchs, C. Egger-Danner, M. Mayerhofer, H. Schobesberger, J. Köfer, and W. Obritzhauser. 2017. Antimicrobial consumption on Austrian dairy farms: an observational study of udder disease treatments based on veterinary medication records. PeerJ 5:e4072–e4072.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. Sequence Modeling: Recurrent and Recursive Nets. 1st ed. MIT Press, Cambridge, MA, USA.

Gröhn, Y.T., D.J. Wilson, R.N. González, J.A. Hertl, H. Schulte, G. Bennett, and Y.H. Schukken. 2004. Effect of Pathogen-Specific Clinical Mastitis on Milk Yield in Dairy Cows. J. Dairy Sci. 87:3358–3374. doi:10.3168/jds.S0022-0302(04)73472-4.

Hajaya, M.K., S. Samarasinghe, D. Kulasiri, and M.L. Benavides. 2019. Detection of dairy cattle mastitis: Modelling of milking features using deep neural networks. Pages 35–41 in 23rd International Congress on Modelling and Simulation - Supporting Evidence-Based Decision Making: The Role of Modelling and Simulation, MODSIM 2019. Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Canberra, ACT, Australia.

Halasa, T., K. Huijps, O. Østerås, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18–31. doi:10.1080/01652176.2007.9695224.

Harford, T. 2014. Big data: A big mistake?. Significance 11:14–19. doi:10.1111/J.1740-9713.2014.00778.X.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Neural Networks. 2nd ed. Springer Science and Business Media, New York, NY.

Hawley, S., C.J. Edwards, N.K. Arden, A. Delmestri, C. Cooper, A. Judge, and D. Prieto-Alhambra. 2020. Descriptive epidemiology of hip and knee replacement in rheumatoid arthritis: An analysis of UK electronic medical records. Semin. Arthritis Rheum. 50:237–244. doi:10.1016/J.SEMARTHRIT.2019.08.008.

Herzberg, D., P. Strobel, A. Ramirez-Reveco, M. Werner, and H. Bustamante. 2020. Chronic inflammatory lameness increases cytokine concentration in the spinal cord of dairy cows. Front. Vet. Sci. 7:125. doi:10.3389/FVETS.2020.00125.

Higgins, J.P.T., S.G. Thompson, and D.J. Spiegelhalter. 2009. A re-evaluation of random-effects meta-analysis. J. R. Stat. Soc. Ser. A 172:137–159. doi:10.1111/J.1467-985X.2008.00552.X.

Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. Neural Comput. 9:1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and clinical mastitis-the quest for the perfect alert. Sensors 10:7991–8009. doi:10.3390/s100907991.

Hueston, W.D. 1988. Evaluation of alternatives for calculating clinical mastitis prevalence and incidence from prospective animal health monitoring schemes. Acta Vet. Scand. 84:442–445.

Hueston, W.D., L.E. Heider, W.R. Harvey, and K.L. Smith. 1987. The use of high somatic-cell count prevalence in epidemiologic investigations of mastitis control practices. Prev. Vet. Med. 4:447–461. doi:10.1016/0167-5877(87)90030-4.

Hunter, J.E., and F.L. Schmidt. 2000. Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. Int. J. Sel. Assess. 8:275–292. doi:10.1111/1468-2389.00156.

Jamali, H., H.W. Barkema, M. Jacques, E.M. Lavallée-Bourget, F. Malouin, V. Saini, H. Stryhn, and S. Dufour. 2018. Invited review: Incidence, risk factors, and effects of clinical mastitis recurrence in dairy cows. J. Dairy Sci. 101:4729–4746. doi:10.3168/JDS.2017-13730.

Jiang, B., Q. Wu, X. Yin, D. Wu, H. Song, and D. He. 2019. FLYOLOv3 deep learning for key parts of dairy cow body detection. Comput. Electron. Agric. 166:104982. doi:10.1016/j.compag.2019.104982.

Jing, Y., Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song. 2020. Neural Style Transfer: A Review.

IEEE Trans. Vis. Comput. Graph. 26:3365–3385. doi:10.1109/TVCG.2019.2921336.

Kafetzis, D.A., I.M. Velissariou, P. Nikolaides, M. Sklavos, M. Maktabi, G. Spyridis, D.D. Kafetzis, and E. Androulakakis. 2005. Procalcitonin as a predictor of severe appendicitis in children. Eur. J. Clin. Microbiol. Infect. Dis. 24:484–487. doi:10.1007/s10096-005-1360-4.

Kamphuis, C., D. Pietersma, R. van der Tol, M. Wiedemann, and H. Hogeveen. 2008. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Comput. Electron. Agric. 62:169–181. doi:10.1016/j.compag.2007.12.009.

Kassania, S.H., P.H. Kassanib, M.J. Wesolowskic, K.A. Schneidera, and R. Detersa. 2021. Automatic detection of coronavirus disease (COVID-19) in X-ray and CT Images: a machine learning based approach. Biocybern. Biomed. Eng. 41:867–879. doi:10.1016/J.BBE.2021.05.013.

Keceli, A.S., C. Catal, A. Kaya, and B. Tekinerdogan. 2020. Development of a recurrent neural networks-based calving prediction model using activity and behavioral data. Comput. Electron. Agric. 170:105285. doi:10.1016/j.compag.2020.105285.

King, M.T.M., K.M. Dancy, S.J. LeBlanc, E.A. Pajor, and T.J. DeVries. 2017. Deviations in behavior and productivity data before diagnosis of health disorders in cows milked with an automated system. J. Dairy Sci. 100:8358–8371.

King, M.T.M., S.J. LeBlanc, E.A. Pajor, T.C. Wright, and T.J. DeVries. 2018. Behavior and productivity of cows milked in automated systems before diagnosis of health disorders in early lactation. J. Dairy Sci. 101:4343–4356. doi:10.3168/jds.2017-13686.

Kopitar, L., P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic. 2020. Early detection of type 2 diabetes mellitus using machine learning-based prediction models. Sci. Rep. 10:1–12.

doi:10.1038/s41598-020-68771-z.

Lesueur, J., S. Barbey, N. Cebron, S. Walachowski, R. Lefebvre, P. Germon, D. Boichard, F. Corbière, and G. Foucras. 2019. Development of a bovine/ovine cytokine 15-plex assay for immunoprofiling of the cellular response in ruminants. Page 176 in International Veterinary Immunology Symposium, Seattle, WA,USA.

Levison, L.J., E.K. Miller-Cushon, A.L. Tucker, R. Bergeron, K.E. Leslie, H.W. Barkema, and T.J. DeVries. 2016. Incidence rate of pathogen-specific clinical mastitis on conventional and organic Canadian dairy farms. J. Dairy Sci 99:1341–1350. doi:10.3168/jds.2015-9809.

Li, L., B. Wagner, H. Freer, M. Schilling, J.P. Bannantine, J.J. Campo, R. Katani, Y.T. Grohn, J. Radzio-Basu, and V. Kapur. 2017. Early detection of Mycobacterium avium subsp. paratuberculosis infection in cattle with multiplex-bead based immunoassays. PLoS One 12:e0189783. doi:10.1371/JOURNAL.PONE.0189783.

Linkov, F., R.L. Ferris, Z. Yurkovetsky, A. Marrangoni, L. Velikokhatnaya, W. Gooding, B. Nolan, M. Winans, E.R. Siegel, A. Lokshin, and B.C. Stack. 2008. Multiplex analysis of cytokines as biomarkers that differentiate benign and malignant thyroid diseases. Proteomics – Clin. Appl. 2:1575–1585. doi:10.1002/PRCA.200780095.

Lip, G.Y.H., G. Tran, A. Genaidy, P. Marroquin, C. Estes, and J. Landsheft. 2021. Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. Eur. Hear. J. - Qual. Care Clin. Outcomes. doi:10.1093/EHJQCCO/QCAB037.

Liquet, B., K.-A.L. Cao, H. Hocini, and R. Thiébaut. 2012. A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. BMC Bioinforma. 2012 131 13:1–14. doi:10.1186/1471-2105-13-325.

Lokhorst, C., R.M. De Mol, and C. Kamphuis. 2019. Invited review: Big Data in precision dairy farming. Animal 13:1519–1528. doi:10.1017/S1751731118003439.

Majumder, S., T. Mondal, and M.J. Deen. 2017. Wearable sensors for remote health monitoring. Sensors 2017, Vol. 17, Page 130 17:130. doi:10.3390/S17010130.

Masaoud, E., and H. Stryhn. 2010. A simulation study to assess statistical methods for binary repeated measures data. Prev. Vet. Med. 93:81–97. doi:10.1016/j.prevetmed.2009.10.004.

De Mauro, A., M. Greco, and M. Grimaldi. 2016. A formal definition of Big Data based on its essential features. Libr. Rev. 65:122–135. doi:10.1108/LR-06-2015-0061.

Miekley, B., I. Traulsen, and J. Krieter. 2013. Principal component analysis for the early detection of mastitis and lameness in dairy cows. J. Dairy Res. 80:335–343. doi:10.1017/S0022029913000290.

Mohr, D.C., M. Zhang, and S.M. Schueller. 2017. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu. Rev. Clin. Psychol. 13:23–47. doi:10.1146/ANNUREV-CLINPSY-032816-044949.

De Mol, R.M., G.H. Kroeze, J.M.F.H. Achten, K. Maatje, and W. Rossing. 1997. Results of a multivariate approach to automated oestrus and mastitis detection. Livest. Prod. Sci. 48:219–227. doi:10.1016/S0301-6226(97)00028-6.

Mooney, S.J., D.J. Westreich, and A.M. El-Sayed. 2015. Epidemiology in the Era of Big Data. Epidemiology 26:390. doi:10.1097/EDE.0000000000000274.

Murugan, P., and S. Durairaj. 2017. Regularization and optimization strategies in deep convolutional neural networks. arXiv ID: 1712.04711v1

National Mastitis Council. 2014. Heifer Mastitis Prevention and Control Plan.

Neave, F.K., F.H. Dodd, R.G. Kingwill, and D.R. Westgarth. 1969. Control of Mastitis in the

Dairy Herd by Hygiene and Management. J. Dairy Sci. 52:696–707.

Neethirajan, S. 2020. The role of sensors, big data and machine learning in modern animal farming. Sens. Bio-Sensing Res. 29:100367. doi:10.1016/j.sbsr.2020.100367.

Nobrega, D.B., J. De Buck, S.A. Naqvi, G. Liu, S. Naushad, V. Saini, and H.W. Barkema. 2017. Comparison of treatment records and inventory of empty drug containers to quantify antimicrobial usage in dairy herds.. J. Dairy Sci. 100:9736–9745. doi:10.3168/jds.2017-13116.

Nobrega, D.B., S.A. Naqvi, S. Dufour, R. Deardon, J.P. Kastelic, J. De Buck, and H.W. Barkema. 2020. Critically important antimicrobials are generally not needed to treat nonsevere clinical mastitis in lactating dairy cows: Results from a network meta-analysis. J. Dairy Sci. 103:10585–10603. doi:10.3168/JDS.2020-18365.

O'Mahony, N., S. Campbell, A. Carvalho, S. Harapanahalli, G.V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh. 2019. Deep Learning vs. Traditional Computer Vision. Adv. Intell. Syst. Comput. 943:128–144. doi:10.1007/978-3-030-17795-9_10.

Paajanen, H., A. Mansikka, M. Laato, R. Ristamaki, K. Pulkki, and S. Kostiainen. 2002. Novel serum inflammatory markers in acute appendicitis. Scand. J. Clin. Lab. Invest. 62:579–584. doi:10.1080/003655102764654312.

Piepers, S., S. DeVliegher, A. DeKruif, G. Opsomer, and H. Barkema. 2009. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. Vet. Microbiol. 134:113–120. doi:10.1016/j.vetmic.2008.09.017.

Piepers, S., K. Peeters, G. Opsomer, H.W. Barkema, K. Frankena, and S. De Vliegher. 2011. Pathogen group specific risk factors at herd, heifer and quarter levels for intramammary infections in early lactating dairy heifers. Prev. Vet. Med. 99:91–101.

doi:10.1016/J.PREVETMED.2011.02.010.

Pinto, L.F., and L.J. dos Santos. 2020. Electronic medical records in primary care: management of duplicate records and a contribution to epidemiological studies. Cien. Saude Colet. 25:1305–1312. doi:10.1590/1413-81232020254.34132019.

Powell, M.L., M.A. Pegues, A.J. Szalai, V.K. Ghanta, L.R. D'Abramo, and S.A. Watts. 2015. Effects of the Dietary omega 3:omega 6 Fatty Acid Ratio on Body Fat and Inflammation in Zebrafish (Danio rerio). Comp. Med. 65:289–294.

Pralle, R.S., J.D. Amdall, R.H. Fourdraine, G.R. Oetzel, and H.M. White. 2021. Hyperketonemia predictions provide an on-farm management tool with epidemiological insights. Animals 11:1291. doi:10.3390/ANI11051291.

Rasheed, J., A.A. Hameed, C. Djeddi, A. Jamil, and F. Al-Turjman. 2021. A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. Interdiscip. Sci. Comput. Life Sci. 2021 131 13:103–117. doi:10.1007/S12539-020-00403-6.

Reddy, S.M., S. Patel, M. Weyrich, J. Fenton, and M. Viswanathan. 2020. Comparison of a traditional systematic review approach with review-of-reviews and semi-automation as strategies to update the evidence. Syst. Rev. 9. doi:10.1186/S13643-020-01450-2.

Reyher, K.K., S. Dufour, H.W. Barkema, L. Des Côteaux, T.J. DeVries, I.R. Dohoo, G.P. Keefe, J.-P. Roy, and D.T. Scholl. 2011. The National Cohort of Dairy Farms—A data collection platform for mastitis research in Canada. J. Dairy Sci. 94:1616–1626. doi:10.3168/jds.2010-3180.

Ruegg, P.L. 2003. Investigation of mastitis problems on farms. Vet. Clin. North Am. Anim. Pract. 19:47–73. doi:10.1016/S0749-0720(02)00078-6.

Ruegg, P.L. 2017. Mastitis detection, management, and prevention.. J. Dairy Sci. 100:10381–

10397.

Sabuncu, M.R., and E. Konukoglu. 2015. Clinical prediction from structural brain MRI scans: a large-scale empirical study. Neuroinformatics 13:31–46. doi:10.1007/s12021-014-9238-1.

Sack, U., B. Biereder, T. Elouahidi, K. Bauer, T. Keller, and R.-B. Trobs. 2006. Diagnostic value of blood inflammatory markers for detection of acute appendicitis in children.. BMC Surg. 6:15.

Schukken, Y.H., D. Bar, J. Hertl, and Y.T. Gröhn. 2010. Correlated time to event data: Modeling repeated clinical mastitis data from dairy cattle in New York State. Prev. Vet. Med. 97:150–156. doi:10.1016/J.PREVETMED.2010.09.012.

Senger, P.L. 1994. The estrus detection problem: new concepts, technologies, and possibilities. J. Dairy Sci. 77:2745–2753. doi:10.3168/JDS.S0022-0302(94)77217-9.

Sepúlveda-Varas, P., K.L. Proudfoot, D.M. Weary, and M.A.G. von Keyserlingk. 2016. Changes in behaviour of dairy cows with clinical mastitis. Appl. Anim. Behav. Sci. 175:8–13. doi:10.1016/J.APPLANIM.2014.09.022.

Sha, Y., and M.D. Wang. 2017. Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. ACM-BCB 2017 - Proc. 8th ACM Int. Conf. Bioinformatics, Comput. Biol. Heal. Informatics 233–240. doi:10.1145/3107411.3107445.

Shatte, A.B.R., D.M. Hutchinson, and S.J. Teague. 2019. Machine learning in mental health: a scoping review of methods and applications. Psychol. Med. 49:1426–1448. doi:10.1017/S0033291719000151.

Sherman, E., H. Gurm, U. Balis, S. Owens, and J. Wiens. 2017. Leveraging clinical time-series data for prediction: a cautionary tale. AMIA Annu. Symp. Proc. 2017:1571.

Singh, G., S.J. Al'Aref, M. Van Assen, T.S. Kim, A. van Rosendael, K.K. Kolli, A. Dwivedi, G.

Maliakal, M. Pandey, J. Wang, V. Do, M. Gummalla, C.N. De Cecco, and J.K. Min. 2018. Machine learning in cardiac CT: Basic concepts and contemporary data. J. Cardiovasc. Comput. Tomogr. 12:192–201. doi:10.1016/J.JCCT.2018.04.010.

Slob, N., C. Catal, and A. Kassahun. 2021. Application of machine learning to improve dairy farm management: A systematic literature review. Prev. Vet. Med. 187:105237. doi:10.1016/j.prevetmed.2020.105237.

Snijders, R., A. de Jong, T. Albers, B.D. van der Waaij, and M.R. Vonder. 2019. Interactive insight in big livestock data. Pages 145–151 in Proceedings of 9th European Conference on Precision Livestock Farming, Cork, Ireland.

Steensels, M., A. Antler, C. Bahr, D. Berckmans, E. Maltz, and I. Halachmi. 2016. A decision-tree model to detect post-calving diseases based on rumination, activity, milk yield, BW and voluntary visits to the milking robot. Animal 10:1493–1500. doi:10.1017/S1751731116000744.

Steensels, M., E. Maltz, C. Bahr, D. Berckmans, A. Antler, and I. Halachmi. 2017. Towards practical application of sensors for monitoring animal health: The effect of post-calving health problems on rumination duration, activity and milk yield. J. Dairy Res. 84:132–138. doi:10.1017/S0022029917000176.

Sumathi, S., and G.K. Pugalendhi. 2021. Cognition based spam mail text analysis using combined approach of deep neural network classifier and random forest. J. Ambient Intell. Humaniz. Comput. 12:5721–5731. doi:10.1007/S12652-020-02087-8/TABLES/7.

Supré, K., F. Haesebrouck, R.N. Zadoks, M. Vaneechoutte, S. Piepers, and S. De Vliegher. 2011. Some coagulase-negative Staphylococcus species affect udder health more than others. J. Dairy Sci. 94:2329–2340. doi:10.3168/jds.2010-3741.

Thompson, S.G., and S.J. Sharp. 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat. Med. 18:2693–2708. doi:10.1002/(SICI)1097-0258(19991030)18:20.

Triantafyllidis, A.K., and A. Tsanas. 2019. Applications of machine learning in real-life digital health interventions: Review of the literature. J. Med. Internet Res. 21:e12286. doi:10.2196/12286.

Troia, R., G. Mascalzoni, C. Agnoli, D. Lalonde-Paul, M. Giunti, and R. Goggs. 2020. Cytokine and chemokine profiling in cats with sepsis and septic shock. Front. Vet. Sci. 7:305. doi:10.3389/FVETS.2020.00305.

Tsafnat, G., P. Glasziou, M.K. Choong, A. Dunn, F. Galgani, and E. Coiera. 2014. Systematic review automation technologies. Syst. Rev. 3:74. doi:10.1186/2046-4053-3-74.

Tse, C., H.W. Barkema, T.J. DeVries, J. Rushen, and E.A. Pajor. 2017. Effect of transitioning to automatic milking systems on producers' perceptions of farm management and cow health in the Canadian dairy industry. J. Dairy Sci. 100:2404–2414. doi:10.3168/jds.2016-11521.

Ukoumunne, O.C., M.C. Gulliford, and S. Chinn. 2002. A note on the use of the variance inflation factor for determining sample size in cluster randomized trials. J. R. Stat. Soc. Ser. D 51:479–484. doi:10.1111/1467-9884.00332.

Del Valle, D.M., S. Kim-Schulze, H.-H. Huang, N.D. Beckmann, S. Nirenberg, B. Wang, Y. Lavin, T.H. Swartz, D. Madduri, A. Stock, T.U. Marron, H. Xie, M. Patel, K. Tuballes, O. Van Oekelen, A. Rahman, P. Kovatch, J.A. Aberg, E. Schadt, S. Jagannath, M. Mazumdar, A.W. Charney, A. Firpo-Betancourt, D.R. Mendu, J. Jhang, D. Reich, K. Sigel, C. Cordon-Cardo, M. Feldmann, S. Parekh, M. Merad, and S. Gnjatic. 2020. An inflammatory cytokine signature predicts COVID-19 severity and survival. Nat. Med. 2020 2610 26:1636–1643. doi:10.1038/s41591-020-1051-9.

Vayena, E., and A. Blasimme. 2018. Health research with Big Data: Time for systemic oversight. J. Law, Med. Ethics 46:119–129. doi:10.1177/1073110518766026.

Vermeire, S., G. Van Assche, and P. Rutgeerts. 2004. C-reactive protein as a marker for inflammatory bowel disease. Inflamm. Bowel Dis. Bowel Dis. 10:661–665. doi:10.1097/00054725-200409000-00026.

De Vliegher, S., H.W. Barkema, G. Opsomer, A. De Kruif, and L. Duchateau. 2005. Association between somatic cell count in early lactation and culling of dairy heifers using Cox frailty models. J. Dairy Sci. 88:560–568. doi:10.3168/JDS.S0022-0302(05)72718-1.

De Vliegher, S., L.K. Fox, S. Piepers, S. McDougall, and H.W. Barkema. 2012a. Mastitis in dairy heifers: nature of the disease, potential impact, prevention, and control.. J. Dairy Sci. 95:1025–1040.

De Vliegher, S., L.K. Fox, S. Piepers, S. McDougall, and H.W. Barkema. 2012b. Invited review: Mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. J. Dairy Sci. 95:1025–1040. doi:10.3168/jds.2010-4074.

De Vliegher, S., H. Laevens, H.W. Barkema, I.R. Dohoo, H. Stryhn, G. Opsomer, and A. De Kruif. 2004. Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers. J. Dairy Sci. 87:937–947. doi:10.3168/jds.S0022-0302(04)73238-5.

Wang, L., H. Wang, Y. Song, and Q. Wang. 2019a. MCPL-Based FT-LSTM: Medical representation learning-based clinical prediction model for time series events. IEEE Access 7:70253–70264. doi:10.1109/ACCESS.2019.2919683.

Wang, W., M. Kiik, N. Peek, V. Curcin, I.J. Marshall, A.G. Rudd, Y. Wang, A. Douiri, C.D. Wolfe, and B. Bray. 2020. A systematic review of machine learning models for predicting

outcomes of stroke with structured data. PLoS One 15:e0234722. doi:10.1371/JOURNAL.PONE.0234722.

Wang, X., C. Williams, Z.H. Liu, and J. Croghan. 2019b. Big data management challenges in health research—a literature review. Brief. Bioinform. 20:156–167. doi:10.1093/BIB/BBX086.

Warner, D., E. Vasseur, D.M. Lefebvre, and R. Lacroix. 2020a. A machine learning based decision aid for lameness in dairy herds using farm-based records. Comput. Electron. Agric. 169:105193. doi:10.1016/J.COMPAG.2019.105193.

Warner, D., E. Vasseur, M.V. Robichaud, S. Adam, D. Pellerin, D.M. Lefebvre, and R. Lacroix. 2020b. Development of a benchmarking tool for dairy herd management using routinely collected herd records. Animals 10:1689. doi:10.3390/ANI10091689.

Wu, D., Q. Wu, X. Yin, B. Jiang, H. Wang, D. He, and H. Song. 2020. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. Biosyst. Eng. 189:150–163. doi:10.1016/j.biosystemseng.2019.11.017.

Xudong, Z., K. Xi, F. Ningning, and L. Gang. 2020. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. Comput. Electron. Agric. 178:105754. doi:10.1016/j.compag.2020.105754.

Yin, H., and N.K. Jha. 2017. A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. IEEE Trans. Multi-Scale Comput. Syst. 3:228–241. doi:10.1109/TMSCS.2017.2710194.

CHAPTER 2: Udder health in Canadian dairy heifers during early lactation

2.1. Abstract

Mastitis is the most prevalent and costly disease in dairy cattle worldwide, with implications for animal health and welfare, as well as production and economics. Non-lactating heifers are an often-neglected group of animals concerning mastitis management, as they are assumed to be free of mastitis. An observational field study was conducted between 2007 and 2008 on 91 dairy herds across Canada, representative of provincial averages of bulk milk somatic cell count (BMSCC) and barn type. The aims of the study were to: 1) estimate in early lactating heifers overall and pathogen-specific incidence rate of clinical mastitis (IRCM), prevalence of intramammary infection (IMI), and prevalence of subclinical mastitis (SCM; defined as SCC ≥ 200,000 cells/mL); 2) compare these udder health parameters between heifers and multiparous cows; and 3) determine regional patterns and variation of these udder health parameters across BMSCC categories. During the first day of lactation, IRCM was higher in heifers than in multiparous cows (99 versus 48 cases per 10,000 quarter days at risk, respectively). Clinical mastitis affected 4% of heifers (0.73 cases per 100 quarters) in the first 30 days after calving, with the most common pathogens isolated being *Staphylococcus aureus* and *Escherichia coli*, whereas *S. aureus* and non-*aureus* staphylococci (NAS) were the most commonly isolated pathogens in multiparous cows. The IRCM in heifers was highest in Ontario, did not vary by BMSCC category, and was higher in multiparous cows than heifers. Intramammary infections were present in 33% of heifer quarters, with NAS being the most commonly isolated group of bacteria in both heifers (26% of quarters) and multiparous cows (18% of quarters). Pathogen-

specific prevalence of IMI did not differ between heifers and multiparous cows, but there were regional differences and differences across BMSCC categories in pathogen-specific prevalence of IMI. Prevalence of SCM in heifers was 13.6%, and was lowest in Alberta herds. In all regions, SCM prevalence was higher in multiparous cows than it was in heifers. In conclusion, udder health of Canadian dairy heifers was similar to other countries, demonstrating the importance of the issue. Differences between heifers and multiparous cows early in lactation highlighted the need for management practices to target the pre-calving period in heifers, when exposure to risk factors differs from that in lactating cows.

2.2. Introduction

Dairy heifers are typically not a focus when it comes to mastitis prevention and detection on farms, as they are not lactating and free from the physical and physiological stresses associated with the milking process (De Vliegher et al., 2012). However, the incidence of clinical mastitis (CM) and prevalence of subclinical mastitis (SCM) is high in heifers around calving, and the impact of CM on heifers, is more severe than that in older animals (Piepers et al., 2009). Heifers that calve with SCM or CM are more likely to be culled during their first lactation and persistence of mastitis into lactation has a stronger impact on future udder health and milk production (Piepers et al., 2009). Raising an animal to first calving represents a considerable investment that may not be recuperated over the course of a single lactation, and as such, there is a strong economic incentive for prevention and management of mastitis around calving in heifers (Rollin et al., 2015).

Mean incidence rate of CM (IRCM) on dairy farms ranges from 20 to 30 CM cases/100 cow-years at risk (Olde Riekerink et al., 2008a; Verbeke et al., 2014; Levison et al., 2016). The incidence of new intramammary infections (IMI) and CM in primiparous heifers are highest in the weeks before and immediately after calving, as rates are higher than in multiparous cows in early lactation (Barkema et al., 1998; De Vliegher et al., 2004; Moosavi et al., 2014). The distribution of bacterial species in milk samples from cows with CM varies with geographic region, bulk milk SCC (BMSCC), barn type, herd hygiene and bedding type (Kalmus et al., 2006; Olde Riekerink et al., 2008a; Verbeke et al., 2014; Rowbotham and Ruegg, 2016). In lactating cows, *Staphylococcus* species often comprise a very large proportion of bacteria isolated from IMI and mastitis cases (Kalmus et al., 2006; Verbeke et al., 2014; Yang et al.,

2015). In contrast, in heifers raised on pasture in New Zealand, *Streptococcus uberis* accounted for the majority of CM cases at calving (Pankey et al., 1991). In studies comparing bedding types in heifers housed indoors in the US and China, coagulase-negative staphylococci (CNS) and *Escherichia coli* were the most prevalent pathogens (Yang et al., 2015; Rowbotham and Ruegg, 2016). It is important to understand CM pathogen distribution, as they differ in their long-term impact (Piepers et al., 2009; Supré et al., 2011), as well as have different impacts in multiparous and primiparous cows (Gröhn et al., 2004).

No study has been conducted in Canada to determine species-specific IRCM in heifers. Through determining risks and pathogens to which heifers are exposed, producers will be able to create better targeted and effective management and treatment practices for mastitis. The objectives of this study were, therefore, to: 1) determine the species-specific IRCM and prevalence of SCM (defined as SCC ≥ 200,000 cells/mL) and IMI in Canadian dairy heifers in the first month of lactation; 2) compare these udder health parameters between primiparous heifers and multiparous cows; and 3) determine regional patterns and variation in these udder health parameters across BMSCC categories.

2.3. Material and Methods

2.3.1. Herds and samples

Details and rationale of the selection of farms, as well as details about the sampling series were described by Reyher et al. (2011). In short, data were collected from 91 farms in 6 Canadian provinces (17 in Alberta; 27 in Ontario; 29 in Quebec; 8 in PEI, 5 in Nova Scotia and 5 in New Brunswick, grouped together as Atlantic Canada) during 2007 and 2008. Farms were

selected to represent farms across Canada based on housing type (free-stall versus tie-stall), as well as uniformly encompassing 3 strata of 12-month geometric mean average BMSCC, classified as low, intermediate and high (<150,000, 150,000 to 300,000, and >300,000 cells/mL, respectively). In the present study, BMSCC herds were classified as low, intermediate, and high with cut-offs of <150,000, 150,000 to 250,000, and >250,000 cells/mL, respectively, to account for decreased herd BMSCC in the years following data collection. All farms enrolled in the study had at least 80% Holstein-Friesian cows in their herds, milked twice a day, and were enrolled in their regional DHI recording system.

Farmers were asked to identify cows with CM, based on abnormal milk or more severe signs of mastitis, and submit a quarter-milk sample on the day of CM diagnosis, with 2 follow-up samples during treatment (Reyher et al., 2011). A total of 16 of the 91 herds were excluded from the current study, because when audited, they had recorded < 80% of CM cases (Dufour et al., 2011). A non-clinical lactating longitudinal milk sampling series was also done to determine the incidence of IMI and of SCM in cows considered clinically normal at the beginning of the series. Ten lactating cows were selected randomly and the 5 most recently calved cows were added. Quarter milk samples were collected from these 15 cows by Canadian Bovine Mastitis Research Network (CBMRN) technicians 3 times every 3 weeks during spring of 2007 and during 2008. During the summer of 2007, an intensive sampling period consisted of 7 samples collected every week. In this study, only samples collected during the first 30 DIM were considered; although bacteriological findings were reported from all samples collected during the 30 DIM period, prevalence estimates were based on only the earliest sample from each animal during that period. Prevalence of SCM during the first 30 DIM were estimated from DHI milk records. Subclinical

mastitis was defined as a composite milk SCC ≥ 200,000 cells/mL in DHI composite milk samples.

2.3.2. Laboratory analyses

Details of sample processing protocol were described by Reyher et al. (2011). Briefly, milk samples collected by farm personnel (CM series) were stored on farm until monthly pickup by technicians, whereas samples collected by CBMRN technicians were immediately sent to 1 of 4 CBMRN laboratories for bacteriological culture and analysis. Before culture, non-CM samples were split and SCC was determined using a Fossomatic cell counter (Foss Electric, Hillerød, Denmark). Bacteriological culturing and identification were conducted following a standardized protocol outlined in the National Mastitis Council guidelines (Hogan et al., 1999). If > 2 species were isolated, the sample was considered contaminated, with the exception of *S. aureus* and *S. agalactiae,* in which case the species was still identified and quantified. For the purposes of defining IMI in the present study, an IMI by a particular species was defined as having ≥ 1 colonies formed from a 10-μL inoculum in mixed or pure culture, with no minimum SCC requirement (Dohoo et al., 2011). Pathogens were grouped into 4 categories: contagious pathogens (*S. aureus, Streptococcus dysgalactiae*), environmental pathogens (*E. coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*), non-*aureus* staphylococci (NAS), and other pathogens (other bacteria, yeasts, and fungi). As IMI by up to 2 pathogens was possible, a single quarter could be included in multiple pathogen groups.

2.3.3. Statistical analyses

Data organization and reshaping were done with R version 3.32 (R Development Core Team, 2008), whereas statistical analyses of disease and species distribution used the "geepack" (Halekoh et al., 2006) and the "lme4" (Bates, 2015) packages. A *P*-value < 0.05 was considered significant. Statistical analyses were conducted for 3 outcomes: CM, IMI, and SCM.

Clinical mastitis. The IRCM was calculated at the quarter level, with the population at risk determined from DHI records. Marginal models using generalized estimating equations (GEEs) were used to estimate parameters. Preliminary unconditional analyses detected no evidence of overdisperion, so the Poisson distribution was selected for all analyses. Only the herd-level correlation structure was included, as using the sandwich estimator for variance accounts for any nested clustering effects within the highest cluster (Betensky et al., 2000). Parity (dichotomized between heifers and multiparous cows), and either region (Alberta, Ontario, Quebec, and Atlantic Canada) or BMSCC (< 150,000; 150,000 – 250,000; or >250,000 cells/mL) were included as predictors, as well as interaction terms between parity and the other predictor to allow stratum-specific estimates. For within-BMSCC comparisons, a subgroup analysis stratifying by region was conducted. One model compared overall IRCM between heifers and multiparous cows, whereas the other 4 compared pathogen-specific IRCM between heifers and multiparous cows. The model used for estimates of CM was:

$$\ln(\mathrm{n}|\mathrm{Y_i}) = \beta_0 + \beta_1\mathrm{Parity} + \beta_2\mathrm{Region} + \beta_3\mathrm{BMSCC} + \beta_4\mathrm{Parity} \times \mathrm{Region} + \beta_5\mathrm{Parity} \times \mathrm{BMSCC_i} + \beta_6\,\mathrm{Region} \times \mathrm{BMSCC} + \beta_7\,\mathrm{Region} \times \mathrm{BMSCC} \times \mathrm{Parity} + \mathrm{Corr}(\mathrm{Y_i}) + \ln(\mathrm{p})$$

where $(n|Y_i)$ is the predicted number of overall or pathogen-specific clinical mastitis cases in herd i assumed to follow a Poisson distribution; β_1 to β_7 are regression coefficients for the

predictors and their interactions; $Corr(Y_i)$ is a correction based on the exchangeable correlation structure of observations within the same herd; $ln(p)$ is the offset of the equation, and the natural logarithm of the population at risk. Parameter estimates and predictions from GEEs represent population averaged estimates (average of all herds in Canada), in contrast to subject-specific estimates (for herds included in the study) provided by mixed effects models.

In addition, a model using only parity was used to compare differences in mean daily IRCM between heifers and multiparous cows for the first 15 DIM:

$$\ln(\mathrm{n}|\mathrm{Y_i}, \mathrm{DIM}) = \beta_0 + \beta_1 \mathrm{Parity} + \mathrm{Corr}(\mathrm{Y_i})_{\mathrm{DIM}} + \ln(\mathrm{p_{DIM}})$$

where $(n|Y_i, DIM)$ is the predicted number of CM cases that follow a Poisson distribution in herd i on a given DIM which ranged from 0 to 14; β_1 is the log incidence rate ratio between heifers and multiparous cows; $Corr(Y_i)$ is correction based on the exchangeable correlation structure of observations within the same herd on the given DIM; ln (p_{DIM}) is the offset of the equation and the natural logarithm of the population at risk on the given DIM.

Intramammary infection. Prevalence of IMI in the first 30 DIM was calculated at the quarter level using the first sample taken per animal and lactation as part of the non-clinical longitudinal series during that interval, with population at risk determined from the total number of quarters sampled during the first 30 DIM. Mixed effects logistic regression models were fitted to determine differences in prevalence between groups, with random effects at the herd and animal levels. Estimation of model parameters was conducted via REML with Laplace approximation. Although results were reported as odds, the inverse logit function was used to present prevalence (for ease of interpretation), although differences were still assessed by comparing odds ratios. The latter were not converted to risk ratios, as when dealing with common diseases (prevalence > 10%), converted odds ratio overestimate the strength of the association (McNutt et al., 2003).

The logit link was used instead of the log link, due to higher model convergence rates with the included random effects. Region, BMSCC and parity were included as predictors, as they were for the CM models. Significance of differences across predictors, and their interactions with parity, were forced into the model for overall IMI prevalence and pathogen-specific prevalences to enable reporting of stratum specific estimates. The model used for IMI prevalence estimate was:

$$\ln\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right)_{\mathrm{jk}} = \beta_0 + \beta_1\mathrm{Parity} + \beta_2\mathrm{Region} + \beta_3\mathrm{BMSCC_k} + \beta_4\mathrm{Parity} \times \mathrm{Region} + \beta_5\mathrm{Parity} \times \mathrm{BMSCC_k} + \beta_6\,\mathrm{Region} \times \mathrm{BMSCC_k} + \beta_7\,\mathrm{Region} \times \mathrm{BMSCC_k} \times \mathrm{Parity} + \mathrm{u_k} + \mathrm{v_j}$$

where $\left(\frac{p}{1-p}\right)_{jk}$ is the predicted quarter-level odds of having any or pathogen-specific IMI for animal j in herd k, where p is assumed to follow a binomial distribution; β_1 to β_7 are regression coefficients for the predictors and their; u_k is the random intercept term for herd, assumed to be normally distributed with a mean of 0 and a constant variance; and v_j is the random intercept term for animal, assumed to be normally distributed with a mean of 0 and a constant variance and accounted for correlation between multiple quarters sampled within each animal. Parameter estimates from mixed effect GLMs represent the average of all the subject-specific estimates from model (average of all the herds and animals included in the study).

Subclinical mastitis. Prevalence of SCM was calculated at the cow level using all DHI records available in the first 30 DIM. A mixed effects logistic regression model was used, with random effects at herd and animal levels. Estimation of model parameters was conducted via REML with Laplace approximation. The logit link was used again instead of the log link, so that between-group comparisons were done using odds ratios. Region and parity were included as

predictors as they were for CM and IMI, and the significance of differences across predictors and interactions with parity were included to determine stratum specific estimates of SCM prevalence. The model used for SCM prevalence prediction with stratification by region was:

$$\ln\left(\frac{\mathrm{p}}{1-\mathrm{p}}\right)_{\mathrm{jk}} = \beta_0 + \beta_1 \mathrm{Parity_j} + \beta_2 \mathrm{Region} + \beta_3 \mathrm{Parity_j} \times \mathrm{Region} + \mathrm{u_k} + \mathrm{v_j}$$

where $\left(\frac{p}{1-p}\right)_{jk}$ is the predicted animal-level odds of having SCM for animal j in herd k, where p is assumed to follow a binomial distribution; β_1 to β_3 are regression coefficients for the predictors and their interactions; u_k is the random intercept term for herd and was assumed to be normally distributed with a mean of 0 and a constant variance; and v_j is the random intercept term for animal and was assumed to be normally distributed with a mean of 0 and a constant variance.

All results presented are least-square means predictions from models with Wald confidence intervals; therefore, they are adjusted for clustering. In the case of IMI results, they were also adjusted for imperfect sensitivity and specificity using the following formula presented in Dohoo et al. (2014):

$$\text{True Prevalence} = \frac{\text{Apparent Prevalence} + \text{Specificity} - 1}{\text{Sensitivity} + \text{Specificity} - 1}$$

Values for sensitivity and specificity were derived from Dohoo et al. (2011b), where values for *S. aureus* were used for contagious pathogens; *E. coli* values were used for environmental pathogens (more conservative than the *Streptococcus* values); CNS values were used for NAS; and values for "Any Pathogen" were used for the Other pathogen category.

2.4. Results

2.4.1. Clinical mastitis

At the first DIM, IRCM in heifers (99 per 10,000 cow-days) was higher than in multiparous cows (IRR = 0.67: $P = 0.01$; Figure 2-1). At 4 DIM, multiparous IRCM (49 per 10,000 cow-days) was higher than that in heifers (11 per 10,000 cow-days). After 4 DIM, the incidence curves for the 2 groups converged and were not different for the next 10 DIM. In heifer CM samples, NAS were the most commonly isolated pathogen, whereas in multiparous cow CM samples, *S. aureus* was the most commonly isolated pathogen (21 and 18%, respectively; Table 2-1).

Across Canada, IRCM in the first 30 DIM did not differ between BMSCC categories, with a combined adjusted IRCM of 4.09 cases per 100 cow-months at risk, although there were some differences after stratification by region (Table 2-2). In high BMSCC herds in Ontario, IRCM in the first 30 DIM was higher in multiparous cows than in heifers. No regional differences in pathogen-specific IRCM were present in heifers, and only contagious IRCM was higher in the first 30 DIM in multiparous cows than it was in heifers (Table 2-3). Incidence rate of CM in heifers and multiparous cows was similar in the first 30 DIM for all BMSCC categories, but when combined was higher in multiparous cows than in heifers (Table 2-4). Pathogen-specific IRCM did not vary by BMSCC category, nor did it differ between heifers and multiparous cows for any category.

In a given herd, although IRCM in multiparous cows was predictive of IRCM in heifers (coefficient = 0.30; 95% CI: 0.10 – 0.50), they were very poorly correlated ($R^2 = 0.09$; Figure 2-2).

2.4.2. Intramammary infection

Distribution of pathogens in heifer and multiparous cow samples collected in the first 30 DIM was similar (Table 2-5). Non-*aureus* staphylococci were the most commonly isolated group of pathogens in both heifers and multiparous cows (34 and 31% of total number of samples, respectively).

Overall prevalence of IMI was higher in heifers (33.4%) than it was in multiparous cows (25.8%). IMI prevalence varied among regions (Table 2-6). Within heifers, NAS IMI prevalence was lowest in Atlantic Canada, whereas IMI by the other pathogen group was lowest in Atlantic Canada and Quebec. Prevalence of IMI did not vary between heifers and multiparous cows regionally, although there were some differences across Canada. Environmental IMI prevalence was higher ($P = 0.04$) in multiparous cows (3.0%) compared to heifers (2.0%), whereas NAS IMI was more prevalent in heifers (26.4%) than it was in multiparous cows (18.4%).

Overall prevalence of IMI was not different between BMSCC categories, although regional and pathogen-specific differences were present upon stratification (Table 2-7). Overall IMI prevalence was highest in high SCC Alberta herds, but IMI were more prevalent in multiparous cows in SCC herds in every region. Contagious pathogen IMI was lowest in medium SCC herds (1.7%), and within high SCC herds, was lowest in Ontario and Quebec. Environmental IMI prevalence in heifers did not vary regionally or among BMSCC categories, with a combined adjusted prevalence of 2.0%. Non-*aureus* staphylococci IMI prevalence was highest in high SCC Alberta herds (69.3%), and lowest in low SCC Atlantic Canada herds (1.9%). Prevalence of IMI by the other pathogen group varied both within heifers and between heifers and multiparous cows following stratification by region and BMSCC. In high and medium SCC herds, other IMI

prevalence was higher in multiparous cows than heifers. High SCC heifers from Quebec herds also had the lowest other pathogen group IMI prevalence of all high SCC herds.

Within herds, IMI prevalence in multiparous cows was predictive of IMI prevalence in heifers (coefficient = 0.66; 95% CI: 0.32 – 1.00), and the 2 were not highly correlated (R^2 = 0.13; Figure 2-3).

2.4.2. Subclinical mastitis

Prevalence of SCM in heifers was lowest in low BMSCC herds, with a combined prevalence of 13% of animals after adjustment for clustering and it increased with increasing BMSCC category, reaching 15 and 20% in medium and high BMSCC herds, respectively. Heifers in Alberta herds had the lowest SCM prevalence at 9.8%, whereas SCM prevalence was similar among other regions (Table 2-8). In all regions, SCM was more prevalent in multiparous cows than in heifers.

Prevalence of SCM in heifers in the first 30 DIM was positively associated with SCM prevalence in multiparous cows, where heifer SCM prevalence increased 0.29% (95% CI: 0.22 – 0.35%) for every 1% increase in multiparous SCM prevalence (R^2=0.45; Figure 2-4).

2.5. Discussion

The primary aim of this study was to quantify the occurrence of udder health problems in primiparous heifers on Canadian dairy farms in the early post-calving period. Although IRCM was higher in multiparous cows over the course of the first month of lactation, immediately after calving, IRCM was higher in heifers than multiparous cows. Only IRCM by contagious

pathogens was higher in multiparous cows than in heifers in the first 30 DIM, whereas all other pathogen-specific rates were similar between the 2 groups. Overall prevalence of IMI in the first 30 DIM was higher in heifers than in multiparous cows. Pathogen-specific differences were only present upon stratification by region and BMSCC. Prevalence of SCM was higher in multiparous cows than heifers in all BMSCC categories and regions.

The 30 DIM IRCM at animal level was 4.09 cases per 100 cows at risk, comparable to results published by Elghafghuf et al. (2011) using the same dataset, and yearly incidence in heifers was 21.3 per 100 cow-years at risk after a very high hazard of CM in the first 13 DIM. This IRCM was also consistent with Parker et al. (2007) who reported 13.6% IRCM during the first 4 months of lactation in dairy heifers in New Zealand. In an earlier study in Finland, IRCM in heifers was much higher at 4.4% in the first week of lactation (Myllys and Rautala, 1995). Barkema et al. (1998) reported similar results to the present study with a mean IRCM in heifers of 4.38 quarters per 10,000 cow-days at risk.

During the first day of lactation, overall IRCM was much higher in first-calving heifers than in multiparous cows, consistent with other reports (Barkema et al., 1998; Compton et al., 2007). After the first few DIM, heifer IRCM decreased to rates similar to multiparous cow IRCM, which was also demonstrated by Barkema et al. (1998). In contrast, Compton et al. (2007) reported that heifer IRCM remained consistently higher than older cow IRCM, even after the first week of lactation. In the present study, overall IRCM during the first 30 DIM was slightly lower in heifers (4.09 cases per 100-cow months at risk) than in multiparous cows (5.28 cases per 100-cow months at risk), consistent with results from Belgium (Verbeke et al., 2014), Estonia (Kalmus et al., 2006) and Sweden (Person Waller et al., 2009).

Pathogen distribution from CM highlighted interesting differences between the present study and other studies. In heifers with CM, NAS were the most commonly isolated pathogens, whereas in multiparous cows, *S. aureus* was the most commonly isolated pathogen. In Belgium, *Strep. uberis* and *E. coli* were the most commonly isolated pathogens from heifer CM (Verbeke et al., 2014); *Strep. uberis* was by far the most commonly isolated pathogen in New Zealand (Compton et al., 2007); *S. aureus, Strep. dysgalactiae,* and *E. coli* in Sweden (Persson Waller et al., 2009); *E. coli* and coagulase-negative staphylococci in China (Yang et al., 2015); *S. aureus* and *E. coli* in the Netherlands (Barkema et al., 1998), and *E. coli* and *Strep. uberis* in Estonia (Kalmus et al., 2006). In multiparous cows, *S. aureus* and *E. coli* were the most frequently isolated pathogen from clinical mastitis cases (Verbeke et al., 2014; Levison et al., 2016). In general, environmental pathogens are more commonly isolated from heifers than from multiparous cows with CM, whereas *S. aureus* is typically associated with mastitis in multiparous cows, and is occasionally isolated from heifers. The high number of NAS isolated from heifer CM samples (unlike many previous studies) highlighted their role in udder health, especially for Canadian herds. In the present study, only contagious IRCM was higher in multiparous cows than heifers in the first 30 DIM, whereas all others were very similar, suggesting that contagious pathogens were driving the higher IRCM in multiparous cows. Interestingly, IRCM in heifers did not vary among regions, although a previous study in Canada demonstrated significant regional variation in multiparous cows (Olde Riekerink et al., 2008b), attributed partially to regional differences in housing types and climate. As heifer housing in Canada typically differs from housing of multiparous cows, that difference in housing type may not be as important in young stock. Low within-herd correlations of IRCM between heifers and multiparous cows may have been due to differences in management, although as sample sizes for

this comparison were low, the analysis had limited reliability and should be interpreted with caution.

Overall prevalence of IMI in this study was similar to those conducted in other countries, although variation among those studies exists as well. Early lactation IMI prevalence was as low as 22% in New Zealand (Compton et al., 2007), and at approximately 30% in Belgium and Denmark (Aarestrup and Jensen, 1997; Piepers et al., 2009). Overall IMI prevalence of 34% of quarters in heifers was reported in this study, which seemed slightly higher than those studies (although animal level prevalence may be more similar), but still lower than a prevalence of 57% reported in Brazil (da Costa et al., 1996). Comparing IMI prevalence between studies can be difficult, due to differences in the case definition of IMI after culturing, with different sensitivities and specificities (Dohoo et al., 2011a; b) that can lead to effects on analysis due to misclassification (Dufour et al., 2012). In this study, a diagnosis of IMI diagnosis was a pathogen-specific case definition, based on a single sample with the highest sensitivities and specificities (Dohoo et al., 2011b), followed by correction of prevalence estimates using those values. Overall prevalence of IMI in the first 30 DIM was higher in heifers than in multiparous cows (34 vs 26%).

Pathogen distribution in heifer IMIs was more similar to previously described research. Non-*aureus* staphylococci were the most commonly isolated pathogens from milk samples, being isolated from 34% of samples. Studies in Brazil, Denmark, New Zealand, Belgium and the US, reported prevalence of CNS IMI at 49, 14, 10, 10, and 9%, respectively (da Costa et al., 1996; Aarestrup and Jensen, 1997; Compton et al., 2007; Piepers et al., 2009; Rowbotham and Ruegg, 2016). Overall prevalence of NAS IMI in heifers was 26%, although using a Bayesian framework, adjusted CNS IMI prevalence was 43% (Dufour et al. 2012). This difference may be

attributed to different adjustment methods, and different study populations (heifers in the present study versus both heifers and multiparous cows in Dufour et al. 2012). Regional differences in IMI prevalence were present for only NAS and other pathogens, with Atlantic Canada having the lowest prevalence of both. Differences in regional climate and housing types, and other management practices that were not captured in this study may have contributed to differences in bacterial species present in the environment. There was no variation in prevalence of IMI by any pathogen type across BMSCC categories without further stratification by region. In high BMSCC herds, overall IMI was more prevalent in multiparous cows in all regions. Interestingly, contagious pathogen prevalence in heifers was actually highest in low SCC herds, but was not different from prevalence in multiparous cows. In both high and medium SCC farms, contagious IMI prevalence was much higher in multiparous cows. These observations were counterintuitive as proper milking management practices should decrease prevalence of contagious pathogens and BMSCC.

Low within-herd correlation of heifer IMI prevalence and older cow IMI prevalence may be reflective of the different environment and pathogens the animals are exposed to prior to calving. Again, a low sample size makes this comparison fairly crude, and results should be interpreted with caution.

Subclinical mastitis was present in 13.6% of heifers tested in the first 30 DIM, and was consistent regionally, except for in Alberta where it was only 9.8%. In a previous study (Santman-Berends et al. 2012), prevalence of SCM in heifers was 26% in the first 100 DIM; perhaps this higher estimate was due to a longer observation period. Rowbotham and Ruegg (2016) estimated SCM in heifers to be present in 31% of quarters over the course of the first year in lactation. Comparisons with previous research in heifers was difficult, due to different case

definitions for subclinical mastitis (this study used SCC ≥ 200,000 cells/mL) and follow-up periods, although overall, results seemed consistent. Within-herd correlation of SCM prevalence between heifers and multiparous cows was higher than it was for IRCM and IMI. Sample sizes for SCM prevalence were much higher than the other 2 parameters, suggesting that correlations in the others may not have been significant due to lack of power.

2.6. Conclusions

Similar to previous studies, heifer IRCM was higher than in multiparous cows in the first day after calving but thereafter lower in heifers than in multiparous cows. No regional variation in IRCM by any pathogen was observed, nor was there any major difference among BMSCC categories. IMI prevalence in Canada was similar to described in other studies, highlighting the importance of the issue. Overall prevalence of IMI was higher in heifers than in older cows, with some pathogen-specific differences regionally and across BMSCC. Subclinical mastitis prevalence in heifers was consistent across most of Canada, and was lower than SCM prevalence in multiparous cows. Based on similarities in disease rates between heifers and multiparous cows in the early post-calving period, management practices in the Canadian dairy industry must include the pre-calving period for heifers.

2.7. References

Aarestrup, F. M. and N. E. Jensen. 1997. Prevalence and duration of intramammary infection in Danish heifers during the peripartum period. J. Dairy Sci. 80:307–312. doi:10.3168/jds.S0022-0302(97)75939-3.

Barkema, H. W., Y. H. Schukken, T. J. G. M. Lam, M. L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419. doi:10.3168/jds.S0022-0302(98)75591-2.

Bates, D., M. Maechler, B. Bolker, B., and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software. 67:1-48. doi:10.18637/jss.v067.i01.

Betensky, R.A., Talcott, J.A., and J.C. Weeks. 2000. Binary data with two, non-nested sources of clustering: an analysis of physician recommendations for early prostate cancer treatment. Biostatistics. 1: 219-230. doi: 10.1093/biostatistics/1.2.219.

Compton, C. W. R., C. Heuer, K. Parker, and S. McDougall. 2007. Epidemiology of mastitis in pasture-grazed peripartum dairy heifers and its effects on productivity. J. Dairy Sci. 90:4157–4170. doi:10.3168/jds.2006-880.

da Costa, E. O., P. A. Melville, A. R. Ribeiro, E. Watanabe, F. C. Viani, and C. R. White. 1996. Prevalence of intramammary infections in primigravid Brazilian dairy heifers. Prev. Vet. Med. 29:151–155. doi:10.1016/S0167-5877(96)01072-0.

De Vliegher, S., L. K. Fox, S. Piepers, S. McDougall, and H. W. Barkema. 2012. Invited review: Mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. J. Dairy Sci. 95:1025–1040. doi:10.3168/jds.2010-4074.

De Vliegher, S., H. Laevens, H. W. Barkema, I.R. Dohoo, H. Stryhn, G. Opsomer, and A. de Kruif. 2004. Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers. J. Dairy Sci. 87:937–947. doi:10.3168/jds.S0022-0302(04)73238-5.

Dohoo, I., S. Andersen, R. Dingwell, K. Hand, D. Kelton, K. Leslie, Y. Schukken, and S. Godden. 2011a. Diagnosing intramammary infections: comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94:5515–5522. doi:10.3168/jds.2011-4486.

Dohoo, I. R., J. Smith, S. Andersen, D. F. Kelton, and S. Godden. 2011b. Diagnosing intramammary infections: evaluation of definitions based on a single milk sample. J. Dairy Sci. 94:250–261. doi:10.3168/jds.2010-3559.

Dohoo, I. R., Stryhn, H., and S. W. Martin. 2014. Veterinary Epidemiologic Research. VER Inc., Charlottetown, PE, Canada.

Dufour, S., I. R. Dohoo, H. W. Barkema, L. Descôteaux, T. J. Devries, K. K. Reyher, J.-P. Roy, and D. T. Scholl. 2012. Epidemiology of coagulase-negative staphylococci intramammary infection in dairy cattle and the effect of bacteriological culture misclassification. J. Dairy Sci. 95:3110–3124. doi:10.3168/jds.2011-5164.

Dufour, S., A. Le Plain, K. Reyher, and D. Scholl. 2011. Clinical mastitis: What have we measured? Canadian Bovine Mastitis Research Network Annual Meeting. Montreal, QC, Canada.

Gröhn, Y. T., D. J. Wilson, R. N. González, J. A. Hertl, H. Schulte, G. Bennett, and Y. H. Schukken. 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87:3358–3374. doi:10.3168/jds.S0022-0302(04)73472-4.

Halekoh, U., Højsgaard, S., and J. Yan. 2006. The R package geepack for Generalized Estimating Equations. J. Stat. Soft. 15:1-11. doi: 10.18637/jss.v015.i02.

Hogan, J. S., Gonzalez, R. N., Harmon, R. J., S. C. Nickerson, S. P. Oliver, J. W. Pankey, and K.L. Smith. 1999. Laboratory handbook on bovine mastitis. National Mastitis Council, Madison, WI. 6-10 pp.

Kalmus, P., A. Viltrop, B. Aasmäe, and K. Kask. 2006. Occurrence of clinical mastitis in primiparous Estonian dairy cows in different housing conditions. Acta Vet. Scand. 48:21. doi:10.1186/1751-0147-48-21.

Levison, L. J., E. K. Miller-Cushon, A. L. Tucker, R. Bergeron, K. E. Leslie, H. W. Barkema, and T. J. DeVries. 2016. Incidence rate of pathogen-specific clinical mastitis on conventional and organic Canadian dairy farms. J. Dairy Sci. 99:1341–1350. doi:10.3168/jds.2015-9809.

McNutt, L.-A., C. Wu, X. Xue, and J. P. Hafner. 2003. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am. J. Epidemiol. 157:940–943. doi:10.1093/aje/kwg074.

Moosavi, M., A. Mirzaei, M. Ghavami, and A. Tamadon. 2014. Relationship between season, lactation number and incidence of clinical mastitis in different stages of lactation in a Holstein dairy farm. Vet. Res. Forum. 5:13–19.

Myllys, V. and H. Rautala. 1995. Characterization of clinical mastitis in primiparous heifers. J. Dairy Sci. 78:538–545. doi:10.3168/jds.S0022-0302(95)76664-4.

Olde Riekerink, R. G. M., H. W. Barkema, D. F. Kelton, and D. T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. J. Dairy Sci. 91:1366–1377. doi:10.3168/jds.2007-0757.

Pankey, J. W., P. A. Drechsler, and E. E. Wildman. 1991. Mastitis prevalence in primigravid heifers at parturition. J. Dairy Sci. 74:1550–1552. doi:10.3168/jds.S0022-0302(91)78316-1.

Persson Waller, K., B. Bengtsson, A. Lindberg, A. Nyman, and H. Ericsson Unnerstad. 2009. Incidence of mastitis and bacterial findings at clinical mastitis in Swedish primiparous cows—Influence of breed and stage of lactation. Vet. Microbiol. 134:89–94. doi:10.1016/j.vetmic.2008.09.004.

Piepers, S., S. De Vliegher, A. De Kruif, G. Opsomer, and H.W. Barkema. 2009. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. Vet. Microbiol. 134:113–120. doi:10.1016/j.vetmic.2008.09.017.

Rollin, E., K. C. Dhuyvetter, and M. W. Overton. 2015. The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. Prev. Vet. Med. 122:257–264. doi:10.1016/j.prevetmed.2015.11.006.

Rowbotham, R. F. and P. L. Ruegg. 2016. Associations of selected bedding types with incidence rates of subclinical and clinical mastitis in primiparous Holstein dairy cows. J. Dairy Sci. 99:1–11. doi:10.3168/jds.2015-10675.

Supré, K., F. Haesebrouck, R. N. Zadoks, M. Vaneechoutte, S. Piepers, and S. De Vliegher. 2011. Some coagulase-negative *Staphylococcus* species affect udder health more than others. J. Dairy Sci. 94:2329–2340. doi:10.3168/jds.2010-3741.

Verbeke, J., S. Piepers, K. Supré, S. De Vliegher, and S. De Vliegher. 2014. Pathogen-specific incidence rate of clinical mastitis in Flemish dairy herds, severity, and association with herd hygiene. J. Dairy Sci. 97:6926–6934. doi:10.3168/jds.2014-8173.

Yang, F. L., C. Shen, and B. X. He. 2015. The prevalence of heifer mastitis and its associated risk factors in Huanggang, Central China. Trop. Anim. Health Prod. 47:87–92. doi:10.1007/s11250-014-0689-z.

Table 2-1. Distribution of pathogens in clinical mastitis samples from 75 dairy farms in 4 Canadian regions during the first 30 DIM for primiparous (n=197) and multiparous cattle (n=485).

Pathogen	Primiparous			Multiparous		
	Frequency	Percentage of samples (%)	Percentage of isolates[3] (%)	Frequency	Percentage of samples (%)	Percentage of isolates (%)
Non-*aureus staphylococci*	41	20.8	24.7	77	15.9	17.5
Staphylococcus aureus	26	13.2	15.7	88	18.1	20.0
Escherichia coli	25	12.7	15.1	57	11.8	13.0
Streptococcus spp.	23	11.7	13.9	55	11.3	12.5
Streptococcus uberis	12	6.1	7.2	43	8.9	9.8
Other Gram-positive bacteria[1]	10	5.1	6.0	32	6.6	7.3
Streptococcus dysgalactiae	9	4.6	5.4	35	7.2	8.0
Trueperella pyogenes	6	3.0	3.6	18	3.7	4.1
Other Gram-negative bacteria[2]	5	2.5	3.0	8	1.6	1.8
Klebsiella spp.	4	2.0	2.4	10	2.1	2.3
Corynebacterium spp.	3	1.5	1.8	8	1.6	1.8
Yeasts and Fungi	2	1.0	1.2	4	0.8	0.9
Prototheca spp.	0	0.0	0.0	2	0.4	0.5
Mixed culture	21	10.7	-	68	14.0	-
Culture-negative	39	19.8	-	75	15.5	-
Contaminated	13	6.6	-	39	8.0	-

[1]Includes *Bacillus* spp.
[2]Includes *Serratia* spp., *Citrobacter* spp., *Proteus* spp., *Salmonella* spp., *Pseudomonas* spp., and *Pasteurella multocida*.
[3]Each sample could contribute up to 2 isolates, whereas samples with ≥ 3 isolates were isolated were considered contaminated.

Contaminated and culture-negative samples were not considered when determining percentage of isolates.

Table 2-2. Incidence rate of clinical mastitis (IRCM; cases per 100 cow-months at risk) in the first 30 DIM of heifers in Canadian dairy herds.

	No. Herds	No. clinical mastitis cases	Total quarter-months at risk	Mean IRCM (cases per 100 cow-months)	95% CI	Rate Ratio compared to Multiparous	95% CI
Low BMSCC[1]	14	33	815	4.05	2.79 - 5.68	1.18	0.70 - 1.99
Alberta	3	2	208	0.96	0.12 - 3.46	4.23	0.85 - 21.0
Ontario	3	19	160	11.8[2]	7.12 - 18.5	0.14	0.03 - 0.69
Quebec	3	2	153	1.30	0.16 - 4.69	0.31	0.04 - 2.41
Atlantic Canada	5	10	291	3.43[3]	1.64 - 6.30	0.39	0.07 - 2.12
Medium BMSCC[1]	31	88	2,204	3.99	3.20 - 4.92	1.34	0.69 - 2.60
Alberta	3	9	276	3.26	1.49 - 6.19	0.56	0.07 - 4.30
Ontario	14	38	995	3.82	2.70 - 5.24	4.85[4]	0.56 - 42.4
Quebec	5	10	241	4.13	1.98 - 7.60	2.20	0.20 - 24.72
Atlantic Canada	9	31	690	4.49	3.05 - 6.37	1.53	0.16 - 14.4
High BMSCC[1]	30	64	1,498	4.27	3.29 - 5.45	0.82	0.42 - 1.58
Alberta	5	13	334	3.89[3]	2.07 - 6.64	0.24[4]	0.04 - 1.60
Ontario	9	21	517	4.05	2.51 - 6.20	6.60[5]	0.92 - 47.2
Quebec	13	25	511	4.89	3.16 - 7.21	2.35	0.22 - 25.1
Atlantic Canada	3	5	134	3.72	1.21 - 8.68	4.98	0.55 - 45.6
Total	75	185	4,518	4.09	3.53 - 4.73	1.29	0.98 - 1.69

[1]Bulk milk somatic cell count. Low: <150,000 cells/mL; medium: 150,000 to 250,000 cells/mL; high: >250,000 cells/mL.
[2]Different ($P < 0.05$) within a BMSCC category, using Alberta herds as a baseline.
[3]Tendency for difference ($0.05 < P < 0.10$) within a BMSCC category using Alberta herds as a baseline or from low BMSCC within Alberta.
[4]Tendency for difference ($0.05 < P < 0.10$) between heifers and multiparous cows.
[5]Different ($P < 0.05$) between heifers and multiparous cows.

Table 2-3. Mean incidence rate of clinical mastitis (cases per 100 cow¬-days at risk) in the first 30 DIM of heifers and multiparous cows in dairy herds from 4 regions across Canada, stratified by pathogen.

Pathogen Type	Parity	Alberta (95% CI)	Ontario (95% CI)	Quebec (95% CI)	Atlantic (95% CI)	All (95% CI)
Contagious[1]	1	0[2]	0.84[3] (0.47 - 1.50)	1.54 (0.85 - 2.80)	0.63 (0.30 - 1.29)	0.77[4] (0.52 - 1.15)
	>1	0.91 (0.57 - 1.45)	1.51 (0.92 - 2.46)	1.74 (1.14 - 2.67)	1.34 (0.82 - 2.18)	1.42[4] (1.10 - 1.84)
Environmental[1]	1	1.46 (0.60 - 3.56)	1.61 (0.91 - 2.84)	1.10 (0.54 - 2.25)	1.52 (0.87 - 2.67)	1.46 (1.04 - 2.05)
	>1	2.23 (1.07 - 4.66)	1.86 (1.02 - 3.39)	1.21 (0.76 - 1.93)	2.46 (1.44 - 4.18)	1.92 (1.40 - 2.64)
NAS[1]	1	0.49 (0.20 - 1.22)	1.25[5] (0.70 - 2.26)	0.66 (0.23 - 1.91)	0.54 (0.20 - 1.47)	0.82 (0.53 - 1.27)
	>1	1.12 (0.45 - 2.75)	1.03 (0.54 - 1.95)	0.58 (0.31 - 1.09)	0.47 (0.26 - 0.86)	0.79 (0.54 - 1.17)
Other[1]	1	0.12 (0.02 - 0.68)	0.24 (0.09 - 0.62)	0.55 (0.21 - 1.46)	0.36 (0.14 - 0.89)	0.31 (0.18 - 0.53)
	>1	0.21 (0.09 - 0.50)	0.48 (0.26 - 0.88)	0.44 (0.19 – 1.00)	0.56 (0.28 - 1.11)	0.45 (0.31 - 0.65)
Culture-negative	1	0.73 (0.22 - 2.40)	0.84 (0.48 - 1.44)	0.22 (0.06 - 0.85)	1.07 (0.46 - 2.49)	0.75 (0.48 - 1.17)
	>1	0.84 (0.37 - 1.89)	0.67 (0.37 - 1.23)	0.29 (0.11 - 0.76)	1.47 (0.92 - 2.35)	0.82 (0.58 - 1.16)
Overall	1	2.93[4] (1.36 - 6.30)	4.66[3] (3.07 - 7.07)	4.08 (2.53 - 6.58)	4.12 (2.61 - 6.51)	4.09[4] (3.18 - 5.27)
	>1	5.24[4] (3.11 - 8.82)	5.35 (3.39 - 8.43)	4.12 (2.67 - 6.35)	6.25 (4.21 - 9.28)	5.28[4] (4.18 - 6.67)

[1]Contagious = *Staphylococcus aureus, Streptococcus dysgalactiae*; environmental = *Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*; NAS = non-*aureus* staphylococci; other pathogens = other bacteria, yeasts, and fungi.
[2]Incidence and sample size were too low to reliably estimate confidence intervals and rates.
[3]Different ($P < 0.05$) from Alberta within parity.
[4]Different ($P < 0.05$) between heifers and multiparous cows within the same pathogen group and region.
[5]Tendency for difference ($0.05 < P < 0.10$) from Alberta within parity.

Table 2-4. Mean incidence rate of clinical mastitis (per 100 cow-months at risk) in heifers and multiparous cows in the first 30 DIM in Canada stratified by herd bulk milk somatic cell count (BMSCC) and pathogen category.

	Parity	BMSCC category			
		Low[2] (95% CI)	Medium[2] (95% CI)	High[2] (95% CI)	All Herds (95% CI)
Contagious[1]	1	0.98 (0.49 - 1.98)	0.68 (0.40 - 1.17)	0.80 (0.36 - 1.79)	0.77[3] (0.52 - 1.15)
	>1	1.40 (0.89 - 2.22)	1.70 (1.15 - 2.50)	1.07 (0.72 - 1.58)	1.42[3] (1.10 - 1.84)
Environmental[1]	1	1.59 (0.88 - 2.88)	1.22 (0.73 - 2.05)	1.73 (0.96 - 3.15)	1.46 (1.04 - 2.05)
	>1	1.40 (0.89 - 2.22)	1.70 (1.15 - 2.50)	1.07 (0.72 - 1.58)	1.92 (1.40 - 2.64)
NAS[1]	1	0.61 (0.20 - 1.88)	1.04 (0.59 - 1.85)	0.60 (0.28 - 1.30)	0.82 (0.53 - 1.27)
	>1	1.28 (0.81 - 2.02)	2.63 (1.70 - 4.09)	1.32 (0.89 - 1.94)	0.79 (0.54 - 1.17)
Other[1]	1	0.25 (0.06 – 1.00)	0.32 (0.16 - 0.64)	0.33 (0.13 - 0.87)	0.31 (0.18 - 0.53)
	>1	0.70 (0.37 - 1.33)	0.81 (0.50 - 1.32)	0.81 (0.37 - 1.80)	0.45 (0.31 - 0.65)
Culture-negative	1	0.74 (0.30 - 1.79)	1.00 (0.56 - 1.77)	0.40 (0.18 - 0.91)	0.75 (0.48 - 1.17)
	>1	0.26 (0.05 - 1.20)	0.55 (0.35 - 0.87)	0.41 (0.22 - 0.75)	0.82 (0.58 - 1.16)
Overall	1	4.05 (2.17 - 7.53)	3.99 (2.82 - 5.65)	4.27 (2.72 - 6.70)	4.09[3] (3.18 - 5.27)
	>1	4.79 (3.53 - 6.49)	6.34 (4.49 - 8.96)	4.14 (2.80 - 6.10)	5.28[3] (4.18 - 6.67)

[1]Contagious = *Staphylococcus aureus, Streptococcus dysgalactiae*; environmental = *Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*; NAS = non-*aureus* staphylococci; other pathogens = other bacteria, yeasts, and fungi.
[2]Low: <150,000 cells/mL; medium: 150,000 to 250,000 cells/mL; high: >250,000 cells/mL.
[3]Different ($P < 0.05$) between heifers and multiparous cows within the same pathogen group and BMSCC category.

Table 2-5. Distribution of mastitis pathogens in milk from clinically normal, randomly selected primiparous (n=2,265) and multiparous (n=6,066) lactating cows from 91 dairy farms in 4 Canadian regions in the first 30 DIM.

Pathogen	Primiparous			Multiparous		
	Frequency	Percentage of samples (%)	Percentage of isolates[3] (%)	Frequency	Percentage of samples (%)	Percentage of isolates[3] (%)
Non-*aureus staphylococci*	766	33.8	62.3	1849	30.5	57.6
Other Gram-positive bacteria[1]	229	10.1	18.6	563	9.3	17.5
Streptococcus spp.	73	3.2	5.9	252	4.2	7.8
Corynebacterium spp.	72	3.2	5.9	284	4.7	8.8
Staphylococcus aureus	55	2.4	4.5	150	2.5	4.7
Streptococcus dysgalactiae	5	0.2	0.4	13	0.2	0.4
Streptococcus uberis	5	0.2	0.4	24	0.4	0.7
Other Gram-negative bacteria[2]	4	0.2	0.3	22	0.4	0.7
Yeasts and Fungi	4	0.2	0.3	19	0.3	0.6
Escherichia coli	3	0.1	0.2	16	0.3	0.5
Prototheca spp.	2	0.1	0.2	4	0.1	0.1
Klebsiella spp.	0	0.0	0.0	4	0.1	0.1
Trueperella pyogenes	0	0.0	0.0	5	0.1	0.2
Enterobacter spp.	0	0.0	0.0	7	0.1	0.2
Mixed culture	231	10.2	-	679	11.2	-
Culture-negative	1,066	47.1	-	2,767	45.6	-
Contaminated	200	8.8	-	767	12.6	-

[1]Includes *Bacillus* spp.
[2]Includes *Serratia* spp., *Citrobacter* spp., *Proteus* spp., *Salmonella* spp., *Pseudomonas* spp., and *Pasteurella multocida*
[3]Each sample could contribute up to 2 isolates; samples where ≥ 3 samples were isolated were considered contaminated. Contaminated and culture-negative samples were not considered when determining percentage of isolates.

Table 2-6. Quarter-level prevalence (%) of intramammary infection in heifers and multiparous cows in 4 regions across Canada in the first 30 DIM stratified by pathogen group.

Pathogen Type	Parity	Alberta (95% CI)	Ontario (95% CI)	Quebec (95% CI)	Atlantic (95% CI)	All (95% CI)
Contagious[1]	1	2.6 (1.1-5.8)	1.9 (0.8-3.9)	1.9 (0.9-3.9)	2.9 (1.2-6.9)	2.0 (1.2-3.4)
	>1	2.9 (1.6-5.0)	2.4 (1.4-3.9)	3.9 (2.5-5.9)	1.2 (0.5-2.3)	2.4 (1.6-3.6)
Environmental[1]	1	1.8 (0.7-4.4)	2.9 (1.5-5.6)	1.7 (0.8-3.7)	1.5 (0.5-4.2)	2.0[2] (1.2-3.3)
	>1	3.4 (2.0-5.8)	2.7 (1.6-4.3)	2.6 (1.6-4.3)	4.4 (2.7-7.0)	3.0[2] (2.1-4.4)
NAS[1]	1	33.1 (22.2-45.1)	40.2 (31-49.9)	20.9[3] (13.4-29.5)	9.0[4] (0.7-19.6)	26.4[2] (21.2-32)
	>1	25.1 (17.8-33.2)	26.2 (20.3-32.5)	19.5 (14.2-25.3)	3.3 (0.0-8.7)	18.4[2] (14.8-22.2)
Other[1]	1	11.6 (7.3-18.1)	6.5[3] (4.1-10.0)	5.2[4] (3.2-8.4)	4.1[4] (2.0-7.9)	6.3 (4.7-8.3)
	>1	10.6 (7.3-15.1)	7.3 (5.3-10.0)	3.7 (2.5-5.4)	3.6 (2.3-5.6)	5.7 (4.5-7.1)
Overall	1	41.5 (29-54)	49.4 (39.4-59.2)	24.2[4] (14.9-33.9)	13.2[4] (1.4-26.4)	33.4[2] (27.3-39.6)
	>1	38.4 (30.0-46.8)	32 (25.6-38.5)	25.8 (19.7-32.2)	7.7 (1.3-14.7)	25.8[2] (21.5-30.1)

[1]Contagious = *Staphylococcus aureus, Streptococcus dysgalactiae*; environmental = *Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*; NAS = non-*aureus* staphylococci; other pathogens = other bacteria, yeasts, and fungi.
[2]Difference ($P < 0.05$) between heifers and multiparous cows within the same pathogen type and region.
[3]Tendency for difference ($0.05 < P < 0.10$) from Alberta within same pathogen type and parity.
[4]Different ($P < 0.05$) from Alberta within parity.

Table 2-7. Overall and pathogen-specific prevalence (%) of quarter-level intramammary infection in first-lactation heifers in the first 30 DIM stratified by region and bulk milk SCC (BMSCC) from 91 Canadian dairy herds.

	No. Herds	Overall (95% CI)	Contagious[1] (95% CI)	Environmental[1] (95% CI)	NAS[4] (95% CI)	Other[5] (95% CI)
Low[2]	17	28.9[3] (15.6-42.9)	4.6[3] (2.2-9.2)	1.3 (0.4-3.6)	18.4 (8.2-30.6)	6.8 (3.8-11.9)
Alberta	5	30.9 (10-53.3)	3.9 (1-13.3)	0.7 (0-6.7)	31.3 (13.1-53.2)	7.3 (2.7-18.4)
Ontario	5	43.9 (22.2-65.1)	3.7 (0.9-12.3)	2.1 (0.4-8.8)	30.5 (12.3-52.5)	8.2 (3.0-20.5)
Quebec	3	30.2 (5.1-57.6)	3.7 (0.6-16.1)	1.1 (0-10.6)	15.1 (0.0-39.5)	13.0 (4.9-30.4)
Atlantic	4	19.1[3] (0-44.7)	10.8 (3.2-30)	2.0 (0.2-11)	1.9[4] (0.0-21.7)	3.6[5] (0.9-13.1)
Medium[2]	40	31.5 (22.6-40.7)	1.2[4,5] (0.5-2.5)	2.0 (1.0-3.9)	28.8 (21-37.3)	5.6 (3.7-8.4)
Alberta	9	28.1 (13-44.3)	1.7 (0.4-5.1)	1.4 (0.3-5)	20.8 (8.9-35.4)	11.4 (6-20.6)
Ontario	8	46.5 (33.5-59.3)	1.8 (0.6-4.5)	2.0 (0.7-5.2)	42.9 (30.9-55.5)	5.6[3] (3.0-10.2)
Quebec	15	32.9 (15.7-51)	0.1 (0-2.9)	4.4 (1.5-11.7)	35.0 (19.7-52.3)	3.5[5,6] (1.3-9.1)
Atlantic	8	9.7 (0-28.2)	1.8 (0.3-6.5)	1.6 (0.3-6.6)	10.6 (0.0-26.3)	4.1 (1.5-10.3)
High[2]	35	37.8 (27.8-48)	2.7[5] (1.4-5)	2.3 (1.1-4.4)	27.7 (19.2-37.1)	6.9 (4.4-10.5)
Alberta	4	95.9[4,5] (70.7-100.0)	7.8 (1.7-27.7)	6.3 (1.4-23.9)	69.3[4] (42.1-92.8)	20.8[5] (8.8-41.7)
Ontario	16	58.6[5] (40.5-75.1)	2.1[4] (0.5-6.9)	6.2 (2.4-15.2)	41.6 (25.6-58.8)	7.5[3] (3.5-15.4)
Quebec	9	18.9[5] (8.1-30.5)	3.1[4] (1.4-6.5)	1.1 (0.3-3.1)	16.5 (8.0-26.7)	4.8[4,5] (2.5-8.8)
Atlantic	5	15.9[4,5] (0-42.2)	1.6[3,6] (0.1-10)	1.2 (0.0-10.1)	14.2 (0.0-37.7)	5.1 (1.5-16.4)
Total	91	33.4[5] (27.3-39.6)	2.0 (1.2-3.4)	2.0[5] (1.2-3.3)	26.4[5] (21.2-32)	6.3 (4.7-8.3)

[1]Contagious = *Staphylococcus aureus, Streptococcus dysgalactiae*; environmental = *Escherichia coli, Klebsiella* spp., *Enterobacter* spp., *Nocardia* spp., *Prototheca* spp., *Trueperella pyogenes*; NAS = non-*aureus* staphylococci; other pathogens = other bacteria, yeasts, and fungi.
[2]Low: <150,000 cells/mL; medium: 150,000 to 250,000 cells/mL; high: >250,000 cells/mL.
[3]Tendency for difference ($0.05 < P < 0.10$) between heifers and multiparous cows within the same BMSCC category and region.
[4]Different ($P < 0.05$) from Alberta herds within BMSCC category and heifers, or from low BMSCC herds within heifers.
[5]Different ($P < 0.05$) between heifers and multiparous cows within the same BMSCC category and region.
[6]Tendency for difference ($0.05 < P < 0.10$) from Alberta herds within BMSCC category and heifers; or from low BMSCC herds within heifers.

Table 2-8. Quarter-level prevalence of subclinical mastitis (SCC > 200,000 cells/mL) in herds across 4 regions in Canada.

Region	No. Herds	Parity	Prevalence
Alberta	27	1	9.8[1] (7.8-12.3)
		>1	17.5[1] (14.7-20.7)
Ontario	29	1	15.6[1,2] (12.7-19)
		>1	18.2[1] (15.3-21.4)
Quebec	27	1	13.2[2,3] (11-15.7)
		>1	18.2[3] (15.8-20.9)
Atlantic	17	1	15.9[1,2] (13.2-19.1)
		>1	15.9[1] (13.8-18.3)
Total	91	1	13.6[1] (12.3-15.1)
		>1	17.6[1] (16.2-19.1)

[1]Different ($P < 0.05$) between heifers and multiparous cows within a region.
[2]Different ($P < 0.05$) within a parity using Alberta herds as a baseline.
[3]Tendency for difference ($0.05 < P < 0.10$) between heifers and multiparous cows within a region.

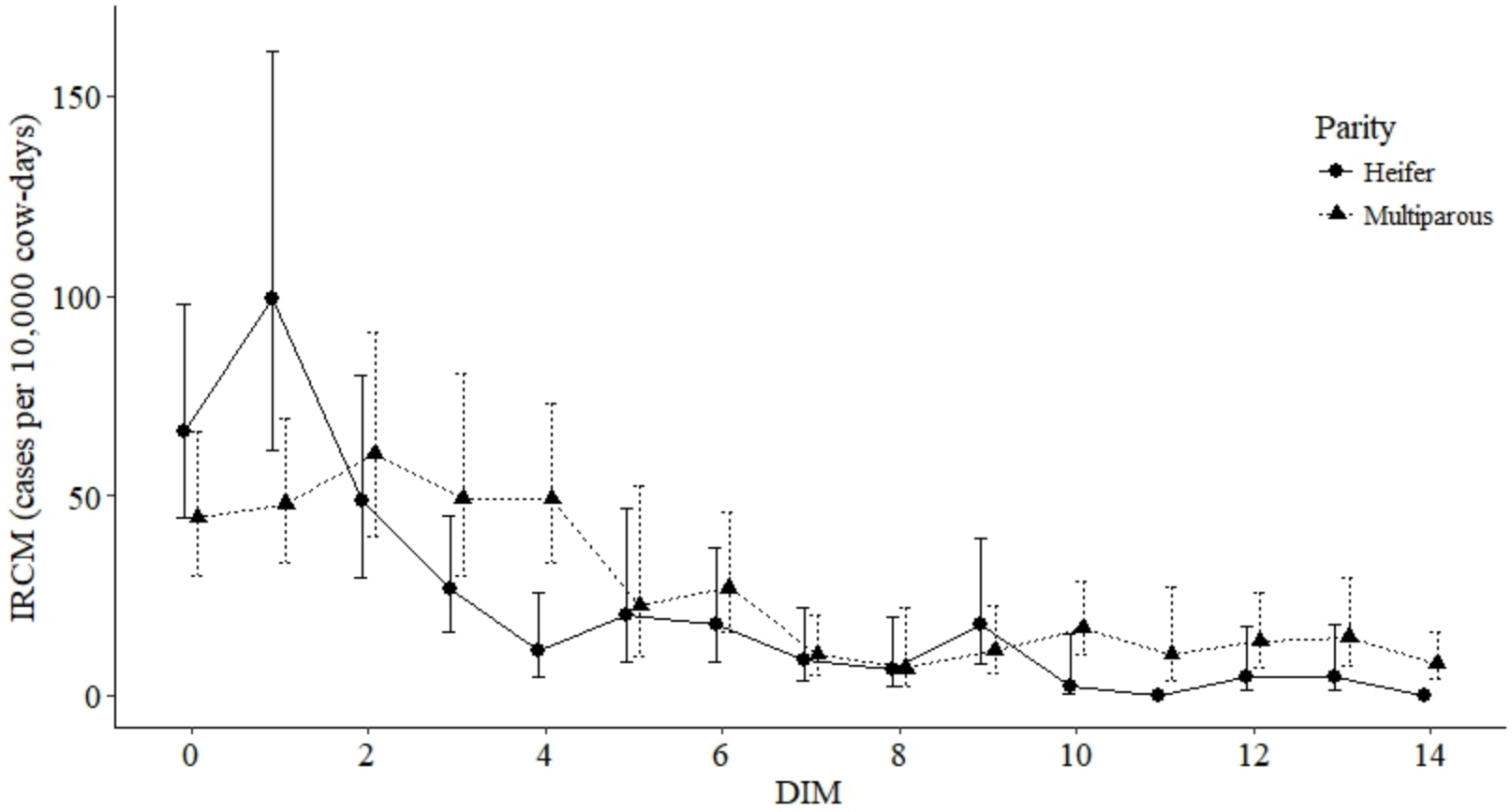


Figure 2-1. Incidence rate of clinical mastitis (IRCM) in the first 14 days of lactation in primiparous (18,157 quarter-months) and multiparous (35,936 quarter-months) dairy cows in Canada. Confidence intervals based on marginal models using generalised estimating equations.

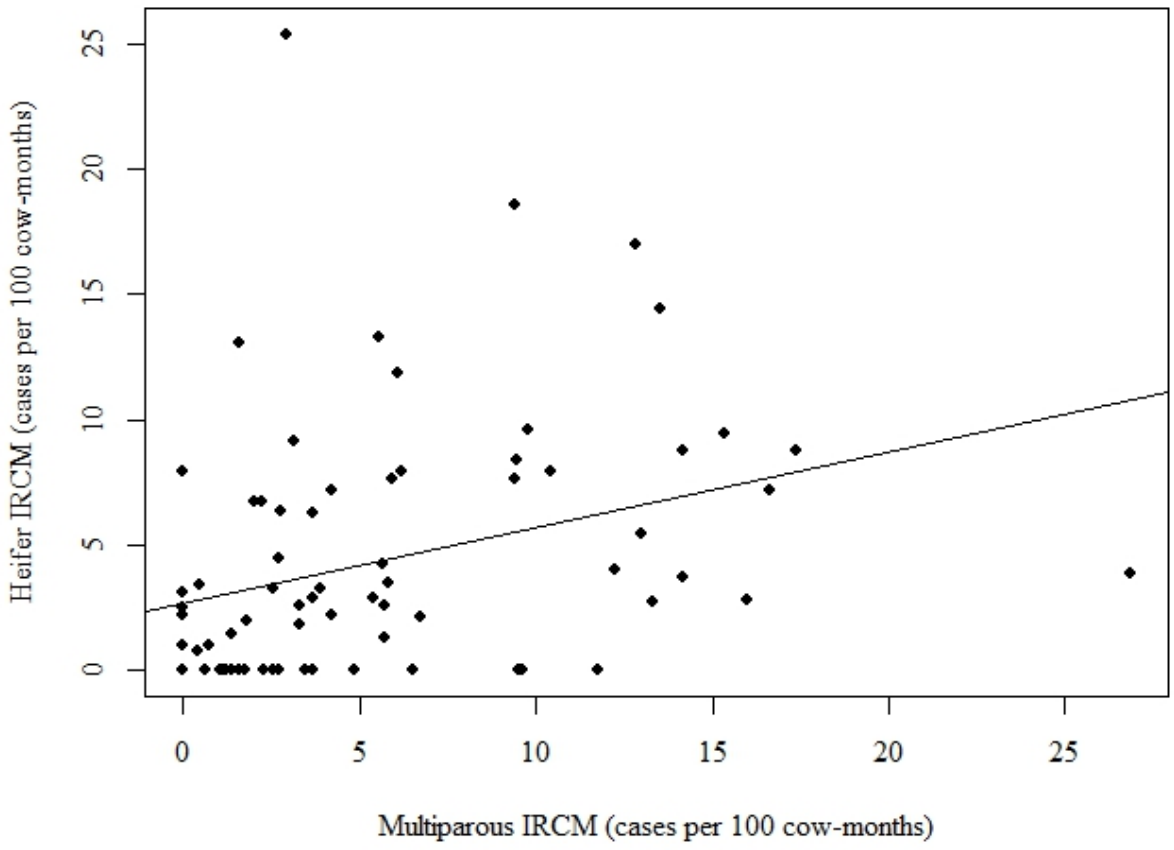


Figure 2-2. Within-herd comparison of incidence rate of clinical mastitis (IRCM) in the first 30 DIM between primiparous and multiparous cows in 75 Canadian dairy herds.

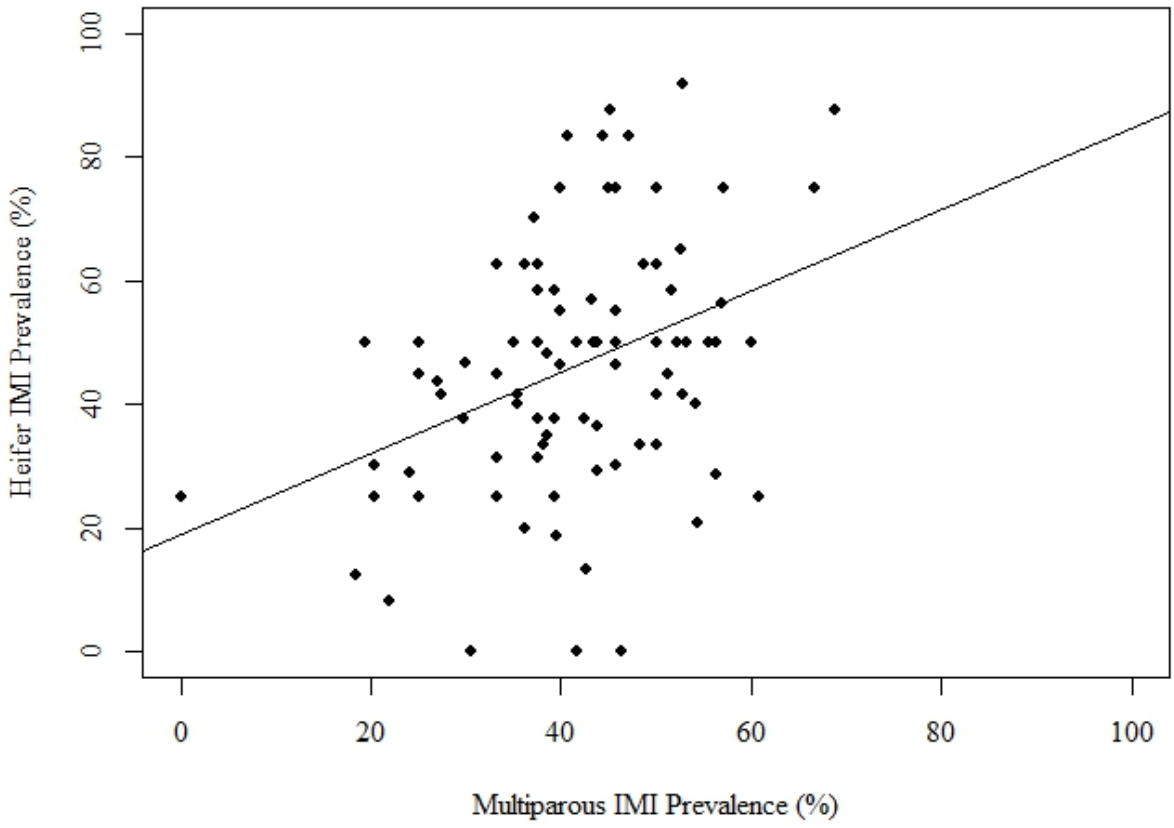


Figure 2-3. Within-herd comparison of the prevalence of intramammary infection (IMI) in the first 30 DIM between primiparous and multiparous cows in 75 Canadian dairy herds ($R^2 = 0.20$).

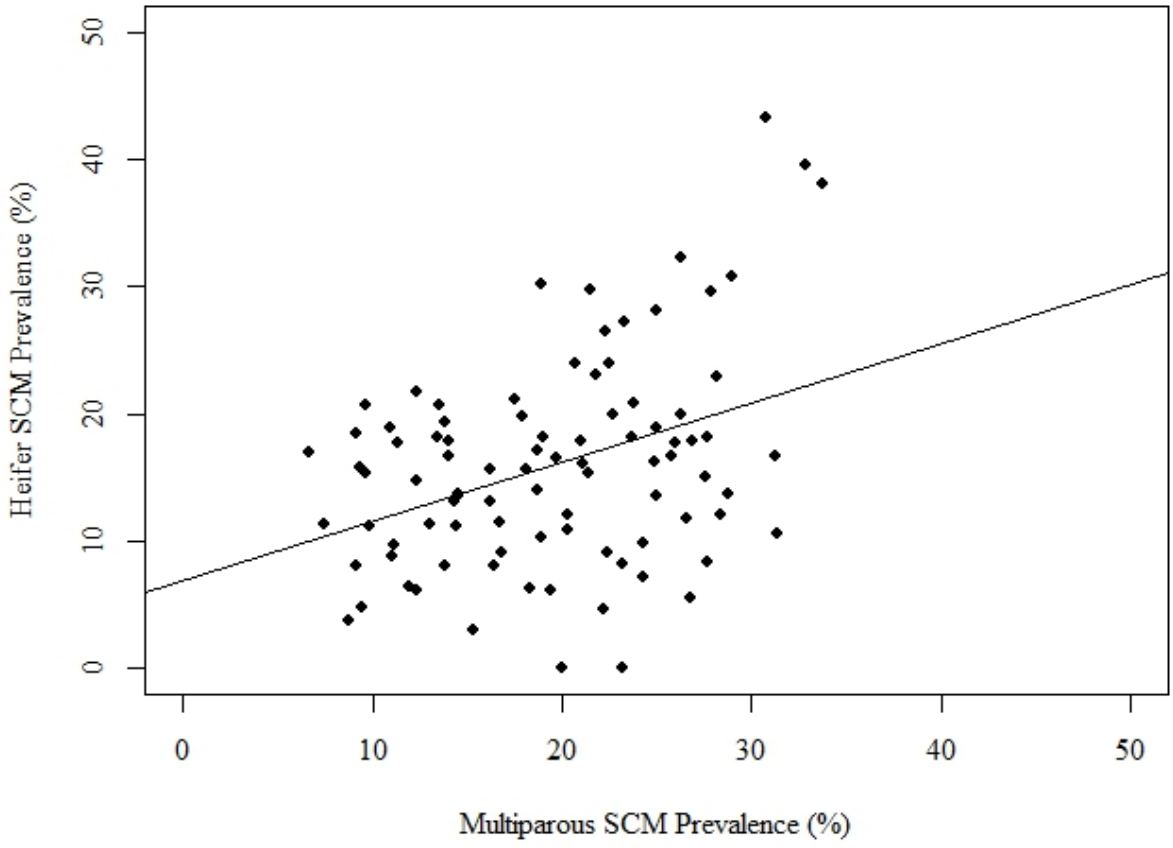


Figure 2-4. Within-herd comparison of the prevalence of subclinical mastitis (SCM) in the first 30 DIM between primiparous and multiparous cows in 91 Canadian dairy herds ($R^2 = 0.45$).

CHAPTER 3: Effectiveness of pre-calving treatment on post-calving udder health in nulliparous dairy heifers: A systematic review and meta-analysis

3.1. Abstract

Clinical mastitis affects 3% of primiparous dairy cattle (heifers) in the first month after calving. Additionally, the prevalence of intramammary infection (IMI) in the months before the first calving is high, resulting in a high prevalence of heifers calving with IMI. Pre-calving therapy is an accepted recommendation for reducing mastitis in multiparous cows, but prophylactic treatment for heifers in North America is uncommon. The objectives of this study were to: 1) quantify the change in post-calving udder health in heifers following the application of a pre-calving treatment; 2) compare effectiveness between different types of treatments; and 3) compare effectiveness of different types of treatments against specific pathogens. A systematic review was conducted comparing interventions aimed at improving udder health in heifers specifically. Of 62 included studies, 48 clinical trials were used in a meta-analysis. Data were synthesized using a random effects model for meta-analysis, followed by sub-group analyses comparing between treatment types, and specific pathogens with statistical testing using meta-regression. Occurrence of mastitis (defined as elevated somatic cell count, clinical mastitis, and IMI) was reduced in treated animals compared to untreated controls with a pooled risk ratio of treated to untreated animals of 0.56 (95% CI: 0.47 to 0.67). Upon stratification by treatment types, teat sealants and combination therapies (vaccines and antimicrobials; antimicrobials and teat sealants; all three) were most effective at improving udder health with pooled risk ratios of 0.40 (95% CI: 0.30 to 0.52), and 0.34 (95% CI: 0.25 to 0.45), respectively. Antimicrobials and

vaccines also reduced occurrence of IMI, subclinical and clinical mastitis when compared to untreated animals. Although there was variation in pathogen-specific effectiveness of treatments at reducing rates of disease, antimicrobials, teat sealants, and combinations of vaccines or teat sealants with antimicrobials were consistently effective, whereas vaccines were only effective for contagious pathogens. Recommendations for use of antibiotics should consider their relative benefit while also considering concern for increasing antimicrobial resistance.

3.2. Introduction

Mastitis is one of the most economically important diseases in the dairy industry (Seegers et al., 2003) as it has implications not only for animal health and welfare, but also for milk quality and production. The occurrence of mastitis is not only high in lactating cows, but also in heifers in the first weeks after calving (De Vliegher et al., 2012; Naqvi et al., 2017). Heifers with (sub)clinical mastitis and/or intramammary infection (IMI) are often not identified because they are not milked, and therefore their teats and milk typically are not examined (De Vliegher et al., 2012). However, the prevalence of IMI in the months before the first calving is high, resulting in a high prevalence of heifers calving with IMI (de Vliegher et al., 2012).

Although non-lactating heifers are not included in standard mastitis prevention plans like the 5-point schedule (Neave et al., 1969) or the relatively recent NMC 10-point plan (NMC, 2011), a variety of treatments to prevent or cure IMI in these heifers have been studied. Some of the commonly used pre-calving treatments are: vaccines, non-antimicrobial teat sealants, short- and long-term antimicrobials, and combinations of these treatments. Many studies have been conducted on the effect of single types of treatment in heifers (Edinger et al., 2000; Pellegrino et al., 2008; Roy et al., 2007; Santos et al., 2004), very few have compared different types of therapies, and their results have been inconsistent (McDougall et al., 2009).

Previous studies have demonstrated and quantified a net positive benefit of many different pre-calving treatments (Duplessis et al., 2014; Passchyn et al., 2014; Ataee et al., 2009), as well as characterizing the pathogen distribution of the ensuing infections. Systematic reviews have been conducted on the effectiveness of antimicrobial treatments in heifers (Nickerson, 2009), non-antibiotic strategies for prevention and control of mastitis in heifers (McDougall et al.,

2009), and vaccines for lactating cows (Middleton et al., 2009), whereas reviews of post-partum mastitis and mastitis control have also described preventive treatments in heifers and lactating cows (De Vliegher et al., 2012; Pyörälä, 2008). However, no review has been carried out that focuses on quantifying treatment effectiveness in heifers, nor comparing the efficacy of pre-calving treatment types through a meta-analysis to estimate overall effect size. The objectives of this study were therefore to: 1) synthesize results from previous studies to determine and compare the effectiveness of different types of pre-calving therapies aimed at improving udder health; and 2) determine and compare effectiveness of different types of pre-calving therapies against different types of pathogens, grouped by etiology.

3.3. Material and Methods

This systematic review and meta-analysis was conducted following a pre-specified protocol created using the PRISMA-P guidelines (Shamseer et al., 2015).

3.3.1. Data sources and literature search

The online databases searched were CAB Abstracts, Web of Science (complete collection), AGRICOLA, MedLine, and SCOPUS, from inception until September 2017. In addition, two individuals (SAN and DBN) hand-searched the references of existing reviews on the subject. The grey literature consisted of a search through all available conference proceedings in relevant conferences, as well as websites of pharmaceutical companies. The International Veterinary Information Service (IVIS) indexes proceedings from the World Buiatrics Congress (2000 – 2014) were included. In addition, proceedings of the American Dairy Scientists Association

(ADSA; 1957 - 2017) and National Mastitis Council (NMC; 1989 - 2017) annual meetings were searched. Experts in the field were also contacted for information about potential ongoing or unpublished studies, and were identified through the review process or through direct communications at conferences (Mastitis Research Worker's meeting in November 2016, and NMC annual meeting in January 2017). No publication date or language filter was applied to the search, and the search was conducted as recommended by Egger et al. (2008).

The search strategy was split into 3 main components corresponding to the population of interest, intervention type and the outcome of the study. The population of interest included primigravid to primiparous dairy cattle, with the intervention having to be applied during pregnancy and udder health measured after calving. The interventions of interest were any kind of treatment applied pre-calving aimed specifically at improving udder health. This was intended to exclude interventions such as feed additives intended to improve overall health and growth as it did not address the primary question of targeted and individual treatment of heifers. Full details of the search terms used are provided in Appendix 1.

3.3.2. Study eligibility and selection

An initial screen of all titles was performed by two individuals (SAN and DBN) to determine if manuscripts were eligible for full text review for inclusion in the systematic review. If the title was not detailed enough to decide about moving forward with full-text screening, the abstract was reviewed by one of the reviewers (SAN) to determine whether it was eligible. A manuscript needed to meet 2 conditions to be considered for a full text review: 1) The manuscript had to report on primary data; and 2) the manuscript had to report on an intervention aimed at improving udder health (no vaccines for other conditions such as leucosis, or treatment

with anthelminthic products), although it may not necessarily be the primary aim of the study (comparing mastitis in 2 regions where a particular product may not be allowed).

This initial screen was designed to be fairly broad to encompass as many possibly relevant manuscripts for full text review. A full text review was then conducted on all of these screened manuscripts to ensure that some aspect of intervention is addressed in the manuscript. The reason for this full text screening step in the systematic review was that the second condition described above could have resulted in manuscripts whose primary goal was not assessing intervention effectiveness, but was included as a covariate in their analysis. As it was not the primary goal of the study, it may not have been reported in the abstract and a full text screening would capture these manuscripts.

3.3.3. Identification of manuscripts for eligibility in the systematic review and meta-analysis

The same two individuals (SAN and DBN)) reviewed full text manuscripts identified after the 2 screening stages described in the previous section using the following criteria to select manuscripts eligible for inclusion in the systematic review and meta-analysis:

Population	=	Primigravid/primiparous dairy cattle (if the primary study population was not primiparous dairy cattle, the manuscript must stratify and report on them separately from older animals)
Intervention	=	Any sort of treatment given after confirmed pregnancy, prior to calving or on the day of calving that was aimed specifically at improving udder health. Studies that did not assess treatment effectiveness as a primary goal (described in the previous section) were excluded from the systematic review and meta-analysis as

		they would not be designed around that variable and pooling would not be appropriate.
Comparison	=	A treatment must be compared to either another treatment, or no treatment at all.
Outcome	=	Post-calving prevalence or incidence of 1) clinical mastitis (CM); 2) subclinical mastitis (SCM; elevated SCC); or 3) IMI (presence of bacteria in the milk) reported within 2 mo after first calving.
Study Design	=	Randomized controlled trials or observational cohort studies.

The screening of manuscripts was designed to capture studies comparing pre-calving treatments in heifers to each other or to untreated heifers in their association with incidence of udder disease post-calving. Only randomized controlled trials were meta-analyzed, whereas results from observational cohort studies were included descriptively in discussion. Agreement between reviewers (SAN and DBN) was quantified using a kappa score and percent agreement, and any disagreements were resolved by the last author (HWB).

3.3.4. Study quality assessment and data extraction

Study quality and risk of bias was assessed using the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) tool (Hooijmans et al., 2014). The SYRCLE tool is based on the Cochrane Risk of Bias Tool (Higgins et al., 2011), with some items modified or added to fit better with risk of bias in experimental trials in animals. Briefly, the SYRCLE tool consists of 10 items, corresponding to 5 different types of bias: selection, performance, detection, attrition, and reporting bias. Risk of selection bias is evaluated through considering allocation

sequence generation, subject baseline characteristics and allocation concealments. Performance bias is evaluated by considering random housing of animals and blinding when assigning treatments. Detection bias is assessed by evaluating randomization of subjects during outcome assessment and blinding of outcome assessors. Attrition bias is assessed by evaluating the study's reporting on incomplete outcome data. Lastly, reporting bias is assessed by evaluating the outcomes that the studies reported on, as well as outcomes that were excluded. A study was considered to be at high risk of a given type of bias if it had failed to address over half of the sources of error contributing to that given type; moderate risk of bias if they failed to address at least one, but not more than half; and low or no risk of bias if they did not have any sources of error from a given type of bias.

Data were extracted using a single electronic form using Microsoft Excel (Microsoft Corp., Redmond, WA). Study characteristics extracted were: 1) country of study; 2) year of publication, 3) breed, 4) intervention type(s); and 5) time of intervention application. As management decisions are based on anticipated calving dates, the intended or median of intended application times was recorded in the extraction form (e.g. for 2 to 3 wk pre-calving, the number taken was 17.5 d). Outcome characteristics recorded, often multiple for a study, were 1) type of outcome defined as cure rate of existing IMI, SCM, IMI (and its definition), and CM; 2) time of outcome measurement or follow-up in the case of CM; and 3) numbers of individuals (animals or quarters depending on the study) and cases per treatment arm of the study. For studies where post-calving samples were collected on multiple days, the earliest sample that was not the day of calving was used for outcome extraction. Numbers in treatment arms were pooled by intervention or control and type of intervention depending on possible stratification in meta-analyses.

3.3.5. Statistical analyses

All statistical analyses were performed using the "metafor" package (Viechtbauer, 2010) in R version 3.4.2 (R Core Team 2017). For all analyses, a P-value < 0.05 was considered significant, and a 0.05 < P-value < 0.10 was considered a tendency for significance. Results from randomized clinical trials were pooled using a random-effects meta-analysis as no observational studies were included in the systematic review. Randomized clinical trials that did not report numbers of individuals and only ratios were not included in the meta-analysis and were included only descriptively. A mixed effects model was selected given the underlying heterogeneity in intervention and population characteristics, and the fact that studies often had multiple treatment types and outcomes. Effect sizes were pooled as risk ratios as it is most intuitively interpretable in terms of showing a benefit/harm. All pooled estimates and stratified meta-analyses were visualized using forest plots, and homogeneity was assessed using the Cochrane Q statistic for significance and I^2 statistic to quantify the amount of variation due to study heterogeneity. The primary question concerning therapy type was statistically tested using meta-regression, and heterogeneity was then explored by stratifying on relevant variables of practical interest: type of outcome and bacterial species of infection. Variation in effectiveness by pathogen type was addressed through a combination of meta-regression and subgroup analyses. Sensitivity and specificity to detect IMI varies considerably between IMI definitions (Andersen et al., 2010; Dohoo et al., 2011), so a meta-regression was conducted to determine whether effect size estimates with rates of IMI differed between studies that described their IMI definition and those that only provided details for culturing. Questions related to effects of study quality or design (abstract vs. full-text, disease definitions) were also addressed using meta-regression. Publication bias for the primary outcome of interest (mastitis of any form) was assessed using a funnel plot

of the risk ratio with the SE of the risk ratio. Asymmetry was tested using Egger's regression test for small study effects and visualized with a graph of study variance against the natural log of the risk ratio.

3.4. Results

3.4.1. Description of studies

Search results. Numbers of studies included at different stages of the review process are presented in Figure 3-1. Briefly, a total of 974 unique records were identified, with most of them coming from the database search (n = 1,245 before duplicate removal) and a small number being identified through the described grey literature review (n=44 before duplicate removal). After screening abstracts, 82 peer-reviewed manuscripts were identified for a more thorough full-text review, along with 31 records from conference proceedings. A small number of non-English manuscripts were identified: Portuguese (n=1), German (n=3), Dutch (n=1), Czech (n=1) and Italian (n=1). These manuscripts were reviewed using the pre-specified criteria by colleagues with fluency in the respective languages.

Studies eligible for inclusion in the systematic review. Forty studies from published manuscripts were included in the systematic review, of which 38 were eligible for inclusion in the meta-analysis. All included peer-reviewed manuscripts reported results from clinical trials; 2 of them could not be included in the meta-analysis as they did not report numbers of outcomes, rather relative measures such as risk ratios and odds ratios. In addition, 22 studies from conference proceedings or abstracts were included in the systematic review. Ten conference proceedings provided enough detail in their results to be eligible for meta-analysis, whereas the

others are discussed only descriptively. Records from 48 studies were included in the meta-analysis, although multiple records were included for many studies since they compared multiple treatment types (Table 3-1). A detailed description of the specific products used in each study is provided in Appendix 2.

Risk of bias in studies included. Design characteristics of all full-text manuscripts that are important for risk of bias are presented in Table 3-2. Based on the SYRCLE guidelines all studies had at least one type of bias that they were at high risk for. The majority of studies were at moderate risk of selection bias (n=27), with 2 being at low/no risk of selection bias. Allocation sequence generation was often systematic rather than completely random, and baseline animal characteristics were described in most studies (n=30). The largest type of bias that this pool of studies was at risk for was detection/performance bias because most did not describe blinding or randomized animal housing. Most (n=31) studies adequately addressed incomplete data by providing numbers of animals at each stage of the study, with reasons provided for their removal or exclusion from the trial minimizing risk of attrition bias. No studies were at risk of reporting bias, as they all either provided justification for the outcomes measured, and they were selected in part based on their reporting of outcomes.

Another aspect of variability between studies and a possible source of bias was how they defined outcomes (Table 3-3), IMI in particular. Many studies (n=16) only described culturing details for their milk samples, without a strict definition of what was considered an IMI. Risk ratio for IMI did not differ between treated and untreated animals when comparing the two groups of studies ($P = 0.93$). Definition of IMI was not used in any subsequent analyses, or considered for stratification.

3.4.2. Effectiveness of treatment

Overall effect size. The overall risk ratio of udder disease of any sort in treated compared to untreated heifers was 0.60 (95% CI: 0.47 to 0.67), demonstrating a reduction in udder diseases after calving in treated compared to untreated heifers ($P < 0.0001$). Effect size estimates from abstracts/proceedings tended to be slightly lower and had more variance than estimates from full-text manuscripts ($P = 0.08$; Figure 3-2), but their inclusion did not have an impact on overall effect size (RR = 0.60 for full-text manuscripts alone; RR = 0.57 for a combined meta-analysis). All further analyses were conducted on results from both full-text manuscripts and abstracts/conference-proceedings without stratification.

Effect size by outcome type. The majority of records (n=34) used IMI as the outcome, whereas 15 records assessed CM and 2 assessed SCM. Effect size estimates for CM tended to be lower than those for IMI ($P = 0.07$), but tended to be higher than those for SCM ($P = 0.06$; Table 3-4). For subsequent meta-analyses and regressions all outcomes other than cure rate were combined through a random-effects term included for outcome type within studies to account for any within-study correlation. Risk of cure of existing IMI following pre-calving treatment with antimicrobials was higher than spontaneous cure with no treatment (RR = 1.63 (95% CI: 1.19 to 2.25); $P = 0.003$).

Effect size by therapy type and pathogen type. A meta-regression on treatment type using CM, SCM and IMI as outcomes resulted in a difference in effectiveness between treatment types (Table 3-5). Teat sealants and combination therapies were the most effective treatments ($P < 0.0001$), although antimicrobials, vaccines and other treatments still reduced risk of CM, SCM or IMI over untreated controls ($P < 0.0001$). Teat disinfection before calving did not reduce udder disease compared to untreated controls. Records not included in the meta-analysis also reported

similar patterns – antimicrobials were effective at preventing CM and IMI, as well as curing existing IMI. Farms observed longitudinally had decreasing prevalence of IMI and incidence of CM around calving after incorporating pre-calving antimicrobial treatment into heifer management (Acuna et al., 2003). Other effective management strategies were the use of physical intramammary devices (Nickerson 1990), implementing fly control for heifers on pasture (Nickerson 1994), and earlier removal of calves after calving to prevent suckling (Compton 2008).

Although studies differentiated culture results to different levels of detail, pathogens were grouped into 3 categories – contagious (*Staphylococcus aureus, Streptococcus agalactiae*), environmental (*Escherichia coli,* non-*agalactiae* streptococci), and non-*aureus* staphylococci. Subgroup analyses including meta-regression on treatment type were conducted using the 3 categories as outcomes (Table 3-6). Antimicrobials, teat sealants, and vaccines all presented similar reductions (RR = 0.51, 0.40, and 0.56, respectively) in contagious pathogen CM and IMI compared to untreated controls. Only one single study reported on teat dips/sprays (Edinger et al., 2000) and one on combination therapies (McDougall et al., 2008); in both studies, the 2 therapies did not change occurrence of CM or IMI compared to controls. Manuscripts not included in the pooled estimate also reported a reduced incidence of CM and IMI by contagious pathogens when animals received pre-calving treatment with antimicrobials, and also when fly-control was implemented for pre-calving heifers on pasture.

Teat sealants and combination therapies reduced occurrence of CM and IMI with environment pathogens over untreated controls (RR = 0.27 and 0.25, respectively; Table 3-6). Antimicrobials, teat disinfection and vaccines did not have any benefit over untreated controls, but 2 studies not included in the meta-analysis (Nickerson et al., 1994; Tenhagen et al., 2001)

reported a reduction of clinical signs during *E. coli* IMI following pre-calving vaccination. Fly control for heifers on pasture reduced *Streptococcus* spp. IMI. Studies reporting on CM and IMI where CNS was isolated reported a reduction compared to controls in animals treated with antimicrobials, teat-sealants, and combination therapies, with combination therapies showing the greatest reduction (RR = 0.34), and similar conclusions were drawn from studies not included in the meta-analysis. Vaccines did not reduce CM and IMI with CNS occurrence compared to untreated controls. Due to only single studies looking at other therapies and teat dips/sprays, pooling was not possible, but neither demonstrated a reduction in occurrence of CM and IMI with CNS compared to untreated controls.

3.4.3. Meta-analysis statistics

Heterogeneity. Significant heterogeneity was present in all analyses. The mixed effects meta-analysis using all records with random effects for outcome type and intervention type had an overall I^2=0.87 (Q = 439.28, P <0.0001). Introducing outcome type as a moderator did not reduce heterogeneity noticeably; I^2=0.85 (Q = 343.54, P < 0.0001). Conducting a meta-regression using outcome type also did not affect heterogeneity; I^2=0.85 (Q = 369.78, P < 0.0001). Sub-group analyses by pathogen type led to reductions in residual heterogeneity, except for the CNS analysis: contagious pathogen analysis I^2=0.29 (Q = 39.36, P = 0.03); environmental pathogen analysis I^2=0.57 (Q = 71.77, P < 0.0001); CNS analysis I^2=0.88 (Q = 212.71, P < 0.0001).

Publication bias. A visual assessment of the funnel plot of SE compared to the effect size estimate (Figure 3-3) from each study suggests some asymmetry due to missing studies showing an increased risk following treatment with small sample sizes (bottom right quadrant). Egger's

regression small study effects indicated a possibility of publication bias as increasing standard error was associated with a larger effect size estimate ($P < 0.0001$; Figure 3-4).

3.5. Discussion

Management of mastitis on dairy farms often starts after the heifer's first calving, too often resulting in less than ideal udder health of first lactation heifers (De Vliegher et al., invited review). Many housing and management practices associated with udder health of older cows are also associated with improved udder health in primiparous heifers (De Vliegher et al., 2012), but the level of implementation in non-lactating heifers is unknown. Prophylactic treatments of the udder before calving such as antimicrobial dry cow therapy or using internal teat sealants are widely adopted for use in the lactating herd (e.g. Olde Riekerink et al., 2010), but these treatments are less commonly used in primiparous heifers. This systematic review and meta-analysis of studies assessing effectiveness of pre-calving therapies for improving post-calving udder health in heifers demonstrates that there is an overall benefit to udder health in heifers when using any kind of pre-calving prophylactic treatment. The treatments vary in both cost to purchase as well as costs and labour associated with application, so an understanding of the effectiveness of treatments is important to inform management decisions. The effectiveness of treatment varied considerably and was also dependent on the type of bacteria causing the IMI.

This is the first review to meta-analyze results from studies of multiple types of pre-calving therapies in primiparous heifers. Recent reviews have looked at heifer mastitis as a whole (de Vliegher et al., 2012), antimicrobial treatment to control heifer mastitis (Nickerson, 2009), vaccination efficacy against staphylococcal infections in all lactating cows (Middleton et

al., 2009), and the effectiveness of treatments other than conventional antimicrobials against CM in all lactating cows (Francoz et al., 2017). The distribution of pathogens in IMI of heifers differs from older cows (Piepers et al., 2009; Supré et al., 2011; Naqvi et al., 2017), as well as their immune response to infection (de Haas et al., 2002; de Haas et al., 2004), so results from studies on older lactating animals may not be generalizable to heifers. A synthesis of studies aimed at heifers in particular is therefore necessary when trying to implement more targeted management strategies. Findings from this review are in line with most of the individual studies and reviews on lactating cows that demonstrate a net positive benefit of any type of treatment for improving udder health. Previous reviews have demonstrated differential effectiveness of antimicrobials (Barkema et al., 2006; Nickerson et al., 2009), and teat sealants (Crispie et al., 2004) based on the pathogen when used in lactating cows, whereas vaccine development tends to focus on a specific type of pathogen such as *Staphylococcus aureus* (Middleton et al., 2009) or *Escherichia coli* (Enviracor® and StartVac®). Comparing with the present review focusing on heifers specifically, variation in treatment effectiveness by pathogen type is similar to what has been reported from studies focusing on older cows.

As mentioned above, management of non-lactating heifers is less than ideal on an average dairy farm (Parker et al., 2007; McDougall et al., 2009). This could be an important reason why the treatments proved to be effective in improving udder health. It is therefore likely that the effectiveness of the treatments, particularly antimicrobial treatment, will be less on farms that pay more attention to the non-lactating heifers.

Although the effectiveness of antimicrobials was consistent across Gram-positive pathogens, their use needs to be balanced with concerns about the rising prevalence of antimicrobial resistance (Barkema et al., 2015; Tang et al., 2017). Other treatments can be just

as, if not more, effective than antimicrobials in heifers, and current research is beginning to focus on their effectiveness in older cows (Francoz et al., 2017; Ruiz et al., 2017). Antimicrobial treatment of non-lactating heifers improved udder health when contagious (Gram-positive) bacteria caused the udder health problems after calving. In contrast, antimicrobial treatment did not improve udder health when environmental pathogens were involved. It is therefore important when udder health problems after calving are an issue on a farm and antimicrobial treatment is considered, to culture heifers after calving. In addition to selecting the best treatment for a given farm, personnel must also be properly trained in its application. For example, improper application of intramammary treatments either by causing trauma after inserting the cannula too far, or using contaminated equipment can actually increase a quarter's risk of infection (du Perez 2000). It is difficult to quantify the impact of iatrogenic infections in studies included in this review since although some mentioned disinfection/preparatory procedures for application, no studies detailed training for personnel in treatment application.

In comparing all types of treatments for heifers, this review faced the challenge of significant between-study variability not only in treatments used, but also in other aspects of study design and setting. Although we tried to explore heterogeneity through subgroup analyses, there were still many other study characteristics that varied considerably to allow for valid subgroup analyses such as timing of treatment application, study location, timing of data collection and geographic location. As a result, significant residual heterogeneity was present even after stratification. Although the significant between-study variability may cast some doubt on the precision of the pooled results, the overall message remains unchanged as the direction of the effect remained consistent between studies, even if the effect size did not: pre-calving treatment of heifers improves udder health after calving, and the degree of improvement varies

by treatment type and pathogen. Variation in characteristics such as breed, geographic location and sampling season or timing of treatment application and specific treatment are inherently linked, making it difficult to be consistent across studies from different countries, and studying different treatments.

An assessment of the risk of bias focuses on characteristics of study design and reporting in a way that is less dependent on factors outside of the researchers' control. The assessment using the SYRCLE tool (Hooijmans et al., 2014) highlighted some areas with significant risk of bias. Although randomization of housing and outcome assessment order are less likely to contribute to detection or performance bias in livestock studies (SYRCLE was designed for lab trials using animals), blinding of both treatment allocator or outcome assessor was not mentioned in most of the studies, nor was allocation concealment. For some outcomes like SCM or IMI, blinding may be inherent since the milk samples are collected by farm personnel and the outcome (SCC or bacterial culture) is assessed by lab technicians. Despite the possible inherent blinding of outcome assessors, lack of blinding of the treatment allocator can still contribute to detection bias, and both types of blinding should be reported, with justification provided if blinding was not done. Similarly, randomization and allocation sequence generation were only described in 6 studies, and presented a possible source of selection bias. Clear outcome definitions, especially for IMI were also inconsistently reported, where different definitions could contribute to residual heterogeneity. Although just defining IMI did not change effect size from studies that did not, looking at specific definitions would likely provide different results as each one would have different operating characteristics (Dohoo et al., 2011a; Dohoo et al., 2011b). This likely is not a reflection of study quality, but more on inconsistency in reporting standards for studies on livestock.

In 2010 a collaboration between many livestock researchers established the Reporting Guidelines for Randomized Controlled Trials for Livestock (REFLECT; O'Connor et al., 2010) to address this issue. Although certain journals have incorporated these into their requirements for authors (Preventive Veterinary Medicine, Journal of Veterinary Internal Medicine, Journal of Swine Health and Production), these guidelines are far from universally incorporated. The small number of journals implementing these guidelines, combined with the fact that many of the studies included were conducted prior to 2010, is likely responsible for the inconsistencies highlighted in this review. Although this review reports on a substantial number of studies, an updated review in the future once the REFLECT statement is more widely implemented would likely show more consistency between studies, leading to more precise pooled estimates of protective effect size.

As the first review to focus on the effectiveness of all existing pre-calving therapies at improving udder health in primiparous dairy heifers, the present study effectively synthesized results from a large number of studies. Heterogeneity was explored through subgroup analyses and meta-regression, focusing primarily on the main goals of the study: understanding variation in therapy success by treatment type and pathogen. Despite this exploration of heterogeneity, even stratification by both treatment type and pathogen still showed significant residuals and unexplained study variability. As discussed, this is likely a reflection of the differences in study settings, design characteristics and reporting. Although this degree of study variability has an impact on the precision of pooled estimates, the direction of the effect clearly shows that there is a positive impact on udder health when pre-calving treatments are applied. Publication bias in a review is another aspect that could impact the reliability of findings, and Egger's regression indicated that there were significant small-study effects in the review. An examination of the

funnel plot shows that there were no studies found showing a negative impact of pre-calving treatment and used a small sample size. Although this may be a reflection of publication bias, in this case it is more clearly reflective of the true heterogeneity in the included studies, attributed primarily to treatment intensity and differences in underlying risk (Egger et al., 2008). Treatments varied considerably in timing of application and dosage, both of which would increase variability between studies. The differences in geographical location, impacting housing, nutrition and other management factors, and animal breed could contribute to different regional risks for mastitis, or differences in susceptibilities of the animals to disease. Asymmetry in the funnel plot may also be an artefact (Egger et al., 2008) as the majority of interventions under study were very unlikely to increase the risk of udder diseases over untreated animals, making it very unlikely that small studies would show harm to udder health by pre-calving treatments. Although there was statistical evidence suggesting publication bias, this was more likely a reflection of study variability as quantified by the residual heterogeneity rather than a bias in study selection.

3.6 Conclusions

This is the first review of pre-calving treatments for improving udder health in primiparous dairy heifers assessed the findings of 60 studies, 48 of which were meta-analyzed for pooled estimates. Overall any type of pre-calving therapy improved udder health by reducing the occurrence of IMI, CM and SCM compared to untreated controls. Variation in success by therapy type demonstrated that studies assessing antimicrobials, teat sealants, vaccines and combination therapies demonstrated a significant reduction in the risk of udder disease, whereas

those looking at teat dips/sprays or other types of therapy did not reduce risk compared to untreated animals. There was also significant variation by pathogens that the studies examined: antimicrobials, teat sealants, vaccines and combination therapies were effective at reducing risk of udder disease caused by contagious pathogens; antimicrobials, teat sealants and vaccines were effective at reducing risk of udder disease caused by environmental pathogens; antimicrobials, teat sealants, vaccines and combinations were effective at reducing udder disease caused by coagulase-negative staphylococci. Once rigid reporting guidelines are more widely implemented by journals and authors, an updated literature review on udder health of dairy heifers will produce more precise pooled estimates and should demonstrate decreased between-study variability.

3.7. References

Acuña Pasqualini, C.N., G.F. Casasnovas, and C.F. Macazaga. 2006. Effectiveness in the reduction of new intramammary infections in the dry period by the use of an external teat sealant in heifers at prepartum. *In* Proc. 45th NMC Ann. Mtg., pp. 258–259.

Acuña, C.N., G.F. Casasnovas, S. Terminiello, A. Moreno, C.G. Santarcangelo, and L.O. Simon. 2015. Use of dry cow therapy with internal teat sealant for the prevention of clinical mastitis occurring after calving in heifers with Staphylococcus aureus. *In* Proc. 54th NMC Ann. Mtg. 171–172.

Acuña, C.N., and R.E. Chertcoff. 2003. Mastitis control program on dairies with high incidence in first calving heifers: Part 2. Proc. 42nd NMC Ann. Mtg.

Alaniz, G.R., T.R. Meinert, W.J. Loveland, G. Gallo, and S.A. Salmon. 2013. Evaluation of *Escherichia coli* bacterin, J-5 strain, as an aid in the control of clinical signs associated with E. coli mastitis in dairy cows. Proc. 52nd NMC Ann. Mtg. 199–200.

Ataee, O., P. Hovareshti, M. Bolourchi, A. Barin, A. Gerami, and A. Niasari-Naslaji. 2009. Effect of systemic antibacterial administration during prepartum period on coagulase negative staphylococcal intramammary infection in Holstein heifers. Iran. J. Vet. Res. 10:255-259. doi: 10.22099/ijvr.2009.1701.

Barkema, H.W., Y.H. Schukken, and R.N. Zadoks. 2006. Invited review: The role of cow, pathogen, and treatment regimen in the therapeutic success of bovine *Staphylococcus aureus* mastitis. J. Dairy Sci. 89:1877–1895. doi:10.3168/jds.S0022-0302(06)72256-1.

Barkema, H., M. von Keyserlingk, J. Kastelic, T. Lam, C. Luby, J. Roy, S. LeBlanc, G. Keefe, and D. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–7445. doi:10.3168/jds.2015-9377.

Barto, P. B. and L. J. Bush. 1978. Effect of prepartum antibiotic infusion on mastitis infection in dairy cows at first calving. Animal Science Research Report, Agricultural Experiment Station, Oklahoma State University (MP-103):164-168.

Bassel, L., D. Kelton, A. Godkin, K. Leslie, and K. Lissemore. 2003. Risk factors for intramammary infection at first calving in heifers from Ontario sentinel herds. *In* Proc. 42nd NMC Ann. Mtg. 282–283.

Bastan, A., M. Cengiz, S. Cengiz, B. Polat, A. Colak, M. Akan, I. Darbaz, and D. B. Acar. 2010. Effects of precalving antibiotic treatment on mastitis and individual somatic cell count in heifers. J. Anim. Vet. Adv. 9:1245-1249. doi: 10.3923/javaa.2010.1245.1249.

Borm, A. A., L. K. Fox, K. E. Leslie, J. S. Hogan, S. M. Andrew, K. M. Moyes, S. P. Oliver, Y. H. Schukken, D. D. Hancock, C. T. Gaskins, W. E. Owens, and C. Norman. 2006. Effects of prepartum intramammary antibiotic therapy on udder health, milk production, and reproductive performance in dairy heifers. J. Dairy Sci. 89:2090-2098. doi: 10.3168/jds.S0022-0302(06)72279-2.

Bowers, S., S. Gandy, K. Graves, S. Eicher, and S. Willard. 2006. Effects of prepartum milking on postpartum reproduction, udder health and production performance in first-calf dairy heifers. J. Dairy Res. 73:257-263. doi: 10.1017/S0022029906001762.

Bray, D. R., F. Elvinger, R. L. de la Sota, R. P. Natzke, P. A. Reed, and J. K. Shearer. 1989. Prevalence of infection in mammary quarters of nulliparous heifers and efficacy of intra-mammary infusion of antibiotics three weeks before parturition. J. Dairy Sci. 72(Supp.1):20.

Bryan, M. and K. Taylor. 2009. Periparturient use of parenteral micronised procaine penicillin to reduce the risk of clinical mastitis in heifers after calving. Vet. Microbiol. 134:143-149. doi: 10.1016/j.vetmic.2008.09.021.

Bryan, M. A. and G. M. Friton. 2005. Stochastic economic modeling of the use of penethamate hydriodide (Mamyzin®) in heifers around calving to control peri-parturient mastitis. Pages 232-234 in Mastitis in Dairy Production: Current Knowledge and Future Solutions. Wageningen Academic Publishers, Wageningen, The Netherlands.

Bush, L. J., P. B. Barto, and G. D. Adams. 1977. Effect of teat dipping on mastitis infection in dairy heifers at 1st calving. J. Dairy Sci. 60:167-167.

Compton, C., and S. McDougall. 2008. Effect of early milking of calved heifers and selenium supplementation on incidence of clinical mastitis in dairy heifers. *In* Proc. of the 25th Jubilee World Buiatrics Congr., p. 48.

Compton, C. W. R., F. R. Emslie, and S. McDougall. 2014. Randomised controlled trials demonstrate efficacy of a novel internal teat sealant to prevent new intramammary infections in dairy cows and heifers. New Zealand Vet. J. 62:258-266. doi: 10.1080/00480169.2014.898201.

Contreras, G.A., and P.M. Sears. 2008. Use of systemic antibiotic therapy in prepartum heifers to reduce the incidence of intramammary infections. *In* Proc. 47th NMC Ann. Mtg. 252–253.

Contreras, G. A., J. David Munoz, and P. M. Sears. 2013. Reducing the incidence of intramammary infection in heifers by using prepartum systemic tylosin therapy: Initial results of a single herd pilot study. Rev. Colomb. Cienc. Pec. 26:119-126.

Crispie, F., J. Flynn, R.P. Ross, C. Hill, and W.J. Meaney. 2004. Dry cow therapy with a non-antibiotic intramammary teat seal - a review. Ir. Vet. J. 57:412-418. doi:10.1186/2046-0481-57-7-412.

De Haas, Y., H.W. Barkema, and R.F. Veerkamp. 2002. The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. J. Dairy Sci. 85:1314–1323. doi:10.3168/jds.S0022-0302(02)74196-9.

De Haas, Y., R.F. Veerkamp, H.W. Barkema, Y.T. Gröhn, and Y.H. Schukken. 2004. Associations between pathogen-specific cases of clinical mastitis and somatic cell count patterns. J. Dairy Sci. 87:95–105. doi:10.3168/jds.S0022-0302(04)73146-X.

De Vliegher, S., L.K. Fox, S. Piepers, S. McDougall, and H.W. Barkema. 2012. Invited review: Mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. J. Dairy Sci. 95:1025–1040. doi:10.3168/jds.2010-4074.

De Vliegher, S., H. Laevens, H.W. Barkema, I.R. Dohoo, H. Stryhn, G. Opsomer, and A. de Kruif. 2004. Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers. J. Dairy Sci. 87:937–947. doi:10.3168/jds.S0022-0302(04)73238-5.

Dohoo, I.R., J. Smith, S. Andersen, D.F. Kelton, and S. Godden. 2011. Diagnosing intramammary infections: evaluation of definitions based on a single milk sample. J. Dairy Sci. 94:250–61. doi:10.3168/jds.2010-3559.

Dohoo, I., S. Andersen, R. Dingwell, K. Hand, D. Kelton, K. Leslie, Y. Schukken, and S. Godden. 2011. Diagnosing intramammary infections: comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94:5515–5522. doi:10.3168/jds.2011-4486.

Duplessis, M., C.L. Girard, D.E. Santschi, J.P. Laforest, J. Durocher, and D. Pellerin. 2014. Effects of folic acid and vitamin B12 supplementation on culling rate, diseases, and

reproduction in commercial dairy herds. J. Dairy Sci. 97:2346–2354. doi:10.3168/jds.2013-7369.

Edinger, D., B. A. Tenhagen, P. Kalbe, G. Klunder, B. Baumgartner, and W. Heuwieser. 2000. Effect of teat dipping with a germicide barrier teat dip in late gestation on intramammary infection and clinical mastitis during the first 5 days post-partum in primiparous cows. J. Vet. Med. Series A 47:463-468. doi: 10.1046/j.1439-0442.2000.00305.x.

Egger, M., G.D. Smith, and D.G. Altman eds. 2008. Systematic reviews in health care: Meta-analysis in context. 2nd ed. BMJ Publishing Group, London, UK.

Francoz, D., V. Wellemans, J.P. Dupré, J.P. Roy, F. Labelle, P. Lacasse, and S. Dufour. 2017. Invited review: A systematic review and qualitative analysis of treatments other than conventional antimicrobials for clinical mastitis in dairy cows. J. Dairy Sci. 100:7751–7770. doi:10.3168/jds.2016-12512.

Galphin, S.P., and E. Patrick. 1996. Evaluation of insecticide impregnated tail tags for preventing intramammary infection in dairy heifers. *In* Proc. ADSA - Dairy foods research division. 78

Gentilini, M. B., L. R. Molina, and A. U. Carvalho. 2012. Use of an Escherichia coli J5 vaccine on immunization of dairy heifers against mastitis caused by *E. coli*. Arq. Bras. Med. Vet. Zootec. 64:67-74.

Giraudo, J. A., A. Calzolari, H. Rampone, A. Rampone, A. T. Giraudo, C. Bogni, A. Larriestra, and R. Nagel. 1997. Field trials of a vaccine against bovine mastitis. 1. Evaluation in heifers. J. Dairy Sci. 80:845-853. doi: 10.3168/jds.S0022-0302(97)76006-5.

Han, H.R., and H.M. Park. 1999. A field trial with a bovine staphylococcal mastitis vaccine in lactating cows and heifers. *In* Proc. 38th NMC Ann. Mtg. 197–198.

Higgins, J.P.T., D.G. Altman, and J.A.C. Sterne eds. 2011. Assessing risk of bias in included studies. *In* Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

Hovareshti, P., M. Bolourchi, and A. H. Tabatabayi. 2007. Comparison of the effect of systemic and local antibacterial therapy to control staphylococcal intramammary infection in prepartum heifers. J. Vet. Res. 62:7-9

Lopez-Benavides, M. G., R. T. Cursons, S. J. Lacy-Hulbert, and J. H. Williamson. 2009. Heifer teats sprayed in the dry period with an iodine teat sanitizer have reduced *Streptococcus uberis* teat-end contamination and less *Streptococcus uberis* intra-mammary infections at calving. Vet. Microbiol. 134:186-191. doi: 10.1016/j.vetmic.2008.09.027.

Lopez-Benavides, M. G., J. H. Williamson, S. J. Lacy-Hulbert, and R. T. Cursons. 2006. Teat spraying prior to calving may reduce the risk of heifer mastitis caused by *Streptococcus uberis*. Proc. New Zealand Soc. Anim. Prod. 66:168-171.

Matthews, K. R., R. J. Harmon, B. E. Langlois, W. L. Crist, and R. W. Hemken. 1988. Use of latex teat dip with germicide during the prepartum period. J. Dairy Sci. 71:1940-1946. doi: 10.3168/jds.S0022-0302(88)79764-7.

McCarthy, K.K., R. Kreft, and P.M. Sears. 2005. Effect of prepartum intramammary treatment in heifers on postpartum intramammary infections, somatic cell count and milk production. *In* Proc. 44th NMC Ann. Mtg. 257–258.

McDougal, S., K. I. Parker, A. M. Weir, and C. W. R. Compton. 2008. Effect of application of an external teat sealant and/or oral treatment with a monensin capsule pre-calving on the prevalence and incidence of subclinical and clinical mastitis in dairy heifers. New Zealand Vet. J. 56:120-129. doi: 10.1080/00480169.2008.36820.

McDougall, S., K. Parker, C. Compton, and C. Heuer. 2005. Reducing subclinical and clinical mastitis in dairy heifers by precalving infusion of a teat sealant and/or parenteral antibiotic therapy. in Mastitis in dairy production: current knowledge and future solutions. H. Hogeveen, ed. Wageningen Academic Publishers, Wageningen, The Netherlands.

McDougall, S., K.I. Parker, C. Heuer, and C.W.R. Compton. 2009. A review of prevention and control of heifer mastitis via non-antibiotic strategies. Vet. Microbiol. 134:177–185. doi:10.1016/j.vetmic.2008.09.026.

Middleton, J. R., C. D. Luby, B. J. Steevens, J. Lakritz, L. L. Timms, and G. R. Bader. 2005. Effect of prepartum intramammary treatment with pirlimycin hydrochloride on prevalence of early first-lactation mastitis in dairy heifers. J. Am. Vet. Med. Assoc. 227:1969-1974. doi: 10.2460/javma.2005.227.1969.

Middleton, J.R., C.D. Luby, and D.S. Adams. 2009. Efficacy of vaccination against staphylococcal mastitis: A review and new data. Vet. Microbiol. 134:192–198. doi:10.1016/j.vetmic.2008.09.053.

Neave, F.K., F.H. Dodd, R.G. Kingwill, and D.R. Westgarth. 1969. Control of mastitis in the dairy herd by hygiene and management. J. Dairy Sci. 52:696–707. doi:10.3168/jds.S0022-0302(69)86632-4.

Newton, H., and G. Chambers. 24AD. Incidence of peri-calving clinical mastitis in maiden heifers treated with an internal teat sealant at varying pre-calving intervals. *In* Proc. 28th World Buiatrics Congr., pp. 67–68.

Nickerson, S.C., and R.L. Boddie. 1990. Effects of novel intramammary devices on incdience of mastitis. *In* Proc. 29th NMC Ann. Mtg. 186.

Nickerson, S. C., W. E. Owens, G. M. Tomita, and P. W. Widel. 1999. Vaccinating dairy heifers with a *Staphylococcus aureus* bacterin reduces mastitis at calving. Large Anim. Pract. 20:16-28.

Nickerson, S. C., E. P. Hovingh, and P. W. Widel. 2008. Immunisation of dairy heifers with a *Staphylococcus aureus* bacterin reduces infection level and somatic cell counts at time of calving. Mastitis Control: From Science to Practice:119-123. Wageningen Academic Publishers, Wageningen, The Netherlands.

Nickerson, S.C. 2009. Control of heifer mastitis: Antimicrobial treatment—An overview. Vet. Microbiol. 134:128–135. doi:10.1016/j.vetmic.2008.09.019.

Nickerson, S.C., V.J. Eubanks, J.D. Chapman, L.O. Ely, K.P. Zanzalari, F.M. Kautz, D.J. Hurley, N.E. Forsberg, and Y.Q. Wang. 2013. Mastitis and Dairy Replacement Heifers: Management Practices to Reduce the Odds. In: Proc. 28th Annu. Southwest Nutr. Manag. Conf., pp. 19–34.

Norman, C.B., W.E. Owens, R.L. Boddie, and C.H. Ray. 2005. Efficacy of prepartum, intramammary lactating cow antibiotics in dairy heifers. *In* Proc. 44th NMC Ann. Mtg., 237–238.

O'Connor, A.M., J.M. Sargeant, I.A. Gardner, J.S. Dickson, M.E. Torrence, C.E. Dewey, I.R. Dohoo, R.B. Evans, J.T. Gray, M. Greiner, G. Keefe, S.L. Lefebvre, P.S. Morley, A. Ramirez, W. Sischo, D.R. Smith, K. Snedeker, J. Sofos, M.P. Ward, and R. Wills. 2010. The REFLECT statement: Methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. J. Vet. Intern. Med. 24:57-64. doi: 10.1111/j.1939-1676.2009.0441.x.

Olde Riekerink, R.G., H.W. Barkema, D.T. Scholl, D.E. Poole, and D.F. Kelton. 2010. Management practices associated with the bulk-milk prevalence of *Staphylococcus aureus* in Canadian dairy farms. Prev. Vet. Med. 97:20-28. doi: 10.1016/j.prevetmed.2010.07.002.

Oliver, S.P., M.J. Lewis, B.E. Gillespie, and H.H. Dowlen. 1997. Influence of prepartum antibiotic administration on intramammary infections in multiparous dairy cows during early lactation. *In* Proc. ADSA - Dairy foods research division. 77.

Oliver, S.P., B.E. Gillespie, M.J. Lewis, E.C. Jaenicke, and H.H. Dowlen. 2001. An effective strategy for the control of mastitis in heifers. *In* Proc. 2nd ISMMQ. 507–508.

Oliver, S. P., H. H. Dowlen, H. Moorehead, J. W. Hallberg, S. T. Chester, K. C. Lamar, S. J. Ivey, B. E. Gillespie, D. L. Johnson, and M. J. Lewis. 2004. Influence of prepartum pirlimycin hydrochloride or penicillin-novobiocin therapy on mastitis in heifers during early lactation. J. Dairy Sci. 87:1727-1731. doi: 10.3168/jds.S0022-0302(04)73326-3.

Oliver, S. P., S. I. Headrick, B. E. Gillespie, M. J. Lewis, D. L. Johnson, K. C. Lamar, H. Moorehead, H. H. Dowlen, and J. W. Hallberg. 2007. Intramammary infections in heifers during early lactation following intramammary infusion of pirlimycin hydrochloride or penicillin-novobiocin at the first milking after parturition. J. Dairy Res. 74:211-217. doi: 10.1017/S0022029906002391.

Oliver, S. P., E. C. Jaenicke, R. K. Roberts, H. H. Dowlen, M. J. Lewis, and B. E. Gillespie. 2003. Prepartum antibiotic treatment of heifers: milk production, milk quality and economic benefit. J. Dairy Sci. 86:1187-1193. doi: 10.3168/jds.S0022-0302(03)73702-3.

Oliver, S. P., S. H. King, M. J. Lewis, and H. H. Dowlen. 1991. Influence of prepartum antibiotic therapy on intramammary infections (IMI) in primigravid heifers during early lactation. J. Dairy Sci. 75:406-414. doi: 10.3168/jds.S0022-0302(92)77776-5.

Oliver, S. P., M. J. Lewis, B. E. Gillespie, and H. H. Dowlen. 1997. Antibiotic residues and prevalence of mastitis pathogen isolation in heifers during early lactation following prepartum antibiotic therapy. J. Vet. Med. Ser. B 44:213-220. doi: 10.1111/j.1439-0450.1997.tb00967.x.

Owens, W. E., S. C. Nickerson, R. L. Boddie, G. M. Tomita, and C. H. Ray. 2001. Prevalence of mastitis in dairy heifers and effectiveness of antibiotic therapy. J. Dairy Sci. 84:814-817. doi: 10.3168/jds.S0022-0302(01)74538-9.

Parker, K., C. Compton, F. Anniss, and S. McDougall. 2007. Management of dairy heifers and its relationships with the incidence of clinical mastitis. N. Z. Vet. J. 55:208–216. doi: 10.1080/00480169.2007.36770.

Parker, K. I., C. Compton, F. M. Anniss, A. Weir, C. Heuer, and S. McOougall. 2007. Subclinical and clinical mastitis in heifers following the use of a teat sealant precalving. J. Dairy Sci. 90:207-218. doi: 10.3168/jds.S0022-0302(07)72622-X.

Parker, K. I., C. Heuer, S. McDougall, C. W. R. Compton, and F. M. Anniss. 2008. Quarter-level analysis of subclinical and clinical mastitis in primiparous heifers following the use of a teat sealant or an injectable antibiotic, or both, precalving. J. Dairy Sci.91 :169-181. doi: 10.3168/jds.2007-0212.

Passchyn, P., S. Piepers, and S. De Vliegher. 2013. Systemic prepartum treatment of end-term dairy heifers with penethamate hydriodide: Effect on udder health, milk yield, and culling until 120 days in milk. J. Dairy Sci. 96:6324-6335. doi: 10.3168/jds.2013-6626.

Passchyn, P., S. Piepers, and S. De Vliegher. 2014. Pathogen group-specific risk factors for intramammary infection in treated and untreated dairy heifers participating in a prepartum antimicrobial treatment trial. J. Dairy Sci. 97:6260–6270. doi:10.3168/jds.2014-8119.

Pellegrino, M., J. Giraudo, C. Raspanti, R. Nagel, L. Odierno, V. Primo, and C. Bogni. 2008. Experimental trial in heifers vaccinated with *Staphylococcus aureus* avirulent mutant against bovine mastitis. Vet. Microbiol. 127(1-2):186-190. doi: 10.1016/j.vetmic.2007.07.028.

Pellegrino, M., J. Giraudo, C. Raspanti, L. Odierno, and C. Bogni. 2010. Efficacy of immunization against bovine mastitis using *a Staphylococcus aureus* avirulent mutant vaccine. Vaccine 28:4523-4528. doi: 10.1016/j.vaccine.2010.04.056.

Piepers, S., S. DeVliegher, A. DeKruif, G. Opsomer, and H.W. Barkema. 2009. Impact of intramammary infections in dairy heifers on future udder health, milk production, and culling. Vet. Microbiol. 134:113–120. doi:10.1016/j.vetmic.2008.09.017.

Pol, M., M. Ebinger, and A. Castelletti. 2015. Efficacy of pre-calving treatment with parenteral danofloxacin in heifers. *In* Proc. 54th NMC Ann. Mtg. 207–208.

Pyörälä, S. 2008. Mastitis in post-partum dairy cows. Reprod. Domest. Anim. 43:252–259. doi:10.1111/j.1439-0531.2008.01170.x.

Robertson, B. G., J. H. Williamson, B. Kuhn-Sherlock, S. J. Lacy-Hulbert, and S. A. Turner. 2017. Use of internal teat sealant in heifers reduces mastitis and may affect milk production. Animal Production Science 57:1494-1498. doi: 10.1071/AN16436.

Roger, N., and L.M. Jiménez. 2013. Case study: udder health in heifers after vaccination against mastitis. *In* Proc. 52nd NMC Ann. Mtg. 197–198.

Roy, J. P., D. Du Tremblay, L. DesCôteaux, S. Messier, D. Scholl, and É. Bouchard. 2007. Effect of precalving intramammary treatment with pirlimycin in nulliparous Holstein heifers. Can. J. Vet. Res. 71:283-291.

Ruiz, R., L.O. Tedeschi, and A. Sepúlveda. 2017. Investigation of the effect of pegbovigrastim on some periparturient immune disorders and performance in Mexican dairy herds. J. Dairy Sci. 100:3305–3317. doi: 10.3168/jds.2016-12003.

Sampimon, O. C., S. de Vliegher, H. W. Barkema, J. Sol, and T. J. G. M. Lam. 2009. Effect of prepartum dry cow antibiotic treatment in dairy heifers on udder health and milk production. J. Dairy Sci. 92:4395-4403. doi: 10.3168/jds.2009-2203.

Santos, J. E. P., R. L. A. Cerri, J. H. Kirk, S. O. Juchem, and A. Villasenor. 2004. Effect of prepartum milking of primigravid cows on mammary gland health and lactation performance. Livest. Prod. Sci. 86(1-3):105-116. doi: 10.1016/S0301-6226(03)00149-0.

Seegers, H., C. Fourichon, and F. Beaudeau. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Vet. Res. 34:475–491. doi:10.1051/vetres:2003027.

Shamseer, L., D. Moher, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, L.A. Stewart, and PRISMA-P Group. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 350:g7647. doi:10.1136/BMJ.G7647.

Supré, K., F. Haesebrouck, R.N. Zadoks, M. Vaneechoutte, S. Piepers, and S. De Vliegher. 2011. Some coagulase-negative *Staphylococcus* species affect udder health more than others. J. Dairy Sci. 94:2329–2340. doi:10.3168/jds.2010-3741.

Tang, K.L., N.P. Caffrey, D.B. Nóbrega, S.C. Cork, P.E. Ronksley, H.W. Barkema, A.J. Polachek, H. Ganshorn, N. Sharma, J.D. Kellner, and W.A. Ghali. 2017. Restricting the use of antibiotics in food-producing animals and its associations with antibiotic resistance in food-producing animals and human beings: a systematic review and meta-analysis. Lancet Planet. Health 1:316–327. doi:10.1016/S2542-5196(17)30141-9.

Tenhagen, B. A., D. Edinger, B. Baumgartner, P. Kalbe, G. Klunder, and W. Heuwieser. 2001. Efficacy of a herd-specific vaccine against *Staplylococcus aureus* to prevent post-partum mastitis in dairy heifers. J. Vet. Med. Series A 48:601-607. doi: 10.1046/j.1439-0442.2001.00397.x.

Timms, L.L., A. Steffens, and L. Allen. 1997. Field trial evaluation of a persistent barrier teat dip for preventing mastitis during the dry period. *In* Proc. ADSA - Dairy foods research division. 173.

Toušová, R., L. Stádník, J. Ducháček, and M. Baráková. 2011. Influence of vaccination against mastitis on its incidence, somatic cells amount and content of solid components in milk of primiparous dairy cows. Výzkum v Chovu Skotu 53:84-90.

Trinidad, P., and S.C. Nickerson. 1989. Prevalence and treatment of intramammary infections in breeding age and in pregnant dairy heifers. *In* Proc. 28th NMC Ann. Mtg. 220.

Trinidad, P., S. C. Nickerson, T. K. Alley, and R. W. Adkinson. 1990. Efficacy of intramammary treatment in unbred and primigravid dairy heifers. J. Am. Vet. Med. Assoc. 197:465-470.

Vargas, R. T., F. N. Souza, M. A. V. P. Brito, J. R. F. Brito, M. O. Leite, L. M. Fonseca, I. B. M. Sampaio, and M. M. O. P. Cerqueira. 2016. Partial budget analysis of prepartum antimicrobial therapy and *Escherichia coli* J5 vaccination of dairy heifers and their effect on milk production and milk quality parameters. Pesqui. Vet. Bras. 36:77-82. doi: 10.1590/S0100-736X2016000200003

Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor Package. J. Stat. Softw. 36:1–48. doi:10.18637/jss.v036.i03.

Table 3-1. Study characteristics of the 62 included studies.

First author (year)	Country	Breed of Cattle	Intervention type[6]						Time of intervention application (days before calving)[5]	Outcome type[1]				Outcome measurement time (DIM)	Pathogens Studied[2]		
			AM	TS	TD	Other	VAC	COM		CM	IMI	SCM	CR		Cont.	CNS	Env.
Acuna (2003)[3, 4]	Argentina	Holando Argentino	•						45	•				3			
Acuna (2015)[3]	Argentina	Holando Argentino						•	60	•				30			
Acuna Pasqualina[3] (2006)	Argentina	Holstein		•					12		•			3-6	•		•
Alaniz (2013)[3, 4]	US	Mixed					•		75	•				21			•
Ataee (2009)	Iran	Holstein	•						10 – 14				•	3		•	
Barto (1978)	US	Mixed	•						10 - 14		•			3			
Bassel (2003)[3, 4]	Canada	Mixed			•						•			1			•
Bastan (2010)	Turkey	Holstein	•						45		•	•	•	10	•	•	•
Borm (2006)	US/Canada	-	•						10 – 21		•		•	4		•	
Bowers (2006)	US	Mixed				•			Starting at 21		•			Up to 60			
Bray (1989)[3]	US	-	•						21		•		•	1			
Bryan (2005)	New Zealand	-	•						0	•				4			
Bryan (2009)	New Zealand	-	•						0	•				4			
Bush (1977)	US	Mixed			•				14		•			4			
Compton (2008)[3]	New Zealand					•			0	•				4 - 21			
Compton (2014)	New Zealand	Mixed	•						21	•	•			3 for IMI; 30 for CM			
Contreras (2013)	US	-	•						12 – 18		•			4		•	•

[1]Outcomes reported on: CM = clinical mastitis; IMI = intramammary infection; SCM = subclinical mastitis; CR = cure rate of existing infections
[2]Presented if the study differentiated infectious pathogens. Cont. = contagious; Env. = environmental; CNS = coagulase-negative staphylococci
[3]Inclusion and analysis based on abstract or conference proceedings
[4] Not included in meta-analysis
[5] Intended application time based on expected parturition date, actual time relative to calving was variable
[6]Type of intervention studied: AM = antimicrobial; TS = teat sealant; TD = teat dips/sprays; VAC = vaccine; COM = combination of multiple types

Table 3-1. Continued

First author (year)	Country	Breed of Cattle	Intervention type[6]						Time of intervention application (days before calving)[5]	Outcome type[1]				Outcome measurement time (DIM)	Pathogens Studied[2]		
			AM	TS	TD	Other	VAC	COM		CM	IMI	SCM	CR		Cont.	CNS	Env.
Edinger (2000)	Germany	Holstein			•				Starting at 15		•			4	•	•	•
Galphin (1996)[3]	US	Holstein				•			300	•				0			
Gentilini (2012)	Brazil	Mixed					•		60	•	•			7 for IMI; 100 for CM	•		•
Gilbert (2014)[3]	Wales	-		•					60	•				30			
Giraudo (1997)	Argentina	Holstein					•		56	•	•	•		15	•		
Han (1999)[3]	Korea	-					•		-	•	•	•		-			
Hovareshti (2007)	Iran	Holstein	•						7 – 14				•	7 – 14	•	•	
Jimenez (2008)[3]	Spain		•						60		•		•	7			
Lopez-Benavides (2009)	New Zealand	-			•				Starting at 21		•			14		•	•
Lopez-Benavides (2006)	New Zealand	-			•				Starting at 21		•			14			•
Matthews (1988)	US	Mixed		•					14		•			8	•	•	•
McCarthy (2005)[3 4]	US		•						21		•			5			
McDougall (2008)	New Zealand		•	•				•	21		•			10	•	•	•
McDougall (2005)[4]	New Zealand	Mixed	•	•				•	40	•	•			4 for IMI; 14 for CM			
Middleton (2005)	US	Holstein	•						10 – 14		•		•	7 – 28	•	•	•
Newton (2014)[3]	New Zealand	-		•					28	•				-			
Lopez-Benavides (2006)	New Zealand	-			•				Starting at 21		•			14			•

[1]Outcomes reported on: CM = clinical mastitis; IMI = intramammary infection; SCM = subclinical mastitis; CR = cure rate of existing infections
[2]Presented if the study differentiated infectious pathogens. Cont. = contagious; Env. = environmental; CNS = coagulase-negative staphylococci
[3]Inclusion and analysis based on abstract or conference proceedings
[4] Not included in meta-analysis
[5] Intended application time based on expected parturition date, actual time relative to calving was variable
[6]Type of intervention studied: AM = antimicrobial; TS = teat sealant; TD = teat dips/sprays; VAC = vaccine; COM = combination of multiple types

Table 3-1. Continued.

First author (year)	Country	Breed of Cattle	Intervention type[6]						Time of intervention application (days before calving)[5]	Outcome type[1]				Outcome measurement time (DIM)	Pathogens Studied[2]		
			AM	TS	TD	Other	VAC	COM		CM	IMI	SCM	CR		Cont.	CNS	Env.
Edinger (2000)	Germany	Holstein			•				Starting at 15		•			4	•	•	•
Galphin (1996)[3]	US	Holstein				•			300	•				0			
Gentilini (2012)	Brazil	Mixed					•		60	•	•			7 for IMI; 100 for CM	•		•
Gilbert (2014)[3]	Wales	-		•					60	•				30			
Giraudo (1997)	Argentina	Holstein					•		56	•	•	•		15	•		
Han (1999)[3]	Korea	-					•		-	•	•	•		-			
Hovareshti (2007)	Iran	Holstein	•						7 – 14				•	7 – 14	•	•	
Jimenez (2008)[3]	Spain		•						60		•		•	7			
Lopez-Benavides (2009)	New Zealand	-			•				Starting at 21		•			14		•	•
Lopez-Benavides (2006)	New Zealand	-			•				Starting at 21		•			14			•
Matthews (1988)	US	Mixed		•					14		•			8	•	•	•
McCarthy (2005)[3 4]	US		•						21		•			5			
McDougall (2008)	New Zealand		•	•				•	21		•			10	•	•	•
McDougall (2005)[4]	New Zealand	Mixed	•	•				•	40	•	•			4 for IMI; 14 for CM			
Middleton (2005)	US	Holstein	•						10 – 14		•		•	7 – 28	•	•	•
Newton (2014)[3]	New Zealand	-		•					28	•				-			
Nickerson (1990)[3 4]	US	-					•		-		•			-	•		•

[1]Outcomes reported on: CM = clinical mastitis; IMI = intramammary infection; SCM = subclinical mastitis; CR = cure rate of existing infections
[2]Presented if the study differentiated infectious pathogens. Cont. = contagious; Env. = environmental; CNS = coagulase-negative staphylococci
[3]Inclusion and analysis based on abstract or conference proceedings
[4] Not included in meta-analysis
[5] Intended application time based on expected parturition date, actual time relative to calving was variable
[6]Type of intervention studied: AM = antimicrobial; TS = teat sealant; TD = teat dips/sprays; VAC = vaccine; COM = combination of multiple types

Table 3-1. Continued.

First author (year)	Country	Breed of Cattle	Intervention type[6]						Time of intervention application (days before calving)[5]	Outcome type[1]				Outcome measurement time (DIM)	Pathogens Studied[2]		
			AM	TS	TD	Other	VAC	COM		CM	IMI	SCM	CR		Cont.	CNS	Env.
Nickerson (1994)[3 4]	US	-					•		-		•			-	•	•	•
Nickerson (1999)	US	Jersey					•		540		•			0	•	•	•
Nickerson (2008)	US	Holstein					•		365		•			0	•	•	•
Norman (2003)[3 4]	US	Mixed	•						14		•			14	•	•	
Oliver (1991)	US	Jersey	•						7		•			7	•	•	•
Oliver (1997)a	US	Jersey	•						14		•			3	•	•	•
Oliver (1997)b[3 4]	US	Mixed	•						0		•			0			
Oliver (2003)	US	Jersey	•						7		•			3	•	•	•
Oliver (2004)	US	Mixed	•						14		•		•	7, 14, 30	•	•	•
Oliver (2007)	US	Mixed	•						0		•			30	•	•	•
Owens (2001)	US	Jersey	•						1st, 2nd and 3rd trimesters		•			0	•		
Parker (2007)	New Zealand	-		•				•	30		•			4	•	•	•
Parker (2008)	New Zealand	-	•	•				•	30		•			5	•	•	•
Passchyn (2013)	Belgium	-	•						14		•			3	•	•	•
Pellegrino (2008)	Argentina	Holando Argentino					•		Pre-pregnancy		•			51	•		
Pellegrino (2010)	Argentina	Holando Argentino					•		30		•			21	•		
Pol (2015)[3]	Argentina	-	•						15	•				50			

[1]Outcomes reported on: CM = clinical mastitis; IMI = intramammary infection; SCM = subclinical mastitis; CR = cure rate of existing infections
[2]Presented if the study differentiated infectious pathogens. Cont. = contagious; Env. = environmental; CNS = coagulase-negative staphylococci
[3]Inclusion and analysis based on abstract or conference proceedings
[4] Not included in meta-analysis
[5] Intended application time based on expected parturition date, actual time relative to calving was variable
[6]Type of intervention studied: AM = antimicrobial; TS = teat sealant; TD = teat dips/sprays; VAC = vaccine; COM = combination of multiple types

Table 3-1. Continued.

First author (year)	Country	Breed of Cattle	Intervention type[6]						Time of intervention application (days before calving)[5]	Outcome type[1]				Outcome measurement time (DIM)	Pathogens Studied[2]		
			AM	TS	TD	Other	VAC	COM		CM	IMI	SCM	CR		Cont.	CNS	Env.
Robertson (2017)[4]	New Zealand	-		•					30	•	•	•		1 for IMI; 30 for CM; 56 for SCM	•	•	•
Roger (2013)[3 4]	Spain	Holstein					•		45		•			-			
Roy (2007)	Canada	Holstein	•						9		•		•	5	•	•	•
Sampimon (2009)	Netherlands	Holstein	•						63	•	•			10	•	•	•
Santos (2004)	US	Holstein				•			15	•	•			28 for IMI; 135 for CM	•	•	
Tenhagen (2001)	Germany	Holstein					•		35	•	•			21 – 28 for IMI; 90 for CM	•	•	•
Timms (1997)[3 4]	US				•				12		•			3		•	
Tousova (2011)	Czech Republic	Holstein					•		45	•				-			
Trinidad (1989)[3 4]	US		•						180-90				•	0	•	•	
Trinidad (1990)[3]	-	Jersey	•						-				•	0	•		
Vargas (2016)	Brazil	Holstein	•				•	•	60	•				190			

[1]Outcomes reported on: CM = clinical mastitis; IMI = intramammary infection; SCM = subclinical mastitis; CR = cure rate of existing infections
[2]Presented if the study differentiated infectious pathogens. Cont. = contagious; Env. = environmental; CNS = coagulase-negative staphylococci
[3]Inclusion and analysis based on abstract or conference proceedings
[4] Not included in meta-analysis
[5] Intended application time based on expected parturition date, actual time relative to calving was variable
[6]Type of intervention studied: AM = antimicrobial; TS = teat sealant; TD = teat dips/sprays; VAC = vaccine; COM = combination of multiple types

Table 3-2. Risk of bias as assessed by SYRCLE guidelines (Hooijmans et al. 2013) from full-text manuscripts included in systematic review (n=40).

	Selection Bias				Detection and Performance Bias			Attrition Bias	Reporting Bias
First author (year)	Allocation concealment	Allocation implementation	Sequence generation	Baseline characteristics	Blinding	Random housing	Random outcome assessment	Incomplete outcome data	Selective outcome reporting
Ataee (2009)		•[1]		•					•
Barto (1978)		•		•				•	•
Bastan (2010)		•		•				•	•
Borm (2006)		•	•	•				•	•
Bowers (2006)		•		•				•	•
Bryan (2005)									•
Bryan (2009)		•		•				•	•
Bush (1977)		•						•	•
Compton (2014)		•	•	•				•	•
Contreras (2013)		•						•	•
Edinger (2000)		•		•				•	•
Gentilini (2012)		•							•
Giraudo (1997)		•		•				•	•
Hovareshti (2007)		•		•					•
Lopez-Benavides (2009)		•		•				•	•
Lopez-Benavides (2006)		•		•					•
Matthews (1988)		•		•				•	•
McDougall (2008)		•		•		•		•	•
McDougall (2005)		•		•				•	•
Middleton (2005)	•	•	•	•				•	•
Nickerson (1999)									•
Nickerson (2008)				•				•	•
Oliver (1991)	•	•		•				•	•
Oliver (1997)a		•		•					•
Oliver (2003)		•		•				•	•
Oliver (2004)		•		•				•	•

[1] • indicates that the feature was thoroughly described and minimized risk of bias in that area.

Table 3-2. Continued.

	Selection Bias				Detection and Performance Bias			Attrition Bias	Reporting Bias
First author (year)	Allocation concealment	Allocation implementation	Sequence generation	Baseline characteristics	Blinding	Random housing	Random outcome assessment	Incomplete outcome data	Selective outcome reporting
Oliver (2007)		•[1]		•				•	•
Owens (2001)		•		•				•	•
Parker (2007)	•	•	•	•				•	•
Parker (2008)	•	•		•	•			•	•
Passchyn (2013)	•	•		•				•	•
Pellegrino (2008)		•						•	•
Pellegrino (2010)			•	•				•	•
Robertson (2017)				•				•	•
Roy (2007)	•	•		•				•	•
Sampimon (2009)	•	•		•				•	•
Santos (2004)	•	•		•				•	•
Tenhagen (2001)			•						•
Tousova (2011)					•			•	•
Vargas (2016)		•							•

[1]• indicates that the feature was thoroughly described and minimized risk of bias in that area.

Table 3-3. Outcome definitions in full-text manuscripts included in the systematic review (n=40).

First Author (study year)	Outcomes defined[1]	Outcome Definition
Ataee (2009)		Culturing protocol only (NMC Guidelines)
Barto (1978)		Culturing protocol only (NMC Guidelines)
Bastan (2010)		Culturing protocol only (NMC Guidelines)
Borm (2006)		Culturing protocol only (NMC Guidelines)
Bowers (2006)	T[1]	MAS-D-TEC test for mastitis (SCC via conductivity)
Bryan (2005)	P[1]	Clinical mastitis – specific signs not described
Bryan (2009)	T	Clinical mastitis – described signs and symptoms
Bush (1977)		Culturing protocol only (NMC Guidelines)
Compton (2014)	T	Culturing protocol (NMC Guidelines), >200 CFU/mL for IMI
Contreras (2013)	T	Culturing protocol (NMC Guidelines), >500 CFU/mL for IMI
Edinger (2000)	P	Culturing protocol provided for IMI; changes in secretion and veterinary treatment for CM
Gentilini (2012)	P	Culturing protocol (NMC Guidelines) for IMI; SCC > 250k cells/mL for SCM; clinical signs for CM
Giraudo (1997)	T	Primary pathogen cultured and SCC > 250k cells/mL for IMI
Hovareshti (2007)		Culturing protocol defined for *S. aureus*
Lopez-Benavides (2006)	P	Culturing protocol for *S. uberis* and CNS; clinical signs for CM
Lopez-Benavides (2009)	T	Culturing protocol defined for *S. uberis,* ≥ 100 CFU/mL for IMI
Matthews (1988)	T	Duplicate quarter samples with same pathogen for IMI; clinical signs for CM
McDougall (2005)		Clinical signs for CM, but signs/symptoms not described
McDougall (2008)	T	Same pathogen isolated from sequential or parallel milk samples for IMI
Middleton (2005)	T	≤ 2 species cultured, and same species in 2 sequential samples for IMI; SCC reported; California Mastitis Test used for SCM
Nickerson (1999)		Culturing protocol only (NMC Guidelines)
Nickerson (2008)		Culturing protocol only
Oliver (1991)	T	Same pathogen isolated at two sequential samplings
Oliver (1997)a		Culturing protocol only (NMC Guidelines)
Oliver (2003)		Culturing protocol only (NMC Guidelines)
Oliver (2004)	T	Culturing protocol (NMC Guidelines), ≥ 100 CFU/mL for IMI
Oliver (2007)	T	Culturing protocol (NMC Guidelines), ≥ 100 CFU/mL for IMI
Owens (2001)		Culturing protocol only
Parker (2007)	T	≥ 300 CFU/mL of same species in sequential samples for IMI
Parker (2008)	T	≥ 300 CFU/mL of same species in parallel samples for IMI

[1]T = Definitions were thoroughly described, P = Definitions were partially described.

Table 3-3. Continued.

First Author (study year)	Outcomes defined	Outcome Definition
Passchyn (2013)	T	≥ 100 CFU/mL for IMI
Pellegrino (2008)		Culturing protocol only (NMC Guidelines)
Pellegrino (2010)	T	Presented CFUs from samples
Robertson (2017)	T	≥ 500 CFU/mL for IMI; ≥ 100 CFU/mL and clinical signs for CM; ≥ 120,000 cells/mL for SCM
Roy (2007)	T	Different IMI definitions by species – very thoroughly detailed
Sampimon (2009)	T	≥ 100 CFU/mL for major pathogen IMI, ≥ 500 CFU/mL for major pathogen IMI
Santos (2004)		Culturing protocol only (NMC Guidelines)
Tenhagen (2001)	P	Culturing protocol only for IMI; clinical changes in secretion for CM; reported SCC
Tousova (2011)	P	Clinical mastitis – specific signs not described
Vargas (2016)		≤ 2 species cultured, ≥ 100 CFU/mL for *S. aureus* and *S. agalactiae* IMI, ≥ 1000 CFU/mL for other species

[1]T = Definitions were thoroughly described, P = Definitions were partially described.

Table 3-4. Meta-regression of records included in a systematic review pre-calving therapies for improving udder health in primiparous dairy heifers on the outcome used to assess udder health.

Outcome Type	Risk Ratio[1] (95% CI)	Number of Records	Heterogeneity[3]
Clinical mastitis	0.50 (0.39 – 0.64)	15	$I^2 = 0.85$
Intramammary Infection	0.61 (0.51 – 0.74)	34	Q = 343.5
Subclinical mastitis[2]	0.30 (0.18 – 0.50)	2	$P < 0.0001$

[1]Ratio of risk of infection in treated animals to risk of infection in untreated controls.
[2]Subclinical mastitis defined by as animal having an elevated SCC (measured by electrical conductivity of milk, the California Mastitis Test or a quantified SCC > 120,000 cells/mL or 250,000 cells/mL).
[3]Heterogeneity refers to residual heterogeneity from meta-regression model attributed to differences in studies.

Table 3-5. Meta-regression of records included in a systematic review pre-calving treatment for improving udder health in primiparous dairy heifers on the type of treatment assessed in the study.

Intervention Type	Risk Ratio[1] (95% CI)	Number of Records	Heterogeneity[2]
Antimicrobial[3]	0.58 (0.46-0.74)	27	
Teat sealant	0.40[5] (0.30-0.52)	7	$I^2 = 0.85$
Teat dips/spray	1.50[5] (0.74-3.1)	3	$Q = 369.78$
Vaccine	0.55 (0.35-0.86)	12	$P < 0.0001$
Combination[4]	0.34 (0.25-0.45)	4	
Other	0.50 (0.11–1.1)	4	

[1]Ratio of risk of infection in treated animals to risk of infection in untreated controls.
[2]Residual heterogeneity from meta-regression model attributed to differences in studies.
[3]Includes antibiotics as well as antiseptic compounds like chlorhexidine.
[4]Combinations in included studies consisted of teat sealants and antimicrobials; vaccines and antimicrobials; teat sealants, vaccines and antimicrobials.
[5]Risk ratio different ($P < 0.05$) from studies assessing antimicrobial effectiveness.

Table 3-6. Meta-regression of records included in a systematic review pre-calving treatment for improving udder health in primiparous dairy heifers on the type of treatment assessed in the study, stratified by type of pathogen.

Pathogen Type[6] Intervention Type	Risk Ratio[1] (95% CI)	Number of Records	Heterogeneity[2]
Contagious			
Antimicrobial[3]	0.41 (0.34-0.77)	15	
Teat sealant	0.40 (0.22-0.72)	4	$I^2 = 0.29$
Teat dips/spray	1.48[7] (0.52-4.27)	1	Q = 39.36
Vaccine	0.56 (0.34-0.92)	8	$P = 0.025$
Combination	0.93[7] (0.33-1.26)	1	
Environmental			
Antimicrobial[3]	0.58 (0.46-0.74)	18	
Teat sealant	0.27[5] (0.15-0.49)	5	$I^2 = 0.57$
Teat dips/spray	0.56 (0.17-1.8)	3	Q = 71.77
Vaccine	0.78 (0.20-2.98)	4	$P < 0.0001$
Combination	0.25[5] (0.13-0.50)	2	
Coagulase-negative staphylococci			
Antimicrobial[3]	0.51 (0.35-0.76)	15	
Teat sealant	0.46 (0.29-0.73)	4	$I^2 = 0.89$
Teat dips/spray	1.48 (0.31–0.73)	2	Q = 212.71
Vaccine	0.89 (0.38-2.07)	4	$P < 0.0001$
Combination	0.34[5] (0.21-0.56)	2	
Other	0.80[7] (0.19-3.34)	1	

[1]Ratio of risk of infection in treated animals to risk of infection in untreated controls.
[2]Residual heterogeneity from meta-regression model attributed to differences in studies.
[3]Includes antibiotics as well as antiseptic compounds like chlorhexidine.
[4]Combinations in included studies consisted of teat sealants and antimicrobials; vaccines and antimicrobials; teat sealants, vaccines and antimicrobials.
[5]Risk ratio is different ($P < 0.05$) from studies assessing antimicrobial effectiveness.
[6]Contagious pathogens include *Staphylococcus aureus* and *Streptococcus agalactiae.* Environmental pathogens include *Escherichia coli,* and non-*agalactiae* streptococci.
[7]Intervention type within pathogen type stratum did not have enough records for a meta-analysis. Risk ratio presented is calculated from the single record within that stratum.

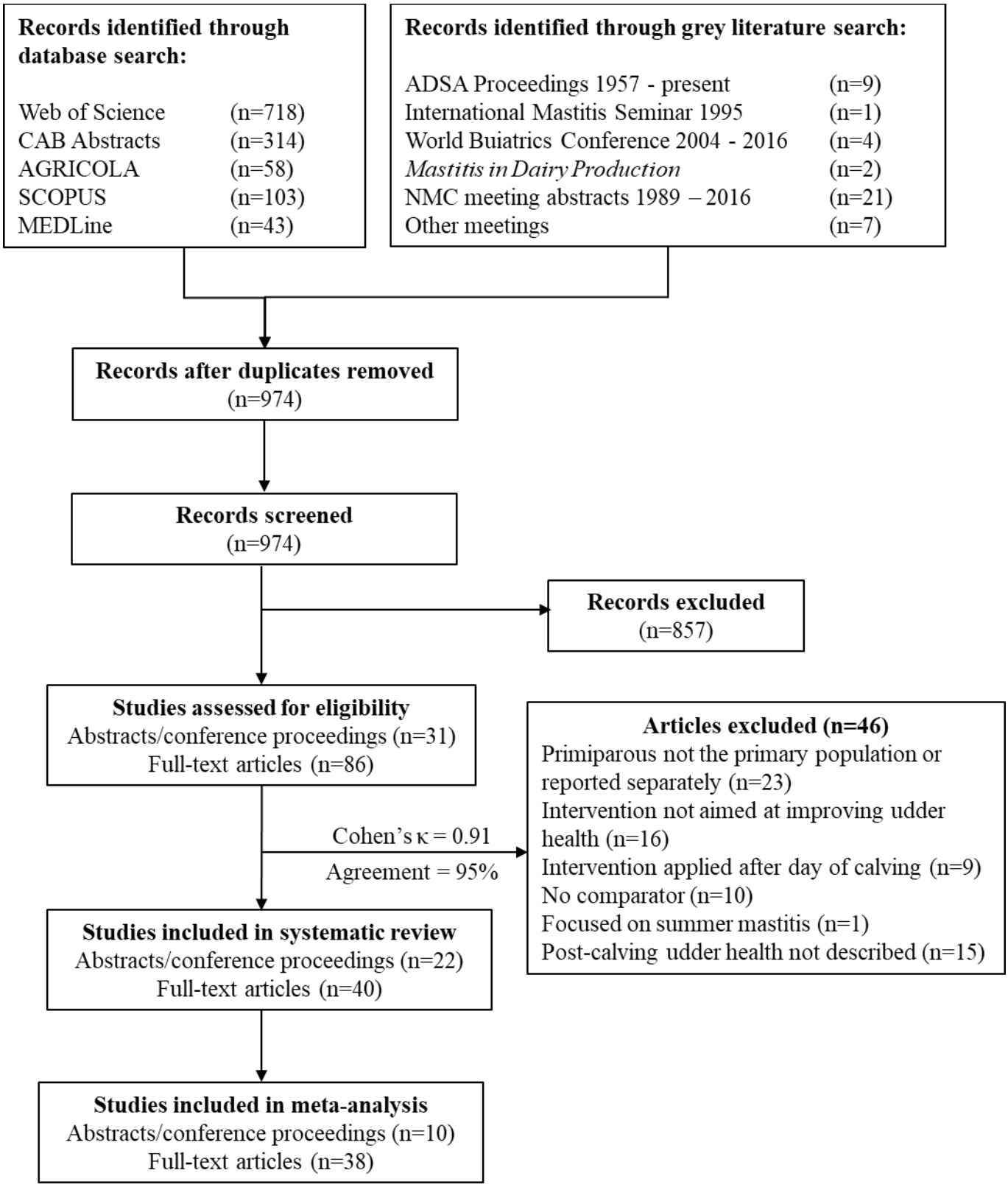


Figure 3-1. Flow diagram of the manuscript selection process.

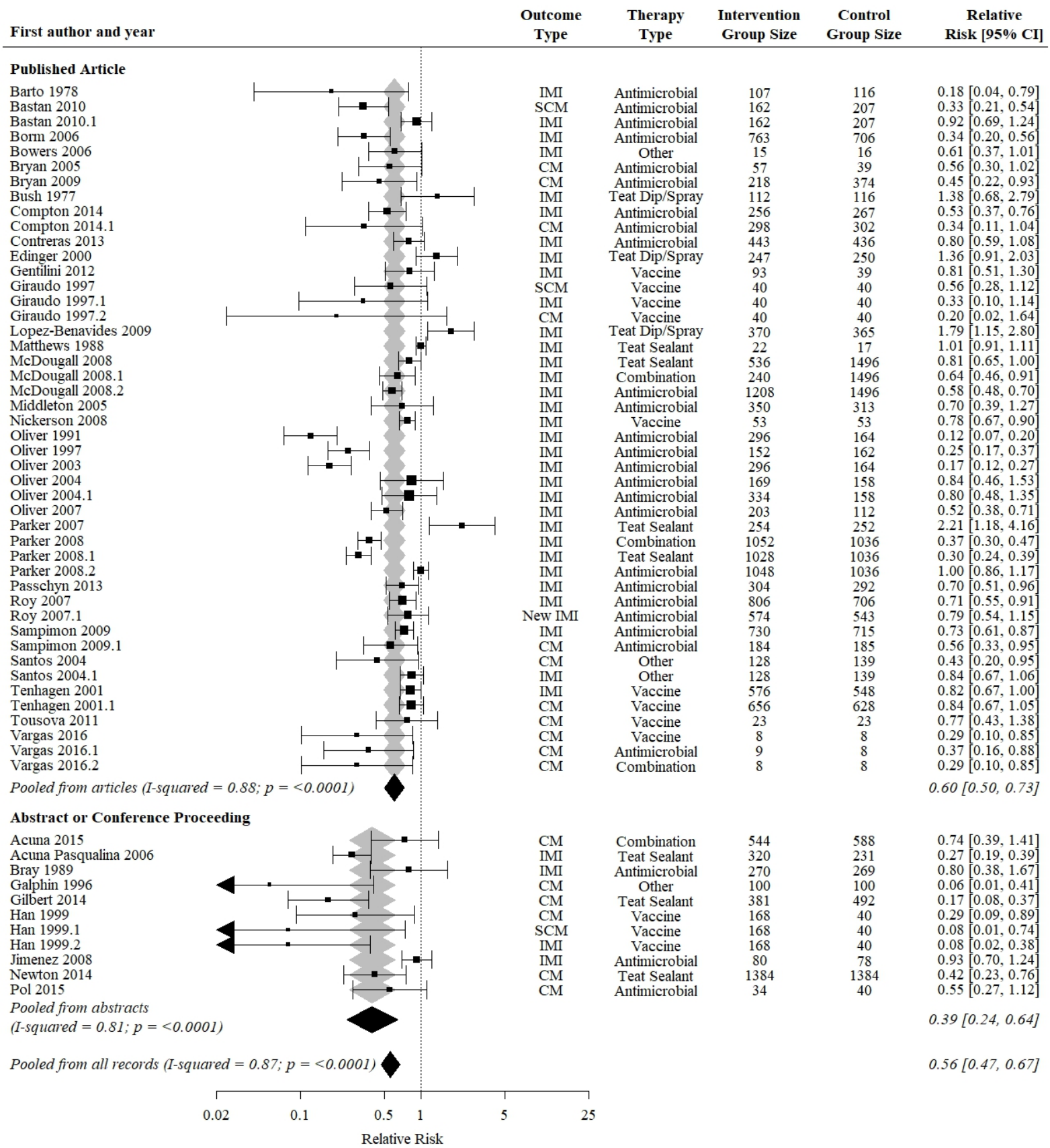


Figure 3-2. Forest plot of effect size estimates from all (n = 48) studies included in the meta-analysis assessing effectiveness of pre-calving therapy at improving post-calving udder health in dairy heifers. Size of boxes reflect relative size of effect estimate; diamonds represent pooled estimates within strata.
CM – Clinical mastitis; IMI – intramammary infection; new IMI – IMI that was not present pre-calving; SCM – subclinical mastitis (defined as elevated SCC)

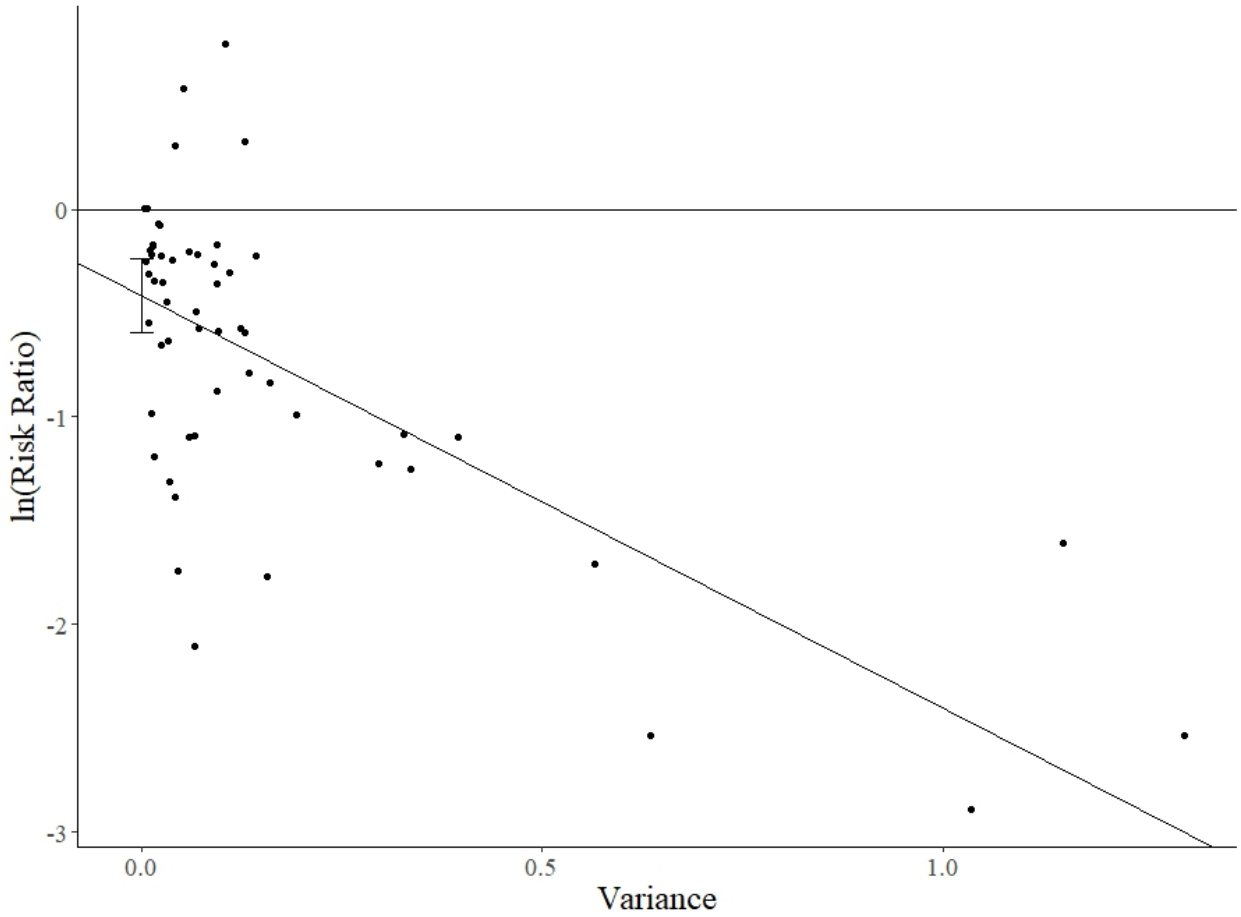


Figure 3-3. Egger's regression for small-study effects in a meta-analysis of studies assessing effectiveness of pre-calving therapies at improving udder health in dairy heifers plotted as the natural log of the risk ratio from a given study against the estimate of variance from the same study. Vertical bars indicate 95% confidence intervals at the intercept (variance = 0; largest possible studies).

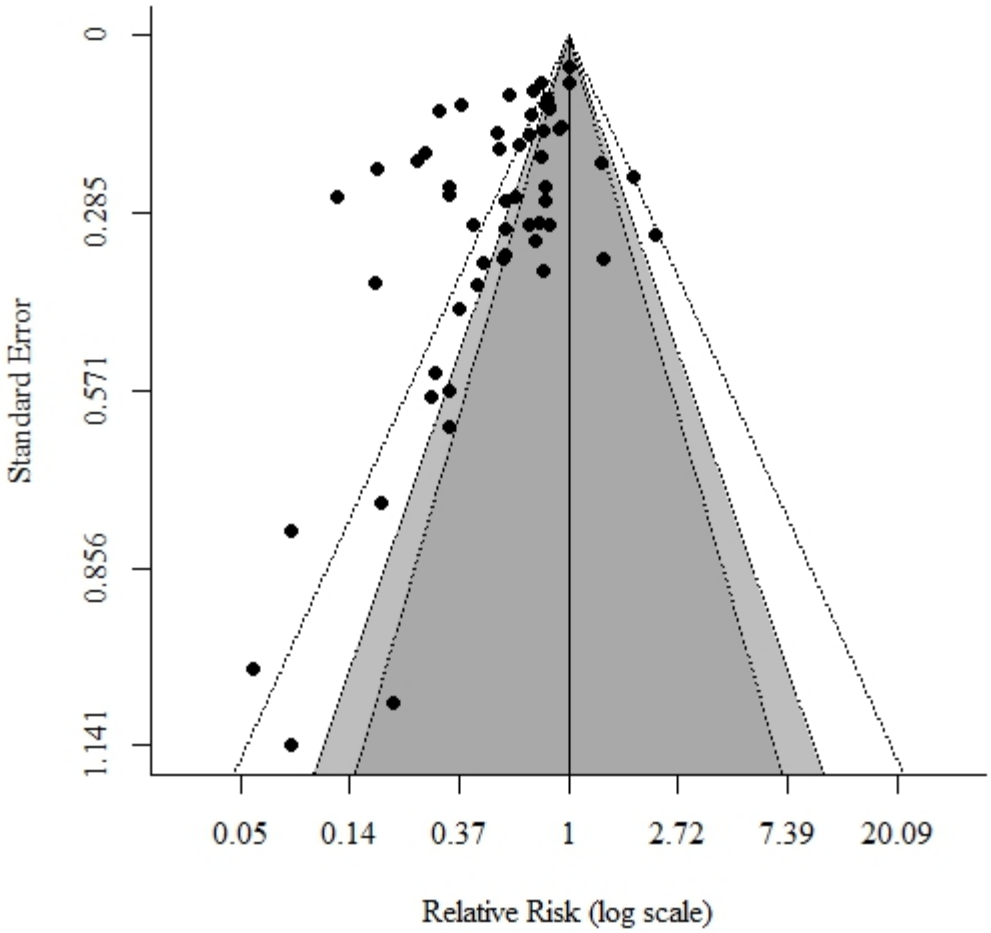


Figure 3-4. Funnel plot to assess asymmetry and publication bias in meta-analysis of studies assessing effectiveness of pre-calving therapies at improving udder health in dairy heifers. Relative risk = 1 indicates equal risk of udder disease in treated and untreated animals (no effect of treatment). Darkest triangle indicates area of 90% confidence, light grey triangle indicates area of 95% confidence and white triangle indicates area of 99% confidence.

3.8. Appendices

Appendix 3-1. Search terms used in database searches. MeSH search is integrated in Web of Science, so the same terms were employed there.

The search terms used were separated into 3 separate components corresponding to the study population, the intervention and the outcome type. Each of these was combined with the Boolean operator AND. Subject heading searches were conducted separately from the text search as the available subject headings were not comprehensive.

1) To identify studies with a study population looking at heifers and not just all cows grouped together, the search was split into two parts (to ensure that the age of animals in the study was appropriate, as well as restrict our search to dairy cows only) which were combined using the Boolean operator AND:
    a. The terms “dairy” and “milk*” combined with the Boolean operator OR. The corresponding subject headings were “dairy herds”, “dairy cows”, “dairy cattle”, and “milking” combined with the operator OR.
    b. The terms “heifer*”, “primiparous”, “primigravid”, “nulliparous” combined with the Boolean operator OR. The corresponding subject headings were “heifers” and “bred heifers” combined with the Boolean operator OR.
2) To identify studies with the relevant interventions, the search was again split into two parts (timing of intervention had to be pre-calving, as well as specify all types of intervention), which were again combined using the Boolean operator AND:

a. The terms “pre-calving”, “pre-partum” and “calving” combined with the Boolean operator OR. The corresponding subject heading search terms were “calving” and “prepartum period” combined using the operator OR.
b. The terms “therap*”, “treat*”, “antimicrob*”, “antibiotic” and “seal*” combined using the operator OR. The corresponding subject headings were “drug therapy", "therapy”, "treatment", and "combination therapy" combined using the Boolean operator OR.

3) To identify all possibly relevant outcomes the search terms used were “mastit*”, “infect*”, “clinical mastit*”, “bacteri*”, “inflam*”, “pathogen”, and “microorganism” combined using the Boolean operator OR. The corresponding subject headings used were "mastitis" , "summer mastitis" , "subclinical mastitis" , "mycotic mastitis" , "bovine mastitis" , and "inflammation" combined using the Boolean operator OR.

Appendix 3-2. Details on product used, primary therapeutic ingredient(s) and method of application for each of the 62 studies.

First Author (year)	Product(s) used
Acuna (2003)	Intramammary cloxacillin benzathine
Acuna (2015)	Enemast dry cow (Merial Argentina) – intramammary cloxacillin benzathine and ampicillin trihydrate
Acuna Pasqualina (2006)	Dry-off (WestfaliaSurge Argentina) – external teat sealant
Alaniz (2013)	Enviracor J-5 – *E. coli* vaccine
Ataee (2009)	Tyloject (Razak Co, Tehran, Iran) - intramuscular tylosin; Cefquinome (Cobactan; Intervet, Boxmeer, The Netherlands) - intramuscular cefquinome
Barto (1978)	Benzathine cephapirin (Bristol Laboratories, Syracuse, NY) – intramammary
Bassel (2003)	Observational study – comparison of farms using and not using pre-calving teat dipping
Bastan (2010)	Rilexine 500 (Virbac, France) – intramammary cephalexin monohydrate and neomycin
Borm (2006)	Cefa-Lak (Fort Dodge Animal Health, Fort Dodge, IA) - intramammary cephapirin sodium
Bowers (2006)	Prepartum milking twice daily beginning 3 wk pre-partum
Bray (1989)	Intramammary penicillin and dihydrostreptomycin
Bryan (2005)	Masticillin (Stockguard Laboratories Ltd., Hamilton, New Zealand) – intramuscular procaine penicillin
Bryan (2009)	Mamyzin (Boehringer Ingelheim Vetmedica GmbH) - parenteral penethemate hydroiodide
Bush (1977)	Iodine teat dip
Compton (2008)	Intercept (Bayer, New Zealand Ltd., Manukau City, NZ) – bismuth subnitrate and chlorhexidine internal teat sealant
Compton (2014)	Reducing interval from calving to first milking; SelPor (Ancare, NZ) - pre-calving, pour-on selenium feed supplementation
Contreras (2013)	Tylosin (Elanco, Greenfield, IN) - intramuscular
Edinger (2000)	Gelstadip (Pfizer GmbH, Karlsruhe, Germany) – iodine teat dip
Galphin (1996)	Insecticide impregnated tail tag
Gentilini (2012)	Rotatec – J5 *E. coli* vaccine
Gilbert (2014)	Orbeseal (Zoetis Animal Health) – internal teat sealant
Giraudo (1997)	Proprietary *S. aureus* vaccine
Han (1999)	Proprietary *S. aureus* vaccine
Hovareshti (2007)	Nafpenzal DC (Intervet International, Holland) – intramammary sodium naficillin, procaine benzyl penicillin and dihydrostreptomycin; Tylan (Elanco, Animal Health, Elililly) - intramammary tylosin; Baytril (Bayer, Bayer Leverkusen, Germany) – subcutaneous enrofloxacin
Jimenez (2008)	Cepravin VS – intramammary cephalonium
Lopez-Benavides (2009)	Teatguard Plus (Ecolab, Hamilton, NZ) – iodine teat spray
Lopez-Benavides (2006)	Teatguard Plus (Ecolab, Hamilton, NZ) – iodine teat spray
Matthews (1988)	Latex external teat sealant with germicide
McCarthy (2005)	Albadry Plus (Pfizer Inc.) – intramammary novobiocin, penicillin G; Orbeseal (Pfizer Inc.) – internal teat sealant; T-Hexx (Hydromer) – teat dip

Appendix 3-2. Continued.

First Author (year)	Product(s) used
McDougall (2005)	Teatseal (Pfizer Animal Health, Auckland, NZ) – bismuth subnitrate internal teat sealant; Tylan 200 (Elanco Animal Health, Manukau City, NZ) – intramuscular tylosin
McDougall (2008)	DryFlex (Delaval, Hamilton, NZ) – external teat sealant; Rumensin ABC (Elanco Animal Health Manuaku City, NZ) – parenteral monensin
Middleton (2005)	Pirsue (Pfizer, Inc. Kalamazoo MI) – intramammary pirlimycin hydrochloride
Newton (2014)	Teat Seal (Zoetis) – internal teat sealant
Nickerson (1990)	5 intramammary device models
Nickerson (1994)	Fly control; dry cow antibiotic
Nickerson (1999)	Lysigin (Boehringer Ingelheim Vetmedica Inc., St. Joseph, MO) - *S. aureus* vaccine
Nickerson (2008)	Lysigin (Boehringer Ingelheim Vetmedica Inc., St. Joseph, MO) - *S. aureus* vaccine
Norman (2003)	Today – intramammary cephapirin sodium; Cefalak – intramammary cephapirin sodium
Oliver (1991)	Intramammary sodium cloxacillin; intramammary cephapirin sodium
Oliver (1997)a	Today (Bristol Meyers, Evansville, IN) – intramammary cephapirin sodium
Oliver (1997)b	Pirsue (Upjohn Co., Kalamazoo, MI) – intramammary pirlimycin hydrochloride; Albacillin (Upjohn Ltd., West Sussex, England) – intramammary penicillin procaine and novobiocin sodium
Oliver (2003)	Intramammary sodium cloxacillin; intramammary cephapirin sodium
Oliver (2004)	Albacillin (Pfizer Animal Health, Kalamazoo, MI) – intramammary novobiocin sodium and penicillin procaine; Pirsue (Pfizer, Inc. Kalamazoo, MI) – intramammary pirlimycin hydrochloride
Oliver (2007)	Albacillin (Pfizer Animal Health, Kalamazoo MI) – intramammary novobiocin sodium and penicillin procaine; Pirsue (Pfizer, Inc. Kalamazoo, MI) – intramammary pirlimycin hydrochloride
Owens (2001)	Tomorrow (Fort Dodge Animal Health, Fort Dodge IA) – intramammary cephapirin benzathine; Quartermaster (Pharmacia Upjohn, Kalamazoo MI) – intramammary penicillin and streptomycin; Albadry Plus (Pharmacia Upjohn, Kalamazoo, MI) – intramammary penicillin and novobiocin; Micotil (Eli Lilly and Co. Indianapolis, IN) – intramammary tilmicosin Cefasafe (Intervet, Boxmeer, The Netherlands) – intramammary cephalonium
Parker (2007)	Teat Seal (Pfizer Animal Health, NZ Ltd., Auckland, NZ) – internal bismuth subnitrate teat sealant
Parker (2008)	Teat Seal (Pfizer Animal Health, NZ Ltd., Auckland, NZ) – internal bismuth subnitrate teat sealant
Passchyn (2013)	Tylan 200 (Elanco Animal Health, Manukau City NZ) – intramammary tylosin Mamyzin Stop M (Boehringer Ingelheim GmbH, Ingelheim, Germany) – intramuscular penethamate hydroiodide
Pellegrino (2008)	Proprietary *S. aureus* vaccine
Pellegrino (2010)	Proprietary *S. aureus* vaccine
Pol (2015)	Advocin 180 (Zoetis) – parenteral danofloxacin
Robertson (2017)	TeatSeal (Zoetis, Auckland, NZ) – internal teat sealant
Roger (2013)	STARTVAC – *E. coli* and *S. aureus* vaccine
Roy (2007)	Pirsue (Pfizer Animal Health, Kirland QC) – intramammary pirlimycin hydrochloride

Appendix 3-2. Continued.

First Author (year)	Product(s) used
Sampimon (2009)	Orbenin Extra Dry Cow (Pfizer, Capelle aan de IJssel, The Netherlands) – intramammary cloxacillin
Santos (2004)	Prepartum milking
Tenhagen (2001)	Proprietary *S. aureus* vaccine
Timms (1997)	Persistent barrier teat dip
Tousova (2011)	StartVac – *E. coli* and *S. aureus* vaccine
Trinidad (1989)	Intramammary penicillin and streptomycin
Trinidad (1990)	Intramammary penicillin and streptomycin
Vargas (2016)	Albadry Plus (Pfizer Animal Health) – novobiocin sodium and procaine penicillin; Enviracor J5 (Pfizer Animal Health) – *E. coli* vaccine

CHAPTER 4: Mastitis detection with recurrent neural networks in farms using automated milking systems

4.1. Abstract

Mastitis is the most important disease in the dairy industry. With widespread adoption of automated milking systems (AMS) in Canada, there is an increasing need for automated detection of mastitis in AMS farms. The main objective of this study was to develop a recurrent neural network (RNN) model for the detection of clinical mastitis (CM) in dairy cows on farms using AMS. Producer-recorded treatment records and AMS data were collected over 3 time periods from a total of 89 dairy farms in 7 provinces across Canada. In addition to developing effective models for the detection of CM, our study also evaluated different windows around the day of diagnosis when the cow would be considered CM-positive to guide practical implementation of models. We also compared numerous subsets of variables including milk and behavioural characteristics, cow traits and farm-level/environmental variables to determine their importance and impact on model performance. Data were randomly divided into a training and a hold-out test set, consisting of all records from 66 and 23 farms, respectively. A 10-fold internal cross-validation was also employed on the training set for model development. When comparing different windows of time around diagnosis, considering animals as CMpositive for 3 d prior to recorded diagnosis resulted in the most timely and effective detection of CM with a per-case sensitivity of 89.8% (range: 83.3-96.0%), and per-day specificity of 84.3% (range: 83.4-85.8%) over the validation folds. These levels of sensitivity and specificity were achieved when using all recorded variables and their daily variances, although the inclusion of behavioural variables and

farm-level/environmental variables provided marginal performance improvement over using milk traits alone. Performance of the model was worse on the hold-out test set with a per-case sensitivity of 83.5% (range: 77.9-86.3%) and a per-day specificity of 80.4 % (range: 78.1-82.4%), likely due to farm-specific heterogeneity not encountered in the training data. Over 90% of cases of severe CM (defined by an increase in milk temperature over the pre-CM baseline) were identified by the model, indicating effective performance for the detection of CM requiring the most immediate treatment. Somatic cell count, daily variance in milking interval and milk temperature were identified as the 3 most important variables defined by their impact on model predictions. In addition to milking traits, 8 of the top 20 variables were behavioural measurements, suggesting they can play a role in the detection of CM. Daily variances also represented 8 of the 20 most important variables indicating that CM onset may be associated with sudden, within-day changes in the animal. In conclusion, this study demonstrates that RNNs are able to effectively detect CM by integrating a number of variables that are regularly measured on AMS farms but have typically been excluded from CM detection models.

4.2. Introduction

Economically, mastitis is considered the most important disease in the dairy industry, not only due to its negative effect not only on animal health and welfare, but also due to reductions in milk production and quality (Halasa et al. 2007). In Canada, clinical mastitis (CM) annually costs dairy farms on average $19,899 per 100 cows, and although culling, mortality and milk yield reduction comprise the bulk of that cost, an average of $1,065 per 100 cows is also spent annually on diagnosis, drugs for treatment, and associated labour (Aghamohammadi et al. 2018). Earlier detection and treatment of CM improves health outcomes and reduces the long-term impacts of mastitis (Milner et al. 1997), indicating that early detection is likely a key aspect of an effective mastitis management program.

Many commercial dairy farms in Canada have transitioned to automated (robotic) milking systems (AMS), with 14% of farms using AMS nationally, and 27% of farms using AMS in Western Canada in 2019 (Barkema et al. 2015; Tse et al. 2017; Canadian Dairy Information Centre - Dairy barns by type 2019). Since AMS producers are no longer attaching milking units to cows manually, there are fewer opportunities for them to assess the udder health of individual animals. As a result, there is a growing need for more effective, automated detection of CM to compensate for fewer direct interactions of the producers with the animals. Using a number of sensors during milking, many AMS regularly collect data such as milking time, milk yield, electrical conductivity, somatic cell count (SCC), and milk components, presenting an opportunity for the integration of these measurements in disease detection models. As an increasing number of farms across Canada adopt AMS, the vast quantities of data collected on-line by these systems may provide new information that can be used specifically for mastitis

detection. Thus, this study focused on developing effective mastitis detection methods in dairy herds using AMS.

The onset of CM is associated with many different changes in the physical characteristics of the milk such as quantity (Jamrozik and Schaeffer 2012; King et al. 2017), pH (Kandeel et al. 2019) and electrical conductivity (Norberg 2005), as well as composition of milk as measured by the concentration of white blood cells (SCC) (Schukken et al. 2003), the presence of various enzymes (Pyörälä 2003), and the ratio of fat to protein (Jamrozik and Schaeffer 2012). Numerous researchers have used a variety of combinations of milk characteristics to predict the occurrence of CM with good results. Somatic cell count, milk yield, lactase dehydrogenase and electrical conductivity can be used to detect CM with a sensitivity and specificity of 75 and 80%, respectively, when used to train a neural network or a generalized additive model using data from a herd using AMS (Ankinakatte et al. 2013). Electrical conductivity, milk production and flow rates, and days in milk (DIM) can also be used to detect subclinical mastitis (elevated SCC) in an AMS with a sensitivity and specificity of 80 and 75%, respectively, when used in a neural network (Cavero et al. 2008). Ebrahimi et al. (2019) provided a comparison of numerous machine learning algorithms (deep learning, naïve Bayes, generalized linear model, decision trees, random forests and gradient boosted trees) for detecting subclinical mastitis in a conventional milking system, but reported that all models performed with fairly low specificity (1.2 - 39.7%). A 2010 review assessed 15 studies published between 1992 and 2010 on mastitis detection on a combination of conventional milking and AMS farms and demonstrated that all of them used some combination of milk yield, temperature, colour, and conductivity as inputs for their predictive models (Hogeveen et al. 2010).

Clinical mastitis is also associated with a number of changes aside from milk characteristics which are seen externally as sickness behaviours. Prior to treatment for CM, feed intake is reduced and animals are more restless during milking and more often kick off the milking units (Fogsgaard et al. 2015). In the 5 d prior to CM cows are also more likely to be displaced by others during feeding, and have reduced feeding activity during peak feeding times for the herd, indicating lowered competitiveness for food (Sepulveda-Varas et al. 2016). These behavioural changes also depend on the severity of the CM; prior to diagnosis with acute CM (pronounced signs of local and systemic inflammation), cows spend less time ruminating and lying on the side where the udder is inflamed, increase their activity and intake of silage compared to concentrate, and reduce their water intake (Siivonen et al. 2011). However, these behavioural changes are not necessarily specific to CM, as they can also be observed prior to diagnosis with subclinical ketosis, lameness, displaced abomasum and metritis (King et al. 2017; King et al. 2018). Thus, using behavioural characteristics alone may not be enough for the detection of CM, but integrating them with milk characteristics could facilitate effective detection. By combining milk conductivity and milk yield with times between milking (milking interval) and duration of milking in a neural network model, Hajaya et al. (2019) achieved a sensitivity and specificity of 97 and 99% respectively, for the detection of CM within a 10-d window on a farm using AMS. Many AMS and associated technologies already collect behavioural data such as feed intake at the AMS, milking intervals, activity and rumination; incorporation of these characteristics into predictive models can be done without considerable investment in additional hardware.

The overall objective of this study was, therefore, to develop a model using a diverse range of variables that most effectively identifies a cow milked by an AMS with CM as measured by sensitivity and specificity. This study aimed to extend on existing research in 3 ways: 1) include

all variables measured and recorded by AMS (including behavioural) as model inputs and assess how they affect model performance in terms of sensitivity and specificity of predictions; 2) make predictions using deep recurrent neural networks (RNNs) which are able to capture more complex and time-varying relationships between variables, than simple, feedforward neural networks which treat data as cross-sectional; and 3) use a much larger dataset than previous studies (89 commercial herds from 7 provinces in Canada) to effectively assess generalizability and robustness to between-herd differences.

4.3. Materials and Methods

4.3.1. Herds, animals, and data

Data used for model development in this study were collected over 3 different time periods as part of three other larger studies.

The first study collected data from 9 commercial dairy herds in southwestern Ontario using Lely (Maassluis, South Holland, the Netherlands) AMS between May 2016 and January 2017, and enrolled 605 cows which were followed for the first 50 days of lactation (King et al. 2018). The second study collected data from 6 commercial dairy herds in southwestern Ontario using Lely AMS between January 2018 and January 2019, and enrolled 390 cows which were followed for the first 60 days of lactation (Moore et al. 2020). The third study was the largest and data used from it included 78 commercial dairy herds from across Canada (2 from Alberta; 4 from British Columbia; 2 from Manitoba; 2 from New Brunswick; 41 from Ontario; 21 from Quebec and 6 from Saskatchewan). Overall data collection occurred between October 2018 and July 2020, although not every farm was followed over the entire time period. The third study included

additional herds over those listed, but these were excluded from the present analysis because they did not use Lely AMS, or did not provide health records for the animals (Matson et al. 2020). The dataset from the third study included 9,115 cow lactations (not complete lactations), and cows were followed for different lengths of time at different stages of lactation. Two herds from Study 1 were also included in Study 2; two herds from Study 1 were included in Study 3; and one herd from Study 2 was included in Study 3.

The combined dataset included 10,109 lactations from 8,152 cows and 627 cases of CM.

4.3.1.1 Input variables

Input variables for the predictive models were collected from 3 sources at different time intervals: 1) AMS data for each cow at every successful and attempted visit to the milking unit; 2) neck-mounted activity and rumination monitors (Lely Qwes-HR collars, Lely Industries N.V., Maassluis, The Netherlands) from each cow every 2 hours for each day; 3) mean daily temperature from Environment Canada based on the herds' postal codes on a daily basis; and 4) Dairy Herd Improvement (Lactanet, Guelph, ON, Canada) records for parity and DIM once per lactation. Details of all input variables and their source are provided in Table 4-1. Health and treatment records were obtained from each of the herds and were presented in the standard output format of T4C (the Lely AMS software) or the DairyComp305 software (VAS, Tulare, CA, USA).

4.3.1.2 Outcome definition

Cases of CM were determined from health and treatment records based on diagnosis by dairy producers. With data in the DairyComp305 format, "MAST" was the only disease option for mastitis, so any cases labelled as such were treated as CM cases. With data from T4C, there

were many more possible options that could correspond to mastitis cases ("Mastitis", "Chronic mastitis", "Acute catarrhal mastitis", or "Mastitis E. coli"): these were all treated as cases of CM without classifying severity. Additionally, there were options for user-inputted diseases ("Mastproto1", or "Mastproto2"), diseases that could be CM ("Bloody milk", "Abnormal swelling of udder") and instances where the disease field was left blank. Cases of any of the above were only considered to be CM if the treatment administered included at least one antibiotic or anti-inflammatory treatment indicated for the treatment of CM.

For each case of CM, only the first recorded treatment was considered as the date of diagnosis and not any follow-up treatments. To address cases treated for longer durations and having follow-up treatments, a case of CM occurring within 14 d of an earlier diagnosis in the same cow was not considered a new case of CM (Barkema et al. 1998). Any observation not corresponding to a day when CM diagnosis recorded, was considered "healthy" (outcome=0) regardless of the presence of other diseases or subclinical mastitis, and CM diagnosis days were considered as CM-positive (outcome=1). Thus, initially each cow was classified as having CM only on the day of diagnosis, and was considered healthy for the rest of the observation period. For the purposes of model training, cows were reclassified as having CM for pre-specified windows before and after the recorded day of diagnosis (3 d before, 0 d after; 3 d before, 3 d after; 7 d before, 0 d after; and 7 d before, 7 d after). This ensured that the model would be able to account for any delay between the onset of clinical symptoms and diagnosis by the producer. This window should not be too long, as identification of sick animals prior to showing clinical symptoms can increase the probability of producers treating alerts as false-positives (Mollenhorst et al. 2012). Model performance was compared between the different windows of time to determine which window led to the most accurate and timely detection of CM.

4.3.1.3 Data editing and preprocessing

The first step in preparing the data for model training was to reconcile the different time intervals at which the data were collected with to make them consistent between all variables.

Activity and rumination data were collected at bi-hourly intervals, and for the purposes of analysis needed to be aligned with data collected during visits to the milking unit which had varying time intervals. To address this, two different variables were created for both activity and rumination: 1) mean of all 2-h periods since the last successful milking; and 2) mean of all 2-h periods since the last visit to the milking unit (which could have resulted in a successful milking, failure or refusal). For each cow's visit to the milking unit, all activity and rumination records that occurred between the time of the previous milking, or previous visit, and the time of the current visit were identified. These times were rounded to the nearest 2-h interval to align with activity and rumination data, and then the mean of the identified activity and rumination records was taken as the value for the new variables.

Health records only identified the day of diagnosis with CM, and not the specific time of day, so it was not possible to attach a healthy or sick classification to each milking individually. Instead, the same classification was applied to all milking records occurring during the day. To do this, milking records needed to be aggregated so that a summary value for each day could be generated, resulting in one record per cow per day. For Boolean variables, this was done by taking the sum of all values within the same day (total number of milkings, total number of refusals, etc.). Measured values and percentages were summarized using the mean of each variable for the day. To capture within-day variation in these variables, which would have been lost if using the mean alone, within-day variances for each of these variables were also calculated. As a result, each measured and percentage variable corresponded to two variables in

the summarized data the mean of that variable during the day and its variance. Once all data were summarized to only have one record per cow per day, the health records indicating CM status and window of time before and after diagnosis, the herd-specific mean temperature obtained from Environment Canada, and the cow-specific DIM and parity were all appended. At this stage, the data were converted into 3-dimensional arrays (number of cow-lactations X days of observation X number of variables) and variables that were not going to be used as inputs for the classification model (farm and cow identifiers, dates and times) were removed.

Data were then split into a training set for model development, and a hold-out test set which was not used for model tuning, but only to evaluate the performance of the optimal model. The test set consisted of all records from 23 randomly selected farms (2,125 lactations from 1,643 cows including 131 CM cases) whereas the remaining 66 farms (7,984 lactations from 6,509 cows including 496 CM cases) were used for model development and tuning. By randomizing across all included farms in the study, the distributions of farm provinces in each subset should match that of the combined dataset.

A 10-fold internal cross-validation was used in model training within the training dataset across all farms included in that set. Training and validation arrays for each cross-validation fold were processed prior to being used as inputs for the predictive models. As cows were observed for different lengths of time, the second dimension of the array (the number of time steps) varied for each lactation. RNN models require that all input arrays have a constant length; therefore the second dimension was expanded to the maximum observed time (527 d) and was padded with zeroes to give all cow-lactations a constant length. These extra, zero-padded time steps were masked during model training so that they would not contribute to the loss function and the final parameter estimates.

The final step prior to model fitting was to standardize the data, to avoid the problem of the different scales of each of the variables (for example feed provision during milking ranged from 0 to 2.4 kg, whereas DIM ranged from 1 to 500 days) unduly impacting the learning algorithm. Values for each of the variables were scaled across the training dataset using mean normalization with the equation:

$$x' = \frac{x - mean(\vec{x}_{train})}{max(\vec{x}_{train}) - min(\vec{x}_{train})}$$

where, x refers to the original value within a given variable in either of the training, validation or test sets; $\vec{x}_{train}$ refers to the entire observed vector of values for that given variable in the training dataset; and x' refers to the transformed value that is returned. Mean normalization was chosen over Z-score normalization to ensure that all values would be between -1 and 1; and was chosen over min-max scaling so that a value of 0 would refer to the mean over the entire dataset. Missing values were replaced by zeros, so that those records could still be included in the neural network model. Using zeros (corresponding to the mean across all animals in the training set) ensured that the missing values would provide no additional information to the model. Although replacing missing values with class-specific means (different for healthy and sick animals) could improve model performance, since CM detection was unlikely to be perfect, this could have introduced further bias into the models.

4.3.2. Statistical methodology

4.3.2.1 Software packages

All data management and analyses were conducted in Python 3.7 (Van Rossum and Drake 2009) using base functions and a number of additional libraries. Initial data management and cleaning were done using Pandas v.1.1.4 (The pandas development team 2020; McKinney 2010),

and, after transforming the data to arrays, all management was done using NumPy v.1.18.5 (Harris et al. 2020). Neural network models were fitted using Keras v.1.1.2 (Chollet 2015) with Tensorflow v.2.3.1 (Abadi et al. 2015) as the backend. All visualizations were created in Tableau v.2020.2 (Mountain View, CA, USA).

4.3.2.2 Recurrent neural networks

Simple feed-forward neural networks can be thought of as an extension of generalized linear regression, where additional derived features (hidden units) are formed using linear combinations of the input features. This allows the model to capture complex inter-variable relationships (Hastie et al. 2009). These units can then be combined to create further derived features (hidden layers), and the numbers of units and layers are specified by the user. Increasing the model complexity by adding additional layers and units allows the model to represent and capture increasingly complex relationships between input and output variables. Although simple neural networks take single observations of cross-sectional data as inputs, RNNs were developed as a way to extend neural networks to handle time-series data where each observation is not independent of the others within a sequence (Hochreiter and Schmidhuber 1997).

When using RNNs, the prediction at each time step takes into account the values of the input variables at all previous time steps within the sequence. Simple RNNs struggle with learning problems involving long-term dependencies between earlier time-steps and current predictions as the gradients for earlier timesteps used for updating parameters either become very small (vanishing gradients) or very large (exploding gradients) (Goodfellow et al. 2016). Long short-term memory cells (LSTMs) can be implemented within a RNN and were developed as way to address this limitation, by retaining and passing on only pertinent information from earlier time steps (Hochreiter and Schmidhuber 1997). Within a LSTM cell, there are multiple "gates" which

control the flow of information from the inputs (input gates), the output of the current time step (output gate) and the information that is passed on to the subsequent time step (forget gate) (Hochreiter and Schmidhuber 1997). By reducing the information retained and passed forward, the LSTM is able to effectively address the issue of vanishing or exploding gradients and train a better performing model. Similar to how multiple hidden layers can be stacked in a feed-forward neural network to extract more complex inter-variable relationships, multiple recurrent layers can also be stacked to extract increasingly complex, time-dependent patterns within the sequences.

4.3.2.3 Model architecture and performance metrics

Model architecture

In the present study, a RNN with LSTMs was developed. The full model had 6 layers:

1) input layer - takes as input a sequence of observations on an individual cow during a single lactation;
2) masking layer - identifies which elements of the sequence were added as padding, and excludes them from the calculation of the loss function;
3) recurrent layer - a LSTM cell with 512 hidden units;
4) recurrent layer - a LSTM cell with 512 hidden units;
5) fully connected layer - a simple neural network hidden layer with 32 units, time distributed to produce a separate outcome at each time step and rectified linear unit activation; and

6) output layer - a simple neural network hidden layer with 2 units (corresponding to each output class - healthy and CM), time distributed to produce a separate output at each time step, with a softmax activation (ensures the two probabilities sum to 1).

Regularization to improve generalizability of the model was incorporated in the recurrent layers with the same settings for both layers. Dropout was used for both the recurrent parameters (gate weightings) and fully connected unit parameters where 10% of the parameters were randomly excluded during each update step. Overfitting was also controlled using an L2 regularization function, which increases the loss function by a factor of 0.01 multiplied by the sum of the squares of the parameters. Regularization controls overfitting by reducing the magnitude of the parameter estimates, forcing low estimates to zero.

Model fitting and optimization

The distribution of the two output classes (healthy and CM) was extremely unbalanced in the dataset. To ensure the model would not always predict that animals have a healthy udder (maximizing accuracy) each observation was weighted based on the output class at that time step. To place equal emphasis on the correct identification of healthy and CM cows, the classes were assigned weights based on the ratio of healthy cow-days to CM cow-days in the training data.

The models were fitted using the categorical cross-entropy loss function and the Adam optimizer (Kingma and Ba 2015). Other optimizers such as SGD (Bottou 1991), RMSProp (Dauphin et al. 2015), and AdaGrad (Lydia and Francis 2019) were also tested, but convergence was fastest using the Adam optimizer. The initial learning rate was set to 0.001, with the learning rate being reduced when the validation set loss would not change by more than 0.0001 for 10

epochs. Models were trained for a maximum of 500 epochs, with early stopping specified when the validation set loss did not change by more than 0.0001 for 20 epochs.

Performance metrics

Due to the imbalance in output classes, using accuracy alone would not effectively represent the performance of the model. In addition to accuracy, sensitivity and specificity were also used to assess performance. The performance metrics had the following equations:

$$Accuracy = \frac{True\ positives + True\ negatives}{All\ positives + All\ negatives}$$

$$Sensitivity = \frac{True\ positives}{All\ positives}$$

$$Specificity = \frac{True\ negatives}{All\ negatives}$$

In addition to being assessed on a cow-day interval, sensitivity was also calculated per-case and per-day around the day of diagnosis. Per-case sensitivity allowed for a better representation of the model performance as it reflected cows that were identified as sick at any point during the CM-positive window. By assessing sensitivity relative to the day of diagnosis, we were able to identify when sick cows were being identified as CM-positive and determine the model's effectiveness for early detection.

Per-case sensitivity was also calculated for a subset of cows considered to have severe CM. Cases of CM were considered to be severe when the mean observed milk temperature during the period of sickness was higher than the mean observed milk temperature of all prior days by a certain threshold, as is observed when severe mastitis is experimentally induced (Steele et al. 2019). Multiple thresholds for the increase in temperature were assessed (0.5, 0.75, 1.00, 1.25 and 1.50°C) and sensitivity was calculated for each of these thresholds.

4.3.2.4 Assessing variable importance

Variable importance was assessed at two different points in the model development process: prior to model training, and by examining the predictions of the fitted models.

The variables described in Table 4-1 were categorized as either milk traits (electrical conductivity; SCC; percent fat, lactose and protein; total yield; dead milk time; milking duration), behavioural characteristics

(activity; rumination; numbers of the different types of visits; preferences for visiting different milking units; feed intake; total visit time; time between visits; time between milking visits), cow traits (DIM; parity), environmental/farm-level characteristics (environmental temperature; number of cows milking per AMS unit) and daily variances (within-day variances calculated for the measured variables). Subsets of these variable categories were then used to train models with the same architecture and hyperparameters. These subsets were:

1) all variables without daily variances - milk traits, behavioural characteristics, cow traits and environmental/farm-level characteristics;
2) all variables and daily variances - milk traits, behavioural characteristics, cow traits, environmental/farm-level characteristics and daily variances for measured variables;
3) behavioural characteristics without daily variances;
4) behavioural characteristics with daily variances;
5) cow-level measurements without daily variances - milk traits, behavioural traits and cow traits;
6) cow-level measurements with daily variances - milk traits, behavioural traits, cow traits and daily variances for measured variables;

7) milk traits without daily variances; and

8) milk traits with daily variances for measured variables.

Performance of these models was then assessed using the previously described metrics to determine how inclusion of different categories of variables impacted the detection of CM.

Once the optimal model with respect to the length of the disease window and variable subset was selected based on the performance metrics, variable importance was again assessed relative to the predictions. For the training and validation sets, the gradient (all the partial derivatives) of the input variables with respect to the prediction outcome was determined at each time step for every cow-lactation observed. These gradients were then averaged across all observations, and their value was used to assess the relative contribution of each variable towards the predicted outcome. To determine which variables could have been leading to false-positive and false-negative predictions, the gradients were also averaged over the relevant subsets of observations.

4.4. Results

4.4.1. Determining the optimal disease window

Performance metrics for models trained using the different disease windows are presented in Table 4-2. The highest validation set accuracy was 86.4% (range: 84.7-88.1%), achieved using a CM-positive window of 3 d before and 3 d after diagnosis. The highest specificity of 86.4% (range: 84.8-88.2%) and per-day sensitivity of 71.2% (range: 63.5-76.8%) were also achieved using the window of 3 d before and 3 d after diagnosis. Although the per-case sensitivity was very similar for all 4 disease windows assessed, the highest mean percase sensitivity of 90.4%

(range: 80.7-97.7%) on the validation sets was achieved when using a window of 7 d before and 7 d after the day of diagnosis.

Proportions of sick animals identified by the model relative to the day of diagnosis are presented in Figure 4-1. The overall pattern was similar for all 4 disease windows, with the sensitivity increasing up to approximately the day of diagnosis, and then decreasing thereafter. In the training set, the highest sensitivity of 91.5% (range: 90.0-93.6%) was achieved with a window of 3 d before and 3 d after (Figure 4-1b), on the first day after the recorded diagnosis. In the validation set, the highest sensitivity of 86.7% (range: 75.9-92.1%) was achieved with a window of 7 d before and 0 d after on the recorded day of diagnosis (Figure 4-1c). By evaluating the highest achieved sensitivity before or on the day of diagnosis, we can assess the models' capacity for early detection. In the training set, the highest sensitivity before or on the day of diagnosis was achieved using a window of 3 d before and 0 d after the diagnosis (Figure 4-1a). In the validation set, the highest sensitivity before or on the day of diagnosis was achieved when using a window of 7 d before and 0 d after diagnosis (Figure 4-1c).

Adding days after the day of diagnosis to the disease window resulted in the higher sensitivities after the day of diagnosis, shifting the peak to later in the 15-d period. To keep the emphasis of the models on earlier detection of CM, a disease window of 3 d before and 0 d after the diagnosis was selected as the optimal, as it achieved similar performance as the model with a 7-d window before.

4.4.2. Identification of the most important variable subsets

Performance metrics and training epochs for models trained using the different subsets of variables are presented in Table 4-3. The highest mean validation set accuracy of 84.9% (range:

80.9%-89.4%)) was achieved when using milk traits and daily variances for measured variables. Both the highest per-day and per-case validation set sensitivities of 68.9% (range: 59.9-75.0%) and 89.8% (range: 83.3%-96.0%) respectively, were achieved when using all variables and the daily variances for the measured variables. The highest validation set specificity of 85.0% (range: 80.9%-89.4%) was achieved using milk traits and their daily variances, although specificity was very similar in all subsets of variables except for behavioural variables alone, or behavioural variables with daily variances, which were both lower. Although there was considerable variation in the number of epochs required for training models using all subsets of variables, the mean numbers of epochs were similar.

The proportions of sick animals identified by the model relative to the day of diagnosis are presented in Figure 4-2. The overall patterns were similar for the training and validation sets, and also for the different subsets of variables. Sensitivity increased to a peak on the day of diagnosis, and decreased thereafter. The highest sensitivities achieved on the day of diagnosis for the training and validation sets were 90.0% (range:

88.4-92.5%) and 85.7% (range: 78.9-94.0%), respectively, when using all variables and their daily variances. The model trained using all variables and their daily variances was selected as the optimal model, although models trained using all variables without daily variances, cow-level measurements with and without daily variances, or milk traits with and without daily variances, all performed similarly.

4.4.3. External test performance and variable importance of the optimal model

Only the optimal model identified in the previous sections was assessed for performance on the hold-out test set. The performance metrics for all three subsets of the data are presented in

Table 4-4. Accuracy on the hold-out test set was 80.3% (range: 78.0-82.4%); per-day and per-case sensitivity were 59.8% (range: 56.1-63.4%) and 83.5% (range: 77.9-86.3%), respectively; and specificity was 80.4% (range: 78.1-82.4%). As expected, test set performance was worse than both the validation and training sets for all metrics.

Sensitivity of the model for detecting cows with severe CM in the three datasets is presented in Figure 4-3. As the threshold for the increase in milk temperature over the baseline was increased, fewer cows were identified as having severe CM in all three datasets. At all thresholds, the model achieved a sensitivity of over 90%. For the training and validation sets, the highest sensitivities were achieved when a cow with a 0.75°C increase in temperature over the baseline was considered as having severe CM, but decreased at higher thresholds. In the test set, however, sensitivities increased with increasing thresholds, such that the highest sensitivity for the detection of cows with severe CM was observed with a 1.5°C over the baseline.

Variable importance was assessed over all observations across the three datasets, and the results are presented in Figure 4-4. Overall, and for both false-negatives and false-positives, the 3 variables with the gradients of largest magnitude with respect to the predictions were mean SCC, within-day variance in milk interval, and the mean milk temperature. There were some differences in the remaining variables between the three subsets of observations, although these were only in the rankings of the variables, because the top 20 variables were the same between the subsets. Of these 20 variables, 11 were milk characteristics, 8 were behavioural characteristics, 1 was a farm-level characteristic, 8 were daily variances of measured variables and 5 were quarter-specific measurements.

4.5. Discussion

A RNN was developed that identified cows with CM with high sensitivity and specificity and was able to generalize between a large number of commercial dairy herds. This is the first study to incorporate such a large number of predictive variables into the detection of CM in herds using AMS. By limiting the variables to either those measured by the AMS or otherwise routinely recorded on-farm, the implementation of this model can be easily incorporated into existing AMS software without the need for additional hardware.

When considering the performance of the model, it is important to go beyond metrics evaluated at individual time-steps to understand how the model performs in a more practical setting. This has not been commonly done in previous studies which tend to evaluate only performance per-day (Xudong et al. 2020; Sun et al. 2010), per-milking (Jensen et al. 2016; Hajaya et al. 2019) or per-sample (Mammadova and Keskin 2015). Conversely, most studies identified tend to include a period of time before and/or after diagnosis where the animal is also considered in the state of CM, although the length of time differs between studies: 3 d before and 3 d after (Ankinakatte et al. 2013); 3 d before and 7 d after (Hajaya et al. 2019); 7 d before and 10 d after (De Mol et al. 1997); and 2 d before and 2 d after CM (Cavero et al. 2008). This variation makes it difficult to compare predictive performance directly between studies, and the length of time is generally justified using biological reasoning rather than evaluating the differences in resulting performance.

Although changes in SCC and electrical conductivity can take up to 3 or 4 wk to return to pre-mastitis levels (Bonestroo et al. 2020), this window of time is too long when the objective of the model is to detect CM earlier for treatment purposes. Khatun et al. (2018) evaluated

sensitivity of a model to detect a case of CM at any point in the 3 d before diagnosis and also assessed sensitivity based on different time intervals prior to diagnosis. Using a generalized linear mixed model, sensitivity of CM detection decreased with increasing number of days considered CM-positive prior to diagnosis, whereas specificity was more consistent (Khatun et al. 2018). The highest sensitivity in Khatun et al. (2018) occurred when considering 0 d before diagnosis was 0.93 and specificity was 0.97, which were both considerably higher than our model's performance in all the windows of time assessed. In the present study, the per-day validation set sensitivity decreased slightly when increasing the number of days prior to diagnosis from 3 to 7, although the per-case validation set sensitivity did not change. The present study is the first to assess per-day sensitivity in the days around diagnosis and to incorporate the pattern over the 15 d period (7 d before and 7 d after CM diagnosis) into model selection. By evaluating when the model identified animals as having CM relative to the day of diagnosis, we identified not only the model with the highest per-day sensitivity, but also the model where the peak sensitivity occurred at or before the day of diagnosis. By excluding the days after diagnosis from when the animal was considered to be in a state of CM, we ensured that the model would be penalized for late identification of CM-positive animals. Our study also assessed sensitivity of the optimal model for the detection of cows with severe CM, although these findings should be interpreted with caution. As data about CM severity were not recorded, severity of CM was determined by changes in milk temperature compared to the pre-CM baseline, as is observed in experimentally induced cases of severe CM (Steele et al. 2019). Although the estimated sensitivity for detection of cows with severe CM was very high (>90% in all data sets), milk temperature was identified as one of the most important predictive variables so it is likely that information from this variable would have been indirectly incorporated into the model.

Although the present study had a very large number of variables, rather than assess each one's effect on model training individually, they were grouped into categories to determine how much they affected the performance of the models. Unsurprisingly, milk characteristics of the cows were the best indicators of CM, but the inclusion of additional behavioural, and farm-level/environmental variables marginally improved performance of the model. Despite behavioural changes being observed prior to diagnosis with CM, the fact that they are also observed with a range of other health disorders (King and DeVries 2018), indicates that they do not offer much additional information not captured by patterns and changes in milk characteristics. Behavioural variables alone resulted in very poor model performance, reinforcing the notion that changes in them are not specific enough to reliably and effectively detect CM onset. Despite this, behavioural variables could provide valuable information for the automatic detection of other diseases.

When developing models for the detection of CM, researchers have previously pre-specified the included variables based on existing knowledge about their association with CM (Sun et al. 2010; Mammadova and Keskin 2013; Jensen et al. 2016; Ankinakatte et al. 2013), or individual variables were first assessed for association with CM before being included in a multivariable model (Khatun et al. 2018). Conversely, the approach in the present study was to include all or subsets of variables recorded by the AMS and as part of other farm records, and determine their importance after model fitting. Feature selection prior to model fitting aims to remove irrelevant or non-informative variables, as well as reduce the number of highly correlated variables. By reducing noise in the data, models are able to converge on a solution faster, and in the case of neural networks, generalize better by preventing overfitting of training data (Kavzoglu and Mather 2002). Overfitting did not appear to be an issue in the present study, as there were only

small reductions in performance in the validation sets compared to the training sets. Increasing noise in the data also did not appear to affect times to convergence, as the number of epochs required for training were very similar for all the different subsets of variables. Leray and Gallinari (1999) demonstrated that when using neural networks on synthetic datasets, model performance did not change significantly when using different kinds of feature selection methods that removed noisy variables, suggesting that neural networks are able to effectively handle extraneous information. For more complex problems with even larger numbers of variables, the computational requirements may make including all possible variables infeasible, but in the context of this study that did not appear to be a problem.

The importance of individual variables was also assessed for the optimal model with a 3-d window before diagnosis. As neural networks can recognize complex non-linear relationships, assessing the gradient of each variable individually is inherently limited by the inability to evaluate interactions between multiple variables even though they could play a large role. The 3 variables with the largest average gradients with respect to the predicted outcome were SCC, milk temperature, and observed variance in the milk intervals observed during the day. Somatic cell count is a direct reflection of inflammation of the udder and has consistently been used as an indicator of udder health (Schukken et al. 2003), whereas milk temperature is typically used in combination with other indicators of inflammation for the diagnosis of CM (Pyörälä 2003). Although milking interval is used as a component in mastitis indices used by AMS software (Khatun et al. 2018), in our analysis the within-day variance in milking intervals ranked as more important than the mean daily milk interval lengths. Additionally, many of the top 20 variables identified were within-day variances of measured variables, even though the within-day means of those variables were not as important. This indicates that the onset of CM may be accompanied

with rather sudden changes in milking behaviour resulting in a change in within-day variability compared to when the animal is healthy. Although behavioural variables only marginally improved model performance over using milk characteristics alone, 8 out of the 20 most important variables were behavioural, in particular relating to behaviour related to milking. This indicates that behavioural variables can be useful for the detection of CM, and could possibly replace milk characteristics that were not identified as the most important in our model. Of note, electrical conductivity was not identified in the 20 variables with gradients of the greatest magnitude, even though it has long been established as a key indicator of mastitis-related inflammation (Milner et al. 1997; Pyörälä 2003). Five of the top 20 variables were also quarter-specific, of which two were for the left rear quarter. One of the quarters contributing more to predictions than others can be expected if CM cases in the dataset were more frequent in one quarter than the others. As CM diagnoses were recorded on the cow level, we were not able to determine whether that was the case, but Barkema et al. (1997) determined that high SCC and intra-mammary infections occurred more frequently in rear quarters than in the front. Ideally, CM diagnosis should have been recorded at the quarter-level rather than cow-level, so that the model could use the deeper level of detail to identify specifically which quarter was positive for CM. Since treatments are often administered at the cow-level as in the case of systemic antibiotics or anti-inflammatories, producers would not record the CM-positive quarter if they were only recording the treatments given.

Unsurprisingly, performance of the optimal model was worse on the validation sets than the training data; however, the performance on the test sets was considerably lower than on the validation sets. This would likely be because of underlying differences in the populations included in the different subsets of the data. The validation sets consisted of randomly selected

animals from herds within the training data, whereas the test set consisted of animals from entirely different herds. This difference in performance then likely reflects herd-specific variability that was not encountered during model training. Khatun et al. (2018) used two different farms for their model training and assessment, and both sensitivity and specificity were slightly different between the two farms. By including a large number of different herds in our training set, generalizability should have been improved, but we still saw a marked decrease compared to the validation sets. In the present study, only two farm-level variables were included: environmental temperature and number of cows milking per AMS unit. There are likely other farm-level variables that may improve model performance that were not collected for the study. In particular, as CM was diagnosed by the producers, differences in their diagnostic protocols could represent a considerable source of variability that would impact the outcome definition. The impact of farm-specific differences on automated detection of CM requires further study, and the development of farm-specific detection models could be the best way to address the between-farm heterogeneity.

Although CM is one of the most frequently occurring diseases in dairy cattle, at any given milking or day the odds of a cow having a new case of CM is very low. As a result, there is a considerable imbalance in the output classes for model development in the observed data. Previously, other researchers have addressed this imbalance by randomly sampling from the healthy cows to ensure a balance between the output classes during model training (Khatun et al. 2018; Ankinakatte et al. 2013). Conversely, our approach included all observed records, and weighted each observation's contribution to the loss function used in model optimization to correct for the imbalance. This approach has the benefit of also being able to estimate the disease burden in the sample population, and estimate the positive predictive value (PPV) or precision

that would be observed during implementation of the model. The PPV represents the proportion of observations classified as positive, that are true-positives, and increases with disease prevalence for a fixed sensitivity and specificity. In our external test set of 2,125 lactations from 1,643 cows including 131 CM cases, out of 318,548 cow-days observed, only 524 were classified as CM-positive, yielding a per cow-day incidence of 0.16%. The PPV achieved on the external test set by the optimal model was 0.50% (range: 0.47 - 0.52%), indicating that even with a CM-positive classification by the model, there is a high probability that the cow would not have CM. Although the PPV may not be as useful as sensitivity and specificity for comparing performance between models, it provides an indication of how the model would perform when implemented in a similar population on farms. A minimum specificity of 99% is recommended for detection of abnormal milk by AMS by the International Standard Organization (Hogeveen et al. 2010). Using the recommended specificity of 99%, with the sensitivity achieved by the optimal model and the observed daily incidence, the model would still only have a PPV of 8.9%, meaning 91.1% of positive alerts would still be false-positives. Even if the model had perfect sensitivity, and a specificity of 99%, with the observed prevalence, the PPV would still be low at 14.5%. It is difficult to put this finding into context, as no previous studies were identified that report on the PPV of their model, but it does highlight that even with considerable improvements in sensitivity and particularly specificity, PPV will likely remain low due to the very low per-day incidence of CM.

Despite good performance of the developed RNN model, it is important to interpret the findings in the context of limitations to our study. As CM diagnosis was recorded by producers, there is a chance that some cases of CM could have been missed by producers resulting in false-negatives relative to the true, unknown status of the cow. This would reduce model performance

in two ways: by considering a cow with CM as being healthy, sensitivity would decrease as other animals presenting CM onset in the same way would also be considered healthy; and by considering that cow as not having CM, even if the model identified the cow as having CM, it would be considered a false-positive and decrease specificity. Additionally, as requirements for treatment records for diseases other than mastitis are not as strict, diagnoses of other diseases were not considered in our analyses. It is possible that cows that the model considered false-positives, were actually sick with another disease which was presenting in a similar way to CM, however without a complete and exhaustive list of reliable diagnoses, there would be no way to verify this. Collapsing all observations within a given day to a single record likely resulted in a loss of information which could have informed better CM detection, as suggested by the importance of within-day variances for the models' predictions. This could only have been avoided if the timing of diagnosis was also recorded, but this could also introduce additional bias as cows are not checked for CM 24 hours a day by producers and most diagnoses would have been recorded during the day when they are awake. Jensen et al. (2016) were able to assign a CM classification to each milking as their study was conducted on farms using conventional milking twice a day. Hajaya et al. (2019) reported model performance per-milking, but still assessed CM on a daily basis, meaning that the per-milking classification was actually inferred and not directly recorded. Monitoring cows 24 hours per day to determine exact timing of CM onset is not feasible on commercial herds, and also subject to considerable inter-person biases without a very strict definition of CM, so would likely only be possible on research farms. Lastly, although the CM diagnoses were recorded, details of the cases were not consistently between all farms. Behavioural changes observed in cows differ between ones with moderate CM and those with severe/acute CM (Siivonen et al. 2011), so being able to differentiate between those animals

could have improved model performance by adding another output class for severe cases. Similarly, pathogen-specific cases of CM also differ in their risk factors, and their impact on milk production and quality (Piepers et al. 2011; Zadoks et al. 2001). Had milk samples from CM-positive cows been cultured, the model could have been trained to also identify the associated pathogen in CM cases. Culture of all CM-positive milk samples however would greatly increase the costs of the study and additional labour required by producers, likely leading to reduced producer participation. Limitations like these are to be expected in a study conducted at such a large scale, and controlling too much for between-farm differences or requiring producers to change their normal management practices would likely reduce the generalizability of the model.

4.6. Conclusions

A RNN model was successfully developed that can identify animals with CM with good sensitivity and specificity. By incorporating all variables recorded by the AMS or on-farm in other ways, the RNN was able to identify complex, time-dependent relationships between these variables resulting in improved performance over using only subsets of these variables. By using a narrow window of CM-positivity (3 d before recorded diagnosis), the model was able to provide timely detection of CM cases while detecting over 90% of animals with severe CM, likely requiring immediate attention. Reduced performance of the model on herds not included in model development and training indicates that the detection models may be further optimized for performance on individual herds, while providing a good starting point before that farm-specific tuning can occur.

4.7. References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2015. {TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems.

Aghamohammadi, M., D. Haine, D.F. Kelton, H.W. Barkema, H. Hogeveen, G.P. Keefe, and S. Dufour. 2018. Herd-level mastitis-associated costs on Canadian dairy farms. Front. Vet. Sci. 5. doi:10.3389/fvets.2018.00100.

Ankinakatte, S., E. Norberg, P. Løvendahl, D. Edwards, and S. Højsgaard. 2013. Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. Comput. Electron. Agric. 99:1–6. doi:10.1016/j.compag.2013.08.024.

Barkema, H.W., Y.H. Schukken, T.J.G.M. Lam, D.T. Galligan, M.L. Beiboer, and A. Brand. 1997. Estimation of Interdependence among quarters of the bovine udder with subclinical mastitis and implications for analysis. J. Dairy Sci. 80:1592–1599. doi:10.3168/jds.S0022-0302(97)76089-2.

Barkema, H.W., M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.-P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–45. doi:10.3168/jds.2015-9377.

Barkema, H.W., Y.H. Schukken, T.J.G.M. Lam, M.L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419. doi:10.3168/jds.S0022-0302(98)75591-2.

Bonestroo, J., M. van der Voort, N. Fall, H. Hogeveen, U. Emanuelson, and I.C. Klaas. 2020. Progression of different udder inflammation indicators and their episode length after onset of inflammation using automatic milking system sensor data. J. Dairy Sci. In press. doi:10.3168/jds.2019-18054.

Bottou, L. 1991. Stochastic Gradient Learning in Neural Networks. In Proceedings of Neuro-Nımes, Holmdel, NJ.

Canadian Dairy Information Centre - Dairy barns by type. 2019. URL: https://www.dairyinfo.gc.ca/eng/dairy-statistics-and-market-information/farm-statistics/dairy-barns-by-type/?id=1502467060775.

Cavero, D., K.H. Tölle, C. Henze, C. Buxadé, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114:280–286. doi:10.1016/j.livsci.2007.05.012.

Chollet, F. 2015. Keras. GitHub Repos.URL: https://github.com/fchollet/keras.

Dauphin, Y.N., H. de Vries, J. Chung, and Y. Bengio. 2015. RMSProp and equilibrated adaptive learning rates for non-convex optimization. arXiv ID: 1502.04390v1

De Mol, R.M., G.H. Kroeze, J.M.F.H. Achten, K. Maatje, and W. Rossing. 1997. Results of a multivariate approach to automated oestrus and mastitis detection. Livest. Prod. Sci. 48:219–227. doi:10.1016/S0301-6226(97)00028-6.

Ebrahimi, M., M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K.R. Petrovski. 2019. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Comput. Biol. Med. 114:103456. doi:10.1016/j.compbiomed.2019.103456.

Fogsgaard, K.K., T.W. Bennedsgaard, and M.S. Herskin. 2015. Behavioral changes in freestall-housed dairy cows with naturally occurring clinical mastitis. J. Dairy Sci. 98:1730–1738. doi:10.3168/jds.2014-8347.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. Sequence Modeling: Recurrent and Recursive Nets. In Deep Learning. 1st ed. MIT Press, Cambridge, MA.

Hajaya, M.K., S. Samarasinghe, D. Kulasiri, and M.L. Benavides. 2019. Detection of dairy cattle mastitis: Modelling of milking features using deep neural networks. Pages 35–41 in 23rd International Congress on Modelling and Simulation - Supporting Evidence-Based Decision Making: The Role of Modelling and Simulation, MODSIM 2019. Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ).

Halasa, T., K. Huijps, O. Østerås, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18–31. doi:10.1080/01652176.2007.9695224.

Harris, C.R., K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del R'\io, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T.E. Oliphant. 2020. Array programming with {NumPy}. Nature 585:357–362. doi:10.1038/s41586-020-2649-2.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Neural Networks. In Elements of Statistical Learning. 2nd ed. Springer Science and Business Media, New York, NY.

Hochreiter, S., and J. Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9:1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10:7991–8009. doi:10.3390/s100907991.

Jamrozik, J., and L.R. Schaeffer. 2012. Test-day somatic cell score, fat-to-protein ratio and milk yield as indicator traits for sub-clinical mastitis in dairy cattle. J. Anim. Breed. Genet. 129:11–19. doi:10.1111/j.1439-0388.2011.00929.x.

Jensen, D.B., H. Hogeveen, and A. De Vries. 2016. Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. J. Dairy Sci. 99:7344–7361. doi:10.3168/jds.2015-10060.

Kandeel, S.A., A.A. Megahed, M.H. Ebeid, and P.D. Constable. 2019. Ability of milk pH to predict subclinical mastitis and intramammary infection in quarters from lactating dairy cattle. J. Dairy Sci. 102:1417–1427. doi:10.3168/jds.2018-14993.

Kavzoglu, T., and P.M. Mather. 2002. The role of feature selection in artificial neural network applications. Int. J. Remote Sens. 23:2919–2937. doi:10.1080/01431160110107743.

Khatun, M., P.C. Thomson, K.L. Kerrisk, N.A. Lyons, C.E.F. Clark, J. Molfino, and S.C. García. 2018. Development of a new clinical mastitis detection method for automatic milking systems. J. Dairy Sci. 101:9385–9395. doi:10.3168/JDS.2017-14310.

King, M.T.M., and T.J. DeVries. 2018. Graduate Student Literature Review: Detecting health disorders using data from automatic milking systems and associated technologies. J. Dairy Sci. 101:8605–8614. doi:10.3168/jds.2018-14521.

King, M.T.M., K.M. Dancy, S.J. LeBlanc, E.A. Pajor, and T.J. DeVries. 2017. Deviations in behavior and productivity data before diagnosis of health disorders in cows milked with an automated system. J. Dairy Sci. 100:8358–8371. doi:10.3168/jds.2017-12723.

King, M.T.M., S.J. LeBlanc, E.A. Pajor, T.C. Wright, and T.J. DeVries. 2018. Behavior and productivity of cows milked in automated systems before diagnosis of health disorders in early lactation. J. Dairy Sci. 101:4343–4356. doi:10.3168/jds.2017-13686.

Kingma, D.P., and J.L. Ba. 2015. Adam: A method for stochastic optimization. Page in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR.

Leray, P., and P. Gallinari. 1999. Feature Selection With Neural Networks. Behaviormetrika 26:145–166. doi:10.2333/bhmk.26.145.

Lydia, A., and S. Francis. 2019. Adagrad-An Optimizer for Stochastic Gradient Descent. Int. J. Comput. Sci. 6: 566-568.

Mammadova, N.M., and I. Keskin. 2015. Application of neural network and adaptive neuro-fuzzy inference system to predict subclinical mastitis in dairy cattle. Indian J. Anim. Res. 49:671–679. doi:10.18805/ijar.5581.

Mammadova, N., and I. Keskin. 2013. Application of the support vector machine to predict subclinical mastitis in dairy cattle. Sci. World J. 2013:603897. doi:10.1155/2013/603897.

Matson, R.D., M.T.M. King, T.F. Duffield, D.E. Santschi, K. Orsel, E.A. Pajor, G.B. Penner, T. Mutsvangwa, and T.J. DeVries. 2021. Benchmarking of automated milking farms in Canada and associations with milk production and quality. J. Dairy Sci.104(7):7971-7983. doi:10.3168/jds.2020-20065.

McKinney, W. 2010. Data structures for statistical computing in Python. Pages 56–61 in Proceedings of the 9th Python in Science Conference.

Milner, P., K.L. Page, and J.E. Hillerton. 1997. The Effects of Early Antibiotic Treatment Following Diagnosis of Mastitis Detected by a Change in the Electrical Conductivity of Milk. J. Dairy Sci. 80:859–863. doi:10.3168/JDS.S0022-0302(97)76008-9.

Mollenhorst, H., L.J. Rijkaart, and H. Hogeveen. 2012. Mastitis alert preferences of farmers milking with automatic milking systems. J. Dairy Sci. 95:2523–2530. doi:10.3168/JDS.2011-4993.

Moore, S.M., M.T.M. King, A.J. Carpenter, and T.J. DeVries. 2020. Behavior, health, and productivity of early-lactation dairy cows supplemented with molasses in automated milking systems. J. Dairy Sci. 103:10506–10518. doi:10.3168/jds.2020-18649.

Norberg, E. 2005. Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. Livest. Prod. Sci. 96:129–139. doi:10.1016/J.LIVPRODSCI.2004.12.014.

Piepers, S., K. Peeters, G. Opsomer, H.W. Barkema, K. Frankena, and S. De Vliegher. 2011. Pathogen group specific risk factors at herd, heifer and quarter levels for intramammary infections in early lactating dairy heifers. Prev. Vet. Med. 99:91–101. doi:10.1016/J.PREVETMED.2011.02.010.

Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res. 34:565–578. doi:10.1051/vetres:2003026.

Schukken, Y.H., D.J. Wilson, F. Welcome, L. Garrison-Tikofsky, and R.N. Gonzalez. 2003. Monitoring udder health and milk quality using somatic cell counts. Vet. Res. 34:579–596. doi:10.1051/vetres:2003028.

Sepúlveda-Varas, P., K.L. Proudfoot, D.M. Weary, and M.A.G. von Keyserlingk. 2016. Changes in behaviour of dairy cows with clinical mastitis. Appl. Anim. Behav. Sci. 175:8–13. doi:10.1016/J.APPLANIM.2014.09.022.

Siivonen, J., S. Taponen, M. Hovinen, M. Pastell, B.J. Lensink, S. Pyörälä, and L. Hänninen. 2011. Impact of acute clinical mastitis on cow behaviour. Appl. Anim. Behav. Sci. 132:101–106. doi:10.1016/J.APPLANIM.2011.04.005.

Steele, N.M., T.H. Swartz, K.M. Enger, H. Schramm, R.R. Cockrum, S.J. Lacy-Hulbert, R.R. White, J. Hogan, and C.S. Petersson-Wolfe. 2019. The effect of J5 bacterins on clinical, behavioral, and antibody response following an Escherichia coli intramammary challenge in dairy cows at peak lactation. J. Dairy Sci. 102:11233–11249. doi:10.3168/jds.2019-16549.

Sun, Z., S. Samarasinghe, and J. Jago. 2010. Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. doi:10.1017/S0022029909990550.

The pandas development team. 2020. pandas-dev/pandas: Pandas. doi:10.5281/zenodo.3509134.

Tse, C., H.W. Barkema, T.J. DeVries, J. Rushen, and E.A. Pajor. 2017. Effect of transitioning to automatic milking systems on producers' perceptions of farm management and cow health in the Canadian dairy industry. J. Dairy Sci. 100:2404–2414. doi:10.3168/jds.2016-11521.

Van Rossum, G., and F.L. Drake. 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Xudong, Z., K. Xi, F. Ningning, and L. Gang. 2020. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. Comput. Electron. Agric. 178:105754. doi:10.1016/j.compag.2020.105754.

Zadoks, R.N., H.G. Allore, H.W. Barkema, O.C. Sampimon, G.J. Wellenberg, Y.T. Gröhn, and Y.H. Schukken. 2001. Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis. J. Dairy Sci. 84:2649–2663. doi:10.3168/JDS.S0022-0302(01)74719-4.

Table 4-1. All variables, their type and source included as inputs for a recurrent neural network to predict the occurrence of clinical mastitis in dairy cattle. For variables that were not measured on all farms, the number of farms where they were measured is indicated.

Variable	**Type**	**Data source**
Days in milk (DIM)	Count	DHI Records
Parity	Count	DHI Records
Visit characteristics		
Milking (cow milked at visit)	Boolean	AMS
Refusal (not enough time between milkings)	Boolean	AMS
Failure (incomplete milking)	Boolean	AMS
Feeding (feed provided during visit)	Boolean	AMS
Milk interval (time since previous milking)	Measured	AMS
Measurements during visit (if a milking visit)		
Feed intake during visit	Measured	AMS
Visit duration	Measured	AMS
Milking duration (overall and quarter-specific)	Measured	AMS
Milk temperature (overall and quarter-specific)	Measured	AMS
Electrical conductivity of milk (quarter-specific)	Measured	AMS
Dead milk time (suction applied, but no milk ejected; quarter-specific)	Measured	AMS
Milk yield (quarter-specific)	Measured	AMS
Fat % (cow-level; 52 out of 89 farms)	Percentage	AMS
Lactose % (cow-level; 35 out of 89 farms)	Percentage	AMS
Protein % (cow-level; 52 out of 89 farms)	Percentage	AMS
Somatic cell count (cow-level; 37 out of 89 farms)	Measured	AMS
Activity per 2-h period	Measured	Neck-mounted collar
Rumination time per 2-h period	Measured	Neck-mounted collar
Mean environmental temperature during day	Measured	Environment Canada

Table 4-2. Mean (range) performance metrics for the detection of CM using all variables recorded by AMS and their daily variances over a 10-fold cross-validation comparing different windows of time before and after CM diagnosis.

Dataset	No. days before diagnosis	No. days after diagnosis	Accuracy	Sensitivity		Specificity
				per-day	per-case of CM	
Training	3	0	0.844 (0.824-0.854)	0.749 (0.725-0.789)	0.936 (0.919-0.959)	0.845 (0.824-0.854)
		3	0.865 (0.858-0.873)	0.780 (0.766-0.799)	0.939 (0.924-0.956)	0.865 (0.858-0.873)
	7	0	0.765 (0.741-0.821)	0.694 (0.623-0.775)	0.931 (0.914-0.947)	0.765 (0.741-0.822)
		7	0.819 (0.776-0.840)	0.724 (0.705-0.757)	0.930 (0.918-0.943)	0.820 (0.776-0.840)
Validation	3	0	0.843 (0.834-0.858)	0.689 (0.599-0.750)	0.898 (0.833-0.960)	0.843 (0.834-0.858)
		3	0.864 (0.847-0.881)	0.712 (0.635-0.768)	0.896 (0.833-0.963)	0.864 (0.848-0.882)
	7	0	0.763 (0.732-0.809)	0.618 (0.492-0.708)	0.902 (0.792-0.963)	0.763 (0.732-0.810)
		7	0.818 (0.772-0.836)	0.678 (0.612-0.766)	0.904 (0.807-0.977)	0.819 (0.773-0.837)

Table 4-3. Mean (range) performance metrics and number of epochs required for model training over a 10-fold cross-validation using different subsets of variables for the detection of CM within 3 d prior to diagnosis.

	Training					Validation			
	No. epochs for training	Accuracy	Sensitivity		Specificity	Accuracy	Sensitivity		Specificity
Variable set			per-day	per-case of CM			per-day	per-case of CM	
All variables	26.4 (21-32)	0.838 (0.812-0.856)	0.742 (0.722-0.763)	0.920 (0.904-0.934)	0.838 (0.813-0.856)	0.838 (0.808-0.867)	0.687 (0.586-0.755)	0.883 (0.838-0.926)	0.838 (0.808-0.867)
All variables + daily variances	29.4 (22-38)	0.844 (0.824-0.854)	0.749 (0.725-0.789)	0.936 (0.919-0.959)	0.845 (0.824-0.854)	0.843 (0.834-0.858)	0.689 (0.599-0.750)	0.898 (0.833-0.960)	0.843 (0.834-0.858)
Behavioural characteristics	25.7 (21-31)	0.783 (0.691-0.848)	0.546 (0.465-0.618)	0.788 (0.744-0.840)	0.783 (0.691-0.849)	0.783 (0.678-0.840)	0.509 (0.454-0.573)	0.758 (0.680-0.911)	0.783 (0.679-0.840)
Behavioural characteristics + daily variances	28.1 (23-43)	0.765 (0.704-0.788)	0.597 (0.576-0.630	0.830 (0.814-0.851)	0.766 (0.704-0.788)	0.765 (0.706-0.792)	0.574 (0.485-0.750)	0.819 (0.720-0.956)	0.765 (0.706-0.792)
Cow-level measurements	27.1 (21-33)	0.838 (0.825-0.849)	0.731 (0.717-0.744)	0.911 (0.902-0.925)	0.838 (0.826-0.849)	0.837 (0.822-0.865)	0.667 (0.578-0.773)	0.873 (0.807-0.926)	0.837 (0.823-0.865)
Cow-level measurements + daily variances	29.2 (21-39)	0.846 (0.826-0.863)	0.740 (0.713-0.773)	0.934 (0.919-0.952)	0.846 (0.826-0.864)	0.843 (0.827-0.869)	0.669 (0.579-0.755)	0.894 (0.838-0.933)	0.844 (0.828-0.869)
Milk traits	26.0 (21-41)	0.841 (0.829-0.863)	0.674 (0.651-0.693)	0.884 (0.859-0.902)	0.841 (0.829-0.864)	0.841 (0.825-0.873)	0.625 (0.543-0.710)	0.85 (0.789-0.900)	0.842 (0.825-0.874)
Milk traits + daily variances	27.5 (21-42)	0.851 (0.820-0.881)	0.691 (0.650-0.713)	0.920 (0.883-0.939)	0.851 (0.820-0.882)	0.849 (0.809-0.894)	0.641 (0.569-0.728)	0.879 (0.807-0.926)	0.850 (0.809-0.894)

Table 4-4. Mean (range) performance metrics for the detection of CM using all variables recorded by AMS and their daily variances over a 10-fold cross-validation with a 3d sick window prior to CM diagnosis.

Dataset	Accuracy	Sensitivity		Specificity
		per-day	per-case of CM	
Training	0.844 (0.824-0.854)	0.749 (0.725-0.789)	0.936 (0.919-0.959)	0.845 (0.824-0.854)
Validation	0.843 (0.834-0.858)	0.689 (0.599-0.750)	0.898 (0.833-0.960)	0.843 (0.834-0.858)
Test	0.803 (0.780-0.824)	0.598 (0.561-0.634)	0.835 (0.779-0.863)	0.804 (0.781-0.824)

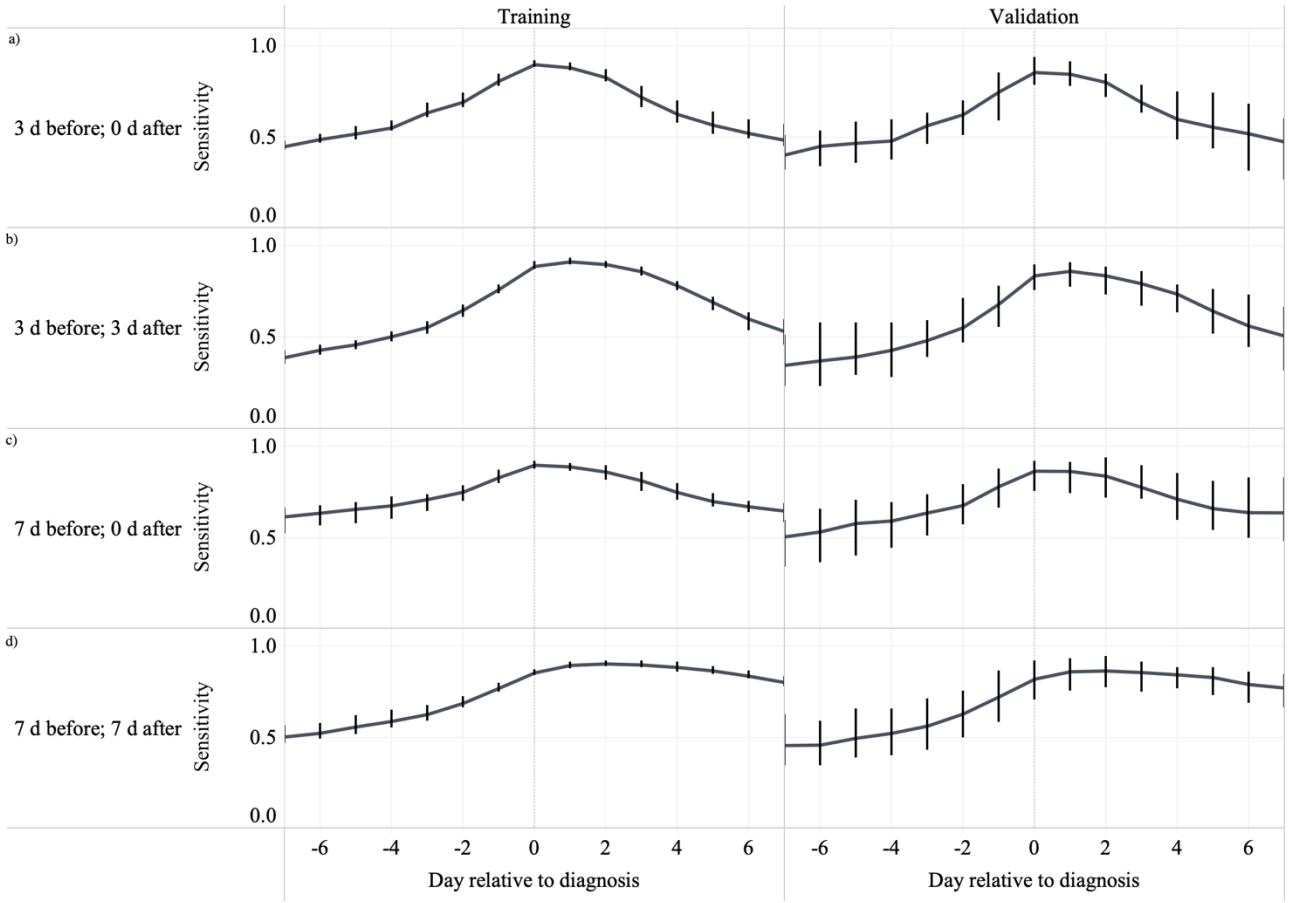


Figure 4-1. Mean ± range sensitivity (proportion of animals with CM, or animals that would be diagnosed with CM identified by the model) relative to the day of diagnosis for the detection of CM in dairy cows using a RNN. Different plots (a, b, c, d) represent the 4 windows of time where the animal was considered to be CM-positive for model training.

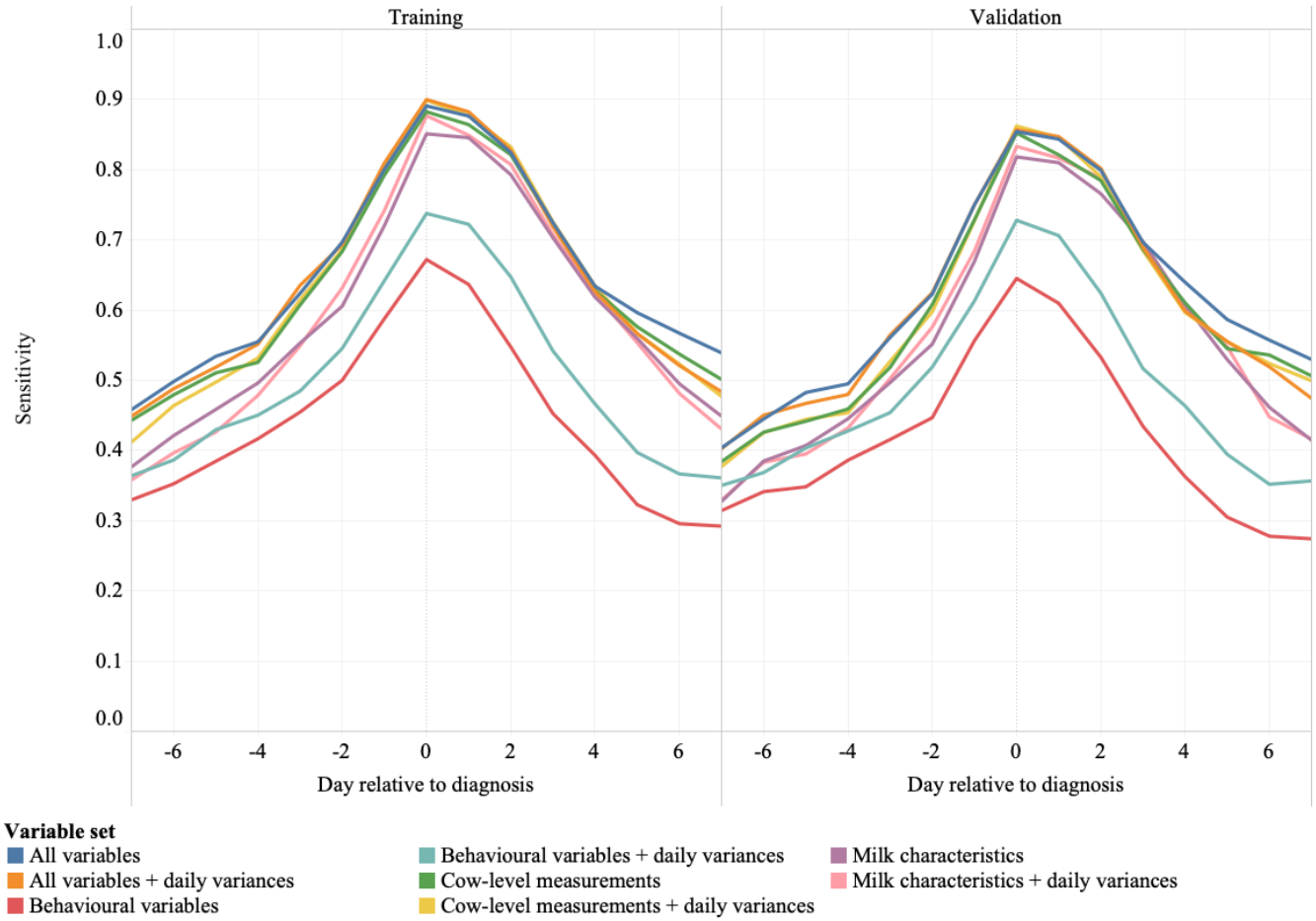


Figure 4-2. Mean sensitivity (proportion of CM animals, or animals that would be diagnosed with CM identified by the model) relative to the day of diagnosis for the detection of CM in dairy cows using a RNN. Different colours represent models trained using various subsets of recorded variables.

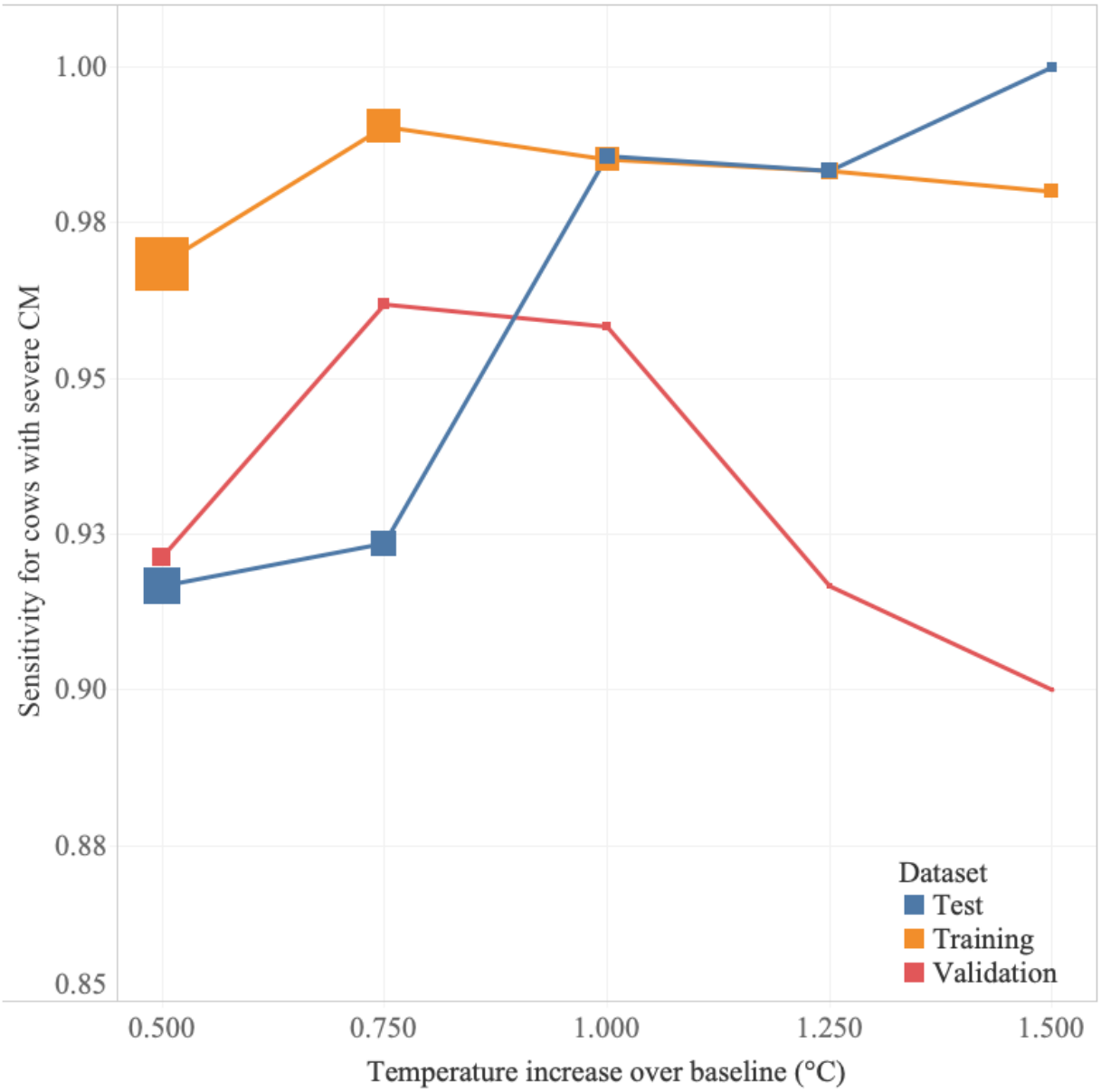


Figure 4-3. Mean sensitivity for the detection of cows with severe CM, as defined by an increase over the baseline in average milk temperature during the sick period. The different lines represent the different datasets; relative size of points reflect the number of animals identified as having severe CM.

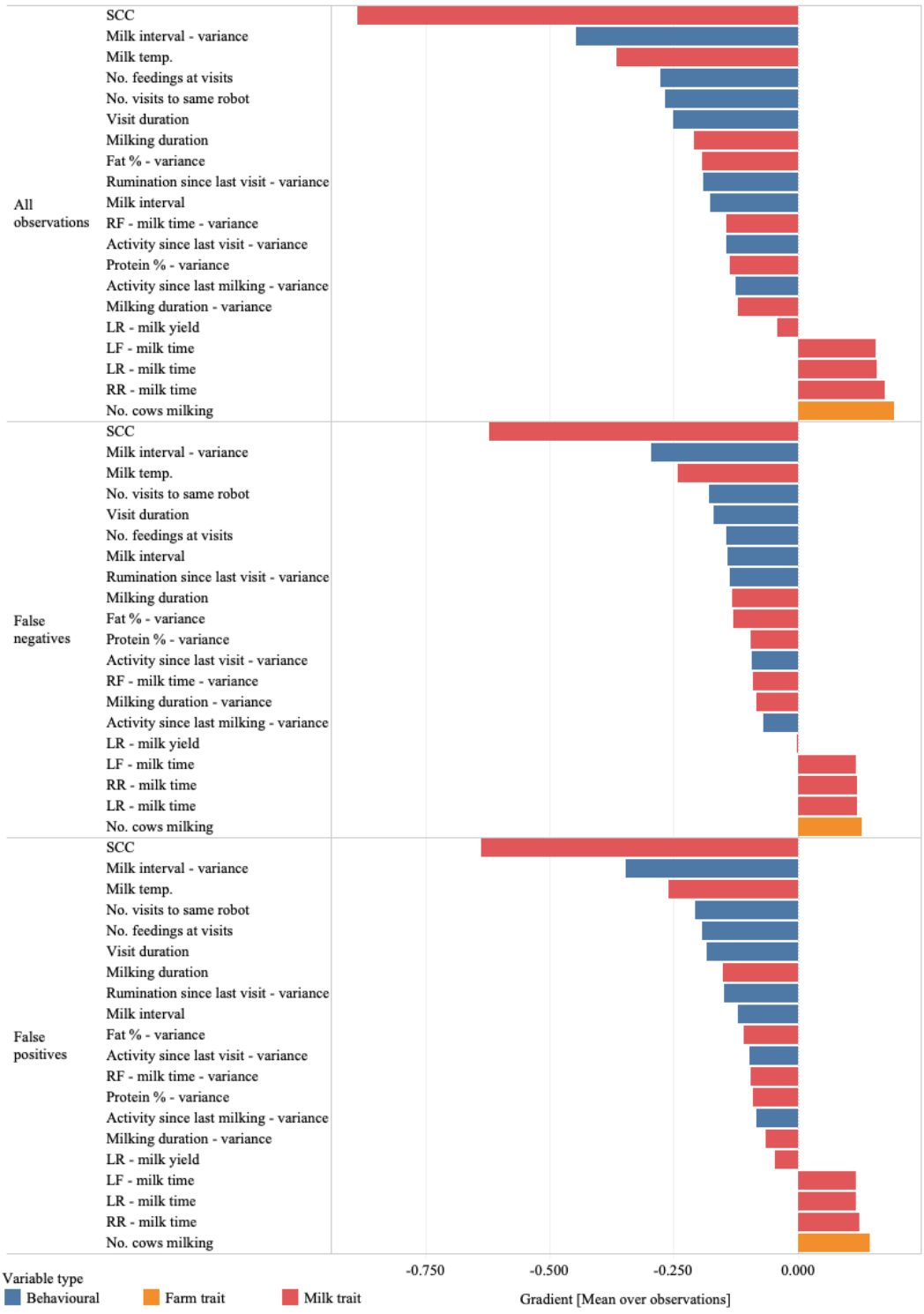


Figure 4-4. Top 20 variables as defined by the absolute value of the gradient, averaged over different subsets of observations for a RNN detecting the onset of CM.

CHAPTER 5: Data considerations for developing deep learning models for dairy applications

5.1. Abstract

With growing adoption of precision dairy technologies, the use of big data is becoming increasingly common in the dairy industry. The speed at which data are generated has led to increased interest in developing detection and predictive models for animal health and disease events using real time records. When combining data from multiple sources, statistical methods exist to account for the underlying heterogeneity in data collected from commercial farms, although its impact on predictive models is not known. We investigated how 4 different issues commonly seen in these large datasets impact the performance of deep recurrent neural networks (RNNs) trained to detect the onset of clinical mastitis (CM) in dairy cows. We simulated data by first sampling from real-world data and adding noise, then defining the association between predictor variables and CM while incorporating parameters to reflect underlying heterogeneity: 1) random effects to reflect unmeasured variability at the farm level; 2) random effects to reflect unmeasured variability at the cow level; 3) missed recording of CM cases; and 4) incomplete observations due to certain farms not having a somatic cell count sensor. Higher levels of unmeasured variability at the farm and cow levels resulted in reduced model performance, indicating that data collection and feature selection should be informed by previous knowledge of the associations between the outcome and predictors when possible, and that model performance may be limited when predictors are selected only from routinely collected data. However, even when 50% of CM cases were incorrectly recorded as CM-negative, model performance did not decrease, demonstrating that deep RNNs are robust to the level of

misclassification that would be typically encountered in dairy datasets. RNNs were also able to accurately detect CM onset even when a highly predictive variable, somatic cell count, was excluded from training and test data, but the models took longer to train. The effect of unmeasured variability on model performance demonstrates how predictors should be selected for RNNs, whereas RNNs appear to be very robust to misclassification in training data as well as missing variables.

5.2. Introduction

Dairy herds in the developed world have increased in size over the past few decades, a change that has been accompanied with increasing adoption of new technologies, such as automated milking systems (AMS; Barkema et al., 2015). In Australia, larger farms in particular increasingly adopt precision dairy technologies (Gargiulo et al., 2018), although this trend varies regionally due to differences in labour costs, milk price, and availability of technical support. Producers adopting precision dairy technologies do so primarily to facilitate management of health and production by monitoring an increasing number of parameters (Rutten et al., 2013; Borchers and Bewley, 2015; Tse et al., 2018). Implementation of these technologies results in the generation of a massive amount of data, which has driven growth in interest and research in applying 'big data' analytics for precision dairy farming, particularly for monitoring animal health and reproductive status (Lokhorst et al., 2019). An aspect of this research is the development of algorithms that are able to synthesize information from multiple sensors to create decision support tools that producers and advisors can use to improve management (Rutten et al., 2013; Cabrera et al., 2020). These algorithms are often developed by using machine learning models and are applied to various aspects of management, such as detecting disease and reproductive status; monitoring animal behaviour and feeding; and predicting water and electricity usage (Cockburn, 2020). Deep neural networks (DNNs) are a type of powerful predictive model that have seen considerable recent improvement due to advances in computational technologies, and have successfully been applied in the dairy context for monitoring feeding behaviour (Kuan et al., 2019), predicting milk yield (Liseune et al., 2021) and detection of clinical mastitis (CM; Ebrahimi et al., 2019; Hajaya et al., 2019; Naqvi et al., 2021), lameness (Wu et al., 2020) and bovine tuberculosis (Denholm et al., 2017).

Epidemiological studies of these diseases and behaviours using large datasets employ statistical methods that can account for the effects of hierarchical clustering in the data, unmeasured variables, imperfect outcome recording, and missing data. Conversely, no commonly used methods in the development of deep neural networks are able to account for this clustered structure and heterogeneity in collected data. To properly validate models, and effectively apply them in real world settings, it is important to first understand how these factors can impact performance and model predictions.

Data for large-scale epidemiological studies are often hierarchical, with correlations present at multiple levels of observation, such as multiple samples within quarters, quarters within cows, and cows within herds (Dohoo et al., 2014). These correlations can be accounted for by using mixed models and quantifying random effects at each level of observation, a methodology that has been extensively applied to understand many different outcomes such as reproductive performance (Dohoo et al., 2001), genetic heritability (Narayana et al., 2018), and diseases such as mastitis (Naqvi et al., 2018), and lameness (Salfer et al., 2018). A 2010 review on CM detection (Hogeveen et al., 2010a) identified 15 studies, of which only 5 used data from multiple farms. In a review of more recent publications (Slob et al., 2021), 7 out of 21 studies focused on CM detection used data from multiple farms. As technology adoption grows and data from different sources are further integrated, the proportion of studies developing models using data from multiple herds is likely to increase, making the between-farm heterogeneity an important factor that these models will have to consider.

Another source of heterogeneity that predictive models may encounter deals with the sampling methodology. Although methodology for microbiological sampling considerably affects diagnostic capability when determining disease outcomes (Dohoo et al., 2011a; Dohoo et

al., 2011b; Reyher & Dohoo, 2011), these are generally held constant within a study. A large study may include commercial herds in their sampling, in which case variation can also be introduced if data is producer-recorded. In the Nordic countries, completeness of producer records for treatment of CM can be as low as 45% when compared to veterinarian-supervised records (Wolff et al., 2012). With imperfect outcome recording, it becomes difficult to evaluate the performance of predictive models, because a false-positive may be a true case that was not recorded. Related to sampling, farms may also differ in the particular sensors they use, determining what variables can be measured and compared against other farms. In the interest of maximizing useable data, it may still be useful to include incomplete observations in model evaluation.

The overall aim of the present study was to determine how different sources of variation that are normally observed in a dairy farm setting affect the performance of a predictive RNN model for CM detection. The specific objectives were to: 1) determine the effect of unexplained variation at the cow and farm levels on model performance; 2) assess how imperfect outcome recording affects model performance; and 3) evaluate the models' robustness to incomplete observations.

5.3. Materials and Methods

5.3.1. Animals, herds, and data

Data used for this study were collected as part of a study with 197 farms using AMS across Canada (Matson et al., 2021). Milking records from each of these farms were available for a 6-month period ending between April and September 2019. Data collection involved AMS milking

records, animal activity and rumination, and treatment records. Of these farms, for consistency in measurements, 37 were selected that used the Lely AMS (Maassluis, South Holland, The Netherlands) and also had a sensor to measure somatic cell count (SCC) at each milking. The resulting dataset included 4,600 unique cow lactations, for a total of 579,842 cow-days observed. Cows of parity > 8 were excluded, as were observations > 365 days in milk (DIM).

Additionally, DHI test-day records from 91 herds across Canada from 2007-2016 were available and used to determine the mean and variance of production and SCC at each DIM, stratified by parity of the animals.

5.3.2. Data simulation

Because the observed data likely already included unmeasured variables, clustering, and imperfect recording, multiple datasets were simulated using these data where known values of the parameters could be specified. To increase the size of the dataset and avoid repetition caused by sampling with replacement from the observed data, noise was added to each observed variable when resampling observations from the original dataset.

5.3.2.1. Covariates

Seven predictor variables were used, of which 5 were at the cow level: mean daily SCC, total daily milk yield, mean milking interval, DIM measured daily, and parity; and 2 variables were at the herd level: mean environmental temperature (recorded at the nearest Environment Canada monitoring station) and number of cows per AMS unit.

The study population consisted of 100 simulated herds, each with 100 cows, created by sampling with replacement from the observed data and adding noise to each observation. Each

simulated herd was based on a single observed herd, randomly selected from the 37 observed herds, with replacement 100 times. After selecting a herd, randomly generated noise was added to the daily values for environmental temperature and lactating cows per AMS unit. Once the herd had been randomly selected, 100 animals were sampled with replacement from all animals observed for that herd. After selecting an animal, randomly generated noise was added to the daily values for milk yield, SCC, milking interval, and DIM. The distributions sampled to generate the values for noise added to each observation are presented in Table 5-1. Once 100 animals were generated from the particular herd, the process was repeated until data for 100 herds had been simulated.

5.3.2.2. Disease outcomes

The relationship between the selected predictor variables and the incidence of CM was defined to be as realistic as possible based on existing research on those variables as risk factors for CM. The association between CM incidence and the predictor variables was defined using a logit transformation of the probability of diagnosis computed using a function of the observed variables. Defining the relationship this way allowed for the incorporation of parameter values from literature, as mastitis epidemiology studies typically employ logistic or Poisson regression. In addition, literature-provided values could be used to quantitatively incorporate the unmeasured effects associated with cow- and farm-level random effects (Table 5-2). The formula to determine probability of CM diagnosis had the form:

$$\ln\left(\frac{p_{j,k,d}}{1-p_{j,k,d}}\right) = \left(\sum \beta_i \cdot g(x_i)\right) + \delta_j + \xi_k + \varepsilon \qquad (1)$$

where $p_{j,k,d}$ is the probability of cow *j* on farm *k* being diagnosed with CM at *d* DIM; $g(x_i)$ is a function applied to the value for predictor variable *i*; β_i is the coefficient for the *g*-transformed

value for predictor variable i; δ_j is the random effects term at the cow-level; ξ_k is the random effects term at the farm-level; and $\mathcal{E}$ is the residual observation-level error.

The functions for the baseline effects of DIM, temperature, and season (determined by observation date) were defined by directly using values from literature (Table 5-2). Temperature was converted to the temperature-humidity index assuming a constant humidity of 75%. The functions for the relationships between CM incidence and SCC, milk yield, number of lactating cows per AMS unit, and parity were also based on literature, although the specific values for the coefficients were chosen in a way to maintain the directionality of the association while yielding realistic rates for CM incidence between 20 and 30 per 100 cow-years (Olde Riekerink et al., 2008; Verbeke et al., 2014; Levison et al., 2016)). The relationship between milking interval and CM incidence was defined to depend on the number of milking intervals <8 h or >15 h in the previous week (Hammer et al., 2012). Although CM incidence depended on the number of milking intervals below and above the thresholds, only mean daily milk interval was used for model training, thus introducing temporal dependencies and latent covariates to better reflect real-world observations. Sampling distributions to generate the cow- and farm-level random effects were also based on literature, although skewed normal distributions were used to account for the skew that results when applying the inverse logit transformation. After applying the inverse logit transformation, cow-days where the predicted probability of CM diagnosis >0.50 were considered CM-positive. To account for imperfect recording, a window of 3 d prior to the identified CM onset day were also considered to be CM-positive (Naqvi et al., 2021). The predictor variables, the functions used to define their associations, and the relevant sources in literature are presented in Table 5-2.

5.3.2.3. Experiments

Four sets of experiments were defined to assess the effects of the following on model performance:

1. different levels of unexplained farm-level variance;
2. different levels of unexplained cow-level variance;
3. different levels of outcome misclassification (missed recordings of CM cases); and
4. incomplete observations due to certain variables not being measured.

For experiments 1 and 2, 3 different levels of farm- and cow-level variance were assessed: low (0), moderate (values from literature), and high (10x values from literature). In addition, training data were split into training and validation sets in 2 ways: by herd (75% of herds of training, 25% of herds for validation) or by time (first 75% of observations for training, last 25% of observations for validation). To simulate misclassification in experiment 3, a proportion of CM-positive days was randomly recoded as CM-negative. Three levels of misclassification proportions were assessed: low (0.10), moderate (0.25) and high (0.50). The misclassification was applied to either the training or testing data, for a total of 6 different datasets. For experiment 4, SCC was chosen as the variable to exclude, and was excluded from either the training or testing data by replacing all values with 0 after scaling.

In addition to these experiments, model performance was also assessed on 3 baseline scenarios: 1) completely random association between predictors and outcome based on a uniform sampling distribution; 2) realistic, moderate noise: moderate farm- and cow-level random effects, mean false-negative rate of CM diagnosis of 0.25 that varied by farm and SCC excluded from 58% of farms (Naqvi et al., 2021); and 3) realistic, high noise: high farm- and cow-level random effects, mean false-negative rate of CM diagnosis of 0.5 that varied by farm, and SCC excluded

from all farms. The varying coefficients associated with disease outcome simulation for each of the experiments and baseline scenarios are presented in Table 5-3.

5.3.3. Model development

5.3.3.1. Software packages

All data management and analyses were conducted in Python 3.7 (Van Rossum and Drake, 2009) using base functions and additional libraries. Initial data management was done using Pandas v.1.1.4 (Reback et al., 2020), followed by array computations using NumPy v.1.18.5 (Harris et al., 2020). Neural network models were developed and fit using Tensorflow v.2.3.1 (Abadi et al., 2016).

5.3.3.2. Recurrent neural networks

Recurrent neural networks are an extension of simple feed-forward neural networks that can handle time-series data where each observation within a sequence is not independent (Hochreiter and Schmidhuber, 1997). In particular, long short-term memory cells can be implemented within RNNs to better recognize long term temporal dependencies and improve model fitting (Hochreiter and Schmidhuber, 1997).

5.3.3.3. Model architecture and fitting

Prior to model fitting, the values for predictor variables were scaled using max-min scaling relative to the training set using the formula:

$$x' = \frac{x - mean(\vec{x}_{train})}{max(\vec{x}_{train}) - min(\vec{x}_{train})} \qquad (2)$$

where x refers to the original value of a variable in either of the training, validation, or test sets; $\vec{x}_{train}$ is the entire observed vector of that variable in the training set; and x' is the scale-transformed value of the observed value for that variable.

The RNNs in the present study had 3 layers:

1. input and masking – takes as input a sequence of observations on an individual cow during a single lactation with padded values being excluded from model optimization;
2. recurrent layer – a LSTM cell with 512 hidden units; and
3. output layer – a simple neural network hidden layer with 2 units, time distributed to yield a separate output for each time step, and softmax activation to ensure that output probabilities sum to 1.

Regularization to improve model generalizability was incorporated in the recurrent layer with parameter dropout set to 0.2, and L2 regularization with a coefficient of 0.001. Dropout randomly excludes the specified proportion of parameters during each model update step to reduce overfitting. L2 regularization reduces the magnitude of parameter estimates by forcing low estimates to zero.

Because the output classes were severely unbalanced (healthy cow-days greatly outnumbered CM-positive cow-days), each observation was weighted according to their output class, based on the ratio of healthy cow-days to CM-positive cow-days. By weighting each observations contribution to the loss function, equal emphasis was placed on predicting CM-negative and CM-positive animals. The categorical cross-entropy loss function was used with the Adam optimizer (Kingma and Ba, 2015). Models were trained for a maximum of 250 epochs, with early stopping employed when the validation balanced accuracy did not improve by at least

0.0001 for 25 epochs. In addition, the learning rate was reduced when validation set balanced accuracy plateaued for 5 epochs.

5.3.3.4. Model comparison and performance metrics

During the model comparison stage, false-positives were cow-days that were CM-negative, when the RNN model predicted CM-positive, while false-negatives were cow-days that were CM-positive, when the RNN model predicted CM-negative. With such a severe imbalance in the output classes, accuracy alone would not be representative of model performance, so additional performance metrics were also computed to compare model performance:

$$Accuracy = \frac{True\ positives + True\ negatives}{All\ positives + All\ negatives} \quad (3)$$

$$Sensitivity = \frac{True\ positives}{All\ positives} \quad (4)$$

$$Specificity = \frac{True\ negatives}{All\ negatives} \quad (5)$$

In addition, balanced accuracy was computed to control the model fitting procedure, calculated using the formula:

$$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2} \quad (6)$$

5.4. Results

5.4.1. Covariate simulation

Simulated values for the predictor variables were similar when compared to the original data (Table 5-4). The largest difference was between the mean number of milking cows in the original dataset and the simulated dataset, as the simulated herds were, on average, much smaller. Two of the 37 original herds were much larger than all the others and considerably skewed the mean to the right, but by only sampling 100 cows per herd, the impact of the larger herds was reduced. As this the mean number of milking cows was not used as a predictor for model training, its effect on the difference between the two datasets on model performance would be negligible.

The simulated means for number of cows per AMS milking unit, environmental temperature, DIM, SCC, and observed days per animal were slightly higher than in the original dataset, but the size of the interquartile ranges was very similar. Conversely, mean parity was lower in the simulated dataset, likely because the effect of extreme outliers on the means was reduced by sampling with replacement.

5.4.2. Experiment 1 – effects of unmeasured farm-level variability

Models trained with different levels of farm-level random effects all performed well (sensitivity and specificity >0.80 on all but one of the test sets; Table 5-5). Test set sensitivity decreased slightly as the magnitude of farm-level random effects increased, while test set specificities were more consistent, and a decrease was only observed at the highest level of farm-level random effects. In all cases, the accuracy metric was very close in value to the specificity,

because of the outcome class imbalance, such that the correct classifications were always more likely to be CM-negative.

When dividing the training data into training and validation sets by time, performance was lower as compared to training data split by herds on both validation and test sets. Model training took longest (122 epochs) on the dataset with no farm-level random effects, whereas training was stopped at 25 epochs when training and validation sets were split by time, and for the highest level of farm-level random effects. Although the number of training epochs required may be interpreted as an indication of model complexity, 25 epochs was also the specified patience parameter for early stopping of model training, so faster training could also indicate that no improvement in model performance was observed on the validation sets in those cases.

5.4.3. Experiment 2 – effects of unmeasured cow-level variability

Models trained with different levels of cow-level random effects also performed well in all cases (sensitivity and specificity > 0.80), except for at the highest level of cow-level random effects (sensitivity = 0.755; Table 5-6). When splitting the training data by time, test set sensitivity was highest when there were no cow-level random effects and decreased with more cow-level random effects; test set specificity decreased at the highest number of? cow-level random effects, and the test set accuracy again corresponded closely with the test set specificity. When splitting the training data by farms, both test set sensitivity and specificity were very similar between the low and moderate levels of cow-level random effects and decreased slightly at the highest level. Validation set sensitivity was slightly lower when the training/validation split was done by time, rather than by farms at all levels of random effects.

The number of epochs required for model training increased with the level of cow-level random effects when training data was split by farms, whereas when training data was split by time, model training only took 25 epochs for all levels. As described previously, 25 epochs were the patience parameter for the early stopping process and were, therefore, the minimum that models could be trained with. This would indicate that there was no improvement on validation set performance over the first 25 epochs of training.

5.4.4. Experiment 3 – effects of outcome misclassification

Model performance across different levels of outcome misclassification in both the training and testing data was good (sensitivity and specificity >0.80; Table 5-7), and by all metrics performance on all subsets of the data were similar. At all levels of false-negative classification in either the training or test data, training stopped at 25 epochs.

5.4.5. Experiment 4 – effects of missing observations in a predictor variable

When SCC was treated as missing in the training or testing datasets, model performance was still very good (sensitivity and specificity >0.90; Table 5-8). Performance metrics on the training, validation, and testing sets were similar when comparing a model trained where SCC was missing from the training data to a model trained including SCC, but excluding it from the testing datasets. Model training took longer when SCC was excluded from the training set (107 epochs) compared to when it was excluded from the testing set (98 epochs).

5.4.6. Baseline scenarios

When the association between CM and the predictors was completely random, the model always predicted cow-days to be CM-positive, resulting in a sensitivity of 1.00 and specificity of 0.00 (Table 5-9) in the training, validation, and test sets. Model training for baseline scenario 1 was completed in 36 epochs. In the moderate scenario, performance was poorer in the validation set, but was similar in the training and test sets with sensitivity values of 0.89 and 0.86 respectively, and specificity values of 0.91 and 0.90 respectively. This performance was also similar to the moderate levels of farm- and cow-level random effects (Tables 5-5 and 5-6), as well as the moderate level of outcome misclassification (Table 5-7), although model training was considerably faster, being completed in 66 epochs. In baseline scenario 3, with high levels of farm- and cow-level random effects, misclassification and SCC being excluded, performance was lower in training, validation, and test sets. Test set performance was poorer than in any of the other experiments, with both sensitivity and specificity <0.80, and model training stopped at 25 epochs.

5.5. Discussion

Models for the detection of disease on dairy farms using AMS developed using RNNs appear to be robust to many of the issues that affect the statistical analysis of these kinds of datasets that are often characterized by missing data. Although the results of this study need to be interpreted in the context of the simulation parameters, the findings provide an indication of how the performance of these models is affected, as increasingly large and varied sources of data are used for their development.

Although the effect of the identified issues could be assessed using completely randomly generated data, to make the findings as realistic and interpretable as possible, the simulation was started with records from real dairy farms. After resampling, data were perturbed by random values based on variability observed from real DHI data. By constraining the resampling size to be constant for all farms, we were able to ensure a balanced distribution of random effects across all herds. The largest difference was observed in the mean value for the number of milking cows per herd. The original data included a few, very large farms which would have disproportionately skewed the mean, but by resampling to include only 100 cows per herd, the effect of these outliers on the computed mean was reduced. Conversely, mean SCC was considerably higher in the simulated than in the original data, because by resampling from herds with less than 100 cows, the rare cows with very high SCC could have been included multiple times. Although these differences likely would not have had a strong effect on model fitting, they are important to keep in mind when interpreting our study findings.

Neural networks, and RNNs using LSTM cells in particular, are able to capture complex, non-linear, and time-dependent relationships between predictor variables and the outcome variable (Hochreiter and Schmidhuber, 1997). The function to relate the probability of CM to the predictor variables was made non-linear using the logit link and polynomial terms, while temporal dependencies were defined by comparing variable values at each time with the mean and variance of the same variable over the previous week. Although temperature and season would be correlated, they were both used as predictors to increase model complexity by including collinearity and different variable types. Variances for farm- and cow-level random effects were based on values from literature, but a skewed normal distribution was used to account for the skew introduced when converting from log odds of disease to probability. The

farm- and cow-level random effects, along with the residual error, added unpredictable noise to the data and represented the effects of unmeasured variables at the farm, cow, and observation levels respectively. When comparing model performance between the different training/validation dataset splits both validation and test set performance , in all cases was worse when the training and validation data were split over time, rather than by farm. When training data were split by time, model training stopped after 25 epochs, indicating that performance did not improve appreciably after the initial pass over the training dataset. This is likely because splitting the data by time may have broken the temporal dependencies required to accurately predict the incidence of CM. Although previous studies for CM prediction have split sequences by time (Miekley et al., 2013; Adriaens et al., 2018), these studies tended to focus on short-term forecasts. For timely detection of disease, the findings from the present study demonstrate that it is preferable to maintain entire sequences and divide the training data by groups (farms or individual animals), rather than over time.

As expected, test set performance decreased with both increasing farm-level random effects and cow-level random effects, while the number of epochs required for model training increased. In both cases, a marked decrease in performance was observed at the highest level compared to low and moderate levels, whereas the difference between the low and moderate levels was smaller. Although a researcher developing predictive models would not explicitly have to account for random effects in model development, it does provide some insight into how data collection should be conducted. Random effects, particularly in the multilevel model setting can be interpreted as the unmeasured effect of different levels of hierarchical clustering on the outcome of interest (Austin et al., 2001; Dohoo et al., 2014). Therefore, when collecting data for model development it is important to consider how much variability at different hierarchical

levels is captured in the measured variables, because farm-specific identifiers cannot be included if the goal is having a model that generalizes well. When analyzing the association between farm-specific variables and culling decisions in dairy herds, Haine et al. (2017) demonstrated that including contextual herd-level variables results in a decrease in unexplained herd-level variance (Haine et al., 2017). In the same study, including cow-level variables actually resulted in an increase in unexplained herd-level variance (Haine et al., 2017), suggesting that there are likely outcome-specific interactions between the farm- and cow-level measurements. In a study of reproductive performance, changing the outcome (intervals between calving and first service, between calving and conception, or first service and conception) resulted in considerable changes in the estimates for residual cow- and farm-level variance, when all other predictors were kept the same (Dohoo et al., 2001). The findings from the present study suggest that unmeasured variability at both the farm- and cow-levels can influence model performance, so variable selection prior to data collection should be informed by previous knowledge about different factors that can affect the outcome of interest. It is, therefore, important when designing studies that leverage data that are routinely collected (such as AMS or DHI data) to develop predictive models, to consider that model performance will ultimately be limited by variables that have been measured, and those that may be relevant but were not measured.

In the field of machine learning, outcome misclassification in training data is commonly referred to as label noise, and its consequences for model training, inference and performance have been extensively studied (Frénay and Verleysen, 2014). In particular, the type of label noise explored in the present study (CM-positive cases being misclassified as CM-negative), would be an example of class-dependent noise where each outcome class has a unique probability of being classified as the other (Algan and Ulusoy, 2020). When applied to image classification using

deep learning, class-dependent label noise has a much smaller impact on model performance when compared to locally-concentrated or feature-dependent noise (Algan and Ulusoy, 2020). Although methods have been proposed to make deep neural networks more robust to label noise (Patrini et al., 2017), sufficiently complex deep neural networks are remarkably robust to mislabelled training data (Rolnick et al., 2017). In the present study, model performance was not affected even at high levels of misclassification, where 50% of CM-positive observations were recorded as CM-negative. Even though for all levels of misclassification the model training stopped after 25 epochs, the models were still able to reach similar performance to the ones trained with low farm- and cow-level random effects. This suggests that even when using data from commercial dairy farms to develop disease prediction models, farm-specific differences in disease diagnosis and recording would only minimally affect model performance if using deep neural networks.

In a previous study using milk, behavioural and farm level characteristics to train RNNs for CM detection (Naqvi et al., 2021), only 37 out of 89 farms had SCC data available, but despite the incomplete observations in both the training and test data, model performance was good with a sensitivity and specificity of approximately 0.80. Excluding SCC measurements from either the training/validation set or the test set, produced models that performed well (sensitivity and specificity of approximately 0.90) on the hold-out test set. The primary difference between the 2 scenarios was the number of epochs required for model training: 107 when SCC was excluded from the training/validation set compared to 98 when SCC was excluded from the test set. This indicates that even when a highly predictive variable is excluded, a deep neural network is still able to detect CM with high accuracy, it just takes longer to learn the weaker associations of the remaining predictors with the outcome. The primary limitation to consider is that in the present

simulation, all the included predictors were associated with probability of CM, so even with the exclusion of SCC, the other predictors had a predictive association with the outcome. If, on the other hand. the remaining predictors were not associated with the probability of CM, it is likely that model performance would have suffered. Since RNNs can robustly ignore the noise in the data due to extraneous predictor variables, model development can proceed without having to carefully exclude unnecessary features or feature engineering (Laksana et al., 2020). Although this can facilitate the use of routinely collected data, it is still necessary to ensure that some of the included predictors contain information about the outcome of interest, even if not all the relevant predictors can be included.

Model performance on the 3 baseline scenarios was as expected: test set performance was best in scenario 2 (realistic, moderate noise) and was similar to performance in the moderate levels in the previous experiments; performance was worse than all other scenarios and experiments for baseline scenario 3 (realistic, high noise); and for scenario 1 (random) the model always predicted cows to be CM-positive, resulting in a sensitivity of 1.00 and a specificity of 0.0. As shown by Naqvi et al. (2021), real world performance of RNNs used for CM detection lies somewhere between scenario 2 (realistic, moderate noise) and scenario 3 (realistic, high noise).

As with any simulation study, the findings of the present study are limited by the parameters of the simulation. By sampling animals from observed data, and perturbing the records, we attempted to make the distribution of predictors as realistic as possible. Although the function relating the predictors to CM incidence was based on previous research, in a real world setting there will always be unexplained variability that is unlikely to be captured no matter how carefully the data collection is designed, such that perfect model performance would be

impossible. Model performance in all experiments in the present study was worse than many previous studies for CM detection (Kamphuis et al., 2008; Hogeveen et al., 2010b; Mammadova and Keskin, 2013), which suggests the simulated association between the predictors and CM was complex enough to not immediately be learned by the model.

5.6. Conclusions

This study demonstrates that RNN models are able to robustly handle many of the common issues encountered in epidemiological studies to produce accurate predictions about the onset of CM in a simulated dataset. We assessed 4 common issues in dairy epidemiological studies: 1) unexplained variability at the farm level; 2) unexplained variability at the cow level; 3) outcome misclassification (or label noise) relating to producer records of disease diagnoses and treatments; and 4) incomplete observations due to differences in available sensors between farms. Unexplained variability at the farm- and cow-level can reduce model performance and should be addressed when designing the data collection and selecting predictor features. Deep neural networks are robust to misclassification in training data, and CM is a rare enough event that incorrectly recorded cases have a negligible effect on model performance. Deep neural networks can accurately detect CM even in the absence of strongly associated predictor variables, given that the remaining predictors have some limited association with the outcome of interest. Recurrent neural networks can be trained for accurate detection of CM without the need to explicitly handle these issues as would be required for traditional statistical analyses.

5.7. References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, and G. Brain. 2016. TensorFlow: A system for large-scale machine learning. Pages 265–283 in Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA.

Adriaens, I., T. Huybrechts, B. Aernouts, K. Geerinckx, S. Piepers, B. De Ketelaere, and W. Saeys. 2018. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. J. Dairy Sci. doi:10.3168/JDS.2018-14696.

Algan, G., and İ. Ulusoy. 2020. Label noise types and their effects on deep learning. arXiv ID: 2003.10471.

Austin, P.C., V. Goel, and C. van Walraven. 2001. An introduction to multilevel regression models. Can. J. Public Heal. 2001 922 92:150–154. doi:10.1007/BF03404950.

Barkema, H.W., M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.-P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–45. doi:10.3168/jds.2015-9377.

Barkema, H.W., Y.H. Schukken, T.J.G.M. Lam, M.L. Beiboer, H. Wilmink, G. Benedictus, and A. Brand. 1998. Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. J. Dairy Sci. 81:411–419. doi:10.3168/jds.S0022-0302(98)75591-2.

Bilby, T. 2014. How Do I Determine?: How Do I Calculate Temperature-Humidity Index (THI)? Accessed June 20, 2021. https://www.progressivedairy.com/topics/herd-health/how-do-i-determine-how-do-i-calculate-temperature-humidity-index-thi.

Borchers, M.R., and J.M. Bewley. 2015. An assessment of producer precision dairy farming technology use, prepurchase considerations, and usefulness. J. Dairy Sci. 98:4198–4205. doi:10.3168/JDS.2014-8963.

Cabrera, V.E., J.A. Barrientos-Blanco, H. Delgado, and L. Fadul-Pacheco. 2020. Symposium review: Real-time continuous decision making using big data on dairy farms. J. Dairy Sci. 103:3856–3866. doi:10.3168/jds.2019-17145.

Cockburn, M. 2020. Review: Application and prospective discussion of machine learning for the management of dairy farms. Animals 10:1–22. doi:10.3390/ani10091690.

Denholm, S.J., T.N. McNeilly, G. Banos, M.P. Coffey, G.C. Russell, A. Bagnall, M.C. Mitchell, and E. Wall. 2017. Estimating genetic and phenotypic parameters of cellular immune-associated traits in dairy cows. J. Dairy Sci. 100:2850–2862. doi:10.3168/jds.2016-11679.

Dohoo, I., S. Andersen, R. Dingwell, K. Hand, D. Kelton, K. Leslie, Y. Schukken, and S. Godden. 2011a. Diagnosing intramammary infections: comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94:5515–22. doi:10.3168/jds.2011-4486.

Dohoo, I., W. Martin, and H. Stryhn. 2014. Introduction to clustered data. 2nd ed. S.M. McPike, ed. VER Inc., Charlottetown, PE, Canada.

Dohoo, I.R., J. Smith, S. Andersen, D.F. Kelton, and S. Godden. 2011b. Diagnosing intramammary infections: evaluation of definitions based on a single milk sample. J. Dairy Sci. 94:250–61. doi:10.3168/jds.2010-3559.

Dohoo, I.R., E. Tillard, H. Stryhn, and B. Faye. 2001. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island. Prev. Vet. Med. 50:127–144. doi:10.1016/S0167-5877(01)00191-X.

Ebrahimi, M., M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K.R. Petrovski. 2019. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Comput. Biol. Med. 114:103456. doi:10.1016/j.compbiomed.2019.103456.

Frénay, B., and M. Verleysen. 2014. Classification in the presence of label noise: A survey. IEEE Trans. Neural Networks Learn. Syst. 25:845–869. doi:10.1109/TNNLS.2013.2292894.

Gargiulo, J.I., C.R. Eastwood, S.C. Garcia, and N.A. Lyons. 2018. Dairy farmers with larger herd sizes adopt more precision dairy technologies. J. Dairy Sci. 101:5466–5473. doi:10.3168/JDS.2017-13324.

Haine, D., H. Delgado, R. Cue, A. Sewalem, K. Wade, R. Lacroix, D. Lefebvre, J. Arsenault, É. Bouchard, and J. Dubuc. 2017. Contextual herd factors associated with cow culling risk in Québec dairy herds: A multilevel analysis. Prev. Vet. Med. 144:7–12. doi:10.1016/J.PREVETMED.2017.05.014.

Hajaya, M.K., S. Samarasinghe, D. Kulasiri, and M.L. Benavides. 2019. Detection of dairy cattle mastitis: Modelling of milking features using deep neural networks. Pages 35–41 in 23rd International Congress on Modelling and Simulation - Supporting Evidence-Based Decision Making: The Role of Modelling and Simulation, MODSIM 2019. Modelling and Simulation Society of Australia and New Zealand Inc. (MSSANZ), Canberra, ACT, Australia.

Hammer, J., J. Morton, and K. Kerrisk. 2012. Quarter-milking-, quarter-, udder- and lactation-level risk factors and indicators for clinical mastitis during lactation in pasture-fed dairy cows managed in an automatic milking system. Aust. Vet. J. 90:167–174. doi:10.1111/j.1751-0813.2012.00917.x.

Harris, C.R., K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del R'\io, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T.E. Oliphant. 2020. Array programming with {NumPy}. Nature 585:357–362. doi:10.1038/s41586-020-2649-2.

Hochreiter, S., and J. Schmidhuber. 1997. Long short-term memory. Neural Comput. 9:1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010a. Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10:7991–8009. doi:10.3390/s100907991.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010b. Sensors and clinical mastitis-the quest for the perfect alert. Sensors. doi:10.3390/s100907991.

Kamphuis, C., D. Pietersma, R. van der Tol, M. Wiedemann, and H. Hogeveen. 2008. Using sensor data patterns from an automatic milking system to develop predictive variables for classifying clinical mastitis and abnormal milk. Comput. Electron. Agric. 62:169–181. doi:10.1016/j.compag.2007.12.009.

Kingma, D.P., and J.L. Ba. 2015. Adam: A method for stochastic optimization. Page in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR.

Kuan, C.Y., Y.C. Tsai, J.T. Hsu, S.T. Ding, and T. Te Lin. 2019. An imaging system based on

deep learning for monitoring the feeding behavior of dairy cows. Page in ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, Boston, MA, USA.

Laksana, E., M. Aczon, L. Ho, C. Carlin, D. Ledbetter, and R. Wetzel. 2020. The impact of extraneous features on the performance of recurrent neural network models in clinical tasks. J. Biomed. Inform. 102:103351. doi:10.1016/J.JBI.2019.103351.

Levison, L.J., E.K. Miller-Cushon, A.L. Tucker, R. Bergeron, K.E. Leslie, H.W. Barkema, and T.J. DeVries. 2016. Incidence rate of pathogen-specific clinical mastitis on conventional and organic Canadian dairy farms. J. Dairy Sci 99:1341–1350. doi:10.3168/jds.2015-9809.

Liseune, A., M. Salamone, D. Van den Poel, B. van Ranst, and M. Hostens. 2021. Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning. Comput. Electron. Agric. 180:105904. doi:10.1016/j.compag.2020.105904.

Lokhorst, C., R.M. De Mol, and C. Kamphuis. 2019. Invited review: Big Data in precision dairy farming. Animal 13:1519–1528. doi:10.1017/S1751731118003439.

Mammadova, N., and I. Keskin. 2013. Application of the support vector machine to predict subclinical mastitis in dairy cattle. Sci. World J. 2013:603897. doi:10.1155/2013/603897.

Matson, R.D., M.T.M. King, T.F. Duffield, D.E. Santschi, K. Orsel, E.A. Pajor, G.B. Penner, T. Mutsvangwa, and T.J. DeVries. 2021. Benchmarking of farms with automated milking systems in Canada and associations with milk production and quality. J. Dairy Sci. doi:10.3168/jds.2020-20065.

Miekley, B., I. Traulsen, and J. Krieter. 2013. Principal component analysis for the early detection of mastitis and lameness in dairy cows. J. Dairy Res. 80:335–343. doi:10.1017/S0022029913000290.

Morse, D., M.A. DeLorenzo, C.J. Wilcox, R.J. Collier, R.P. Natzke, and D.R. Bray. 1988. Climatic effects on occurrence of clinical mastitis. J. Dairy Sci. 71:848–853. doi:10.3168/jds.S0022-0302(88)79626-5.

Naqvi, S.A., J. De Buck, S. Dufour, and H.W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101:3233–3247. doi:10.3168/jds.2017-13579.

Naqvi, S.A., M.T.M. King, R.D. Matson, T.J. DeVries, R. Deardon, and H.W. Barkema. 2021. Mastitis detection with recurrent neural networks in farms using automated milking systems. Submitted for peer review in Comput. Electron. Agric.

Narayana, S.G., F. Miglior, S.A. Naqvi, F. Malchiodi, P. Martin, and H.W. Barkema. 2018. Genetic analysis of subclinical mastitis in early lactation of heifers using both linear and threshold models. J. Dairy Sci. 101:11120–11131. doi:10.3168/jds.2018-15126.

Olde Riekerink, R.G.M., H.W. Barkema, D.F. Kelton, and D.T. Scholl. 2008. Incidence rate of clinical mastitis on Canadian dairy farms. J. Dairy Sci. 91:1366–1377. doi:10.3168/jds.2007-0757.

Patrini, G., A. Rozza, A. Krishna Menon, R. Nock, and L. Qu. 2017. Making deep neural networks robust to label noise: A loss correction approach. arXiv 1609.03683.

Reback, J., W. McKinney, jbrockmendel, J. Van den Bossche, T. Augspurger, P. Cloud, gfyoung, Sinhrks, S. Hawkins, M. Roeschke, A. Klein, T. Petersen, J. Tratner, C. She, W. Ayd, S. Naveh, M. Garcia, J. Schendel, A. Hayden, D. Saxton, V. Jancauskas, A. McMaster, P. Battiston, S. Seabold, chris-b1, h-vetinari, K. Dong, S. Hoyer, W. Overmeire, and M. Gorelli. 2020. pandas-dev/pandas: Pandas 1.1.4. doi:10.5281/zenodo.4161697.

Reyher, K.K., and I.R. Dohoo. 2011. Diagnosing intramammary infections: Evaluation of composite milk samples to detect intramammary infections. J. Dairy Sci. 94:3387–3396.

doi:10.3168/jds.2010-3907.

Riekerink, R.G.M.O., H.W. Barkema, and H. Stryhn. 2007. The effect of season on somatic cell count and the incidence of clinical mastitis. J. Dairy Sci. 90:1704–1715. doi:10.3168/jds.2006-567.

Rolnick, D., A. Veit, S. Belongie, and N. Shavit. 2017. Deep learning is robust to massive label noise. arXiv 1705.10694.

Van Rossum, G., and F.L. Drake. 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Rutten, C.J.J., A.G.J.G.J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. doi:10.3168/jds.2012-6107.

Salfer, J.A., J.M. Siewert, and M.I. Endres. 2018. Housing, management characteristics, and factors associated with lameness, hock lesion, and hygiene of lactating dairy cattle on Upper Midwest United States dairy farms using automatic milking systems. J. Dairy Sci. 101:8586–8594. doi:10.3168/jds.2017-13925.

Slob, N., C. Catal, and A. Kassahun. 2021. Application of machine learning to improve dairy farm management: A systematic literature review. Prev. Vet. Med. 187:105237. doi:10.1016/j.prevetmed.2020.105237.

Tse, C., H.W. Barkema, T.J. DeVries, J. Rushen, and E.A. Pajor. 2018. Impact of automatic milking systems on dairy cattle producers' reports of milking labour management, milk production and milk quality. animal 12:2649–2656. doi:10.1017/S1751731118000654.

Verbeke, J., S. Piepers, K. Supré, S. De Vliegher, and S. De Vliegher. 2014. Pathogen-specific incidence rate of clinical mastitis in Flemish dairy herds , severity , and association with

herd hygiene. J. Dairy Sci. 97:6926–6934. doi:10.3168/jds.2014-8173.

Vitali, A., U. Bernabucci, A. Nardone, and N. Lacetera. 2016. Effect of season, month and temperature humidity index on the occurrence of clinical mastitis in dairy heifers. Adv. Anim. Biosci. 7:250–252. doi:10.1017/S2040470016000315.

Wolff, C., M. Espetvedt, A.K. Lind, S. Rintakoski, A. Egenvall, A. Lindberg, and U. Emanuelson. 2012. Completeness of the disease recording systems for dairy cows in Denmark, Finland, Norway and Sweden with special reference to clinical mastitis. BMC Vet. Res. 8:1–11. doi:10.1186/1746-6148-8-131.

Wu, D., Q. Wu, X. Yin, B. Jiang, H. Wang, D. He, and H. Song. 2020. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. Biosyst. Eng. 189:150–163. doi:10.1016/j.biosystemseng.2019.11.017.

Table 5-1. Sampling distributions for the noise added to each variable when sampling with replacement from observed milking records from 37 herds, and 4,600 cow-lactations.

Variable	**Sampling distribution**
Farm level	
Environmental temperature	$\mathcal{N}(0, 0.25)$
Cows milking per AMS unit	$\mathcal{N}(0, 1.50)$
Cow level	
Milk yield[1]	$\mathcal{N}(0, {}^{\sigma_{Lact,DIM}}/_{2})$
Somatic cell count[1]	$\mathcal{N}(0, {}^{\sigma_{Lact,DIM}}/_{2})$
Milking interval	$\mathcal{N}(0, 1.50)$
Days in milk[2]	$unif\{-20, 20\}$

[1]The observed variance was based on parity- and DIM-specific values in the 2006-2016 DHI dataset.
[2]Only one adjustment was made per-animal, the first observed DIM of the animal was shifted by the randomly generated value to a minimum of 0, and the rest were adjusted sequentially.

Table 5-2. Sampling distributions used for the predictor variables in equation (1) and the functions used to compute their contribution to the simulated log odds of mastitis diagnosis.

Variable	Function/Sampling distribution	Reference
DIM	$-0.002 \cdot \mathrm{DIM}^2 - 1.2$	(Barkema et al., 1998)
Temperature[1]		
Temperature humidity index Relative humidity = 75%	$THI = 0.8 \cdot Temp℃ + 0.75 \cdot (Temp℃ - 14.4) + 46.4$	(Bilby, 2014)
Temperature	$0.755 \cdot THI - 60$	(Morse et al., 1988)
Season (each season is an indicator variable)	$0.00258 \cdot winter + 0.00283 \cdot spring + 0.00315 \cdot summer + 0.00254 \cdot fall$	(Vitali et al., 2016)
Parity	$0.025 \cdot \mathrm{parity}$	(Hammer et al., 2012)
No. cows milking per AMS unit	$0.005 \cdot \mathrm{no.\,cows\ milking\ per\ AMS\ unit}$	
Milking interval		
Milking intervals < 8h	$-1 \cdot \sum_{i=DIM-7}^{DIM} \mathrm{no.\,milking\ intervals} < 8\mathrm{h\ in}\ i$	(Hammer et al., 2012)
Milking intervals > 15h	$0.1 \cdot \sum_{i=DIM-7}^{DIM} \mathrm{no.\,milking\ intervals} > 15\mathrm{h\ in}\ i$	(Hammer et al., 2012)
Somatic cell count (SCC)[2]	$\mathrm{if}\ \ln(\mathrm{SCC})_{\mathrm{DIM}} > Mean_{\ln(SCC)-7d} \Rightarrow 2 \cdot \frac{(\ln(\mathrm{SCC})_{\mathrm{DIM}} - Mean_{ln(SCC)-7d})}{SD_{ln(SCC)-7d}}$	(Hammer et al., 2012)
Milk yield[2]	$\mathrm{if\ Milk\ yield}_{\mathrm{DIM}} < Mean_{MY-7d} \Rightarrow -0.01 \cdot \frac{(\mathrm{Milk\ yield}_{\mathrm{DIM}} - Mean_{MY-7d})}{SD_{MY-7d}}$	(Hammer et al., 2012)
Farm-level random effect	$\mathcal{N}(0, \sigma_j, -1.5)$	(Riekerink et al., 2007)
Cow-level random effect	$\mathcal{N}(0, \sigma_k, -1.5)$	(Riekerink et al., 2007)
Residual error	$\mathcal{N}(0, \pi^2/3, -1.5)$	(Dohoo et al., 2014)

[1]The effect of temperature is based on the temperature humidity index (THI) assuming a relative humidity of 75%.
[2]The effect of milk yield and SCC are based on comparisons with the arithmetic and geometric means respectively over the previous 7-d period.

Table 5-3. Varying parameters for the simulation experiments to determine the effects of the parameters on model performance. SD = standard deviation; j = farm identifier; k = cow identifier.

Variable	**Experiment**						
	Baseline 1 (random)[1]	**Baseline 2 (realistic, moderate noise)**[2]	**Baseline 3 (realistic, high noise)**[2]	**1**	**2**	**3**	**4**
σ_j *SD of farm-level random effects*	-	0.56	5.60	Low: 0 Moderate: 0.56 High: 5.60	0.56	0.56	0.56
σ_k *SD of cow-level random effects*	-	0.89	8.86	0.56	Low: 0 Moderate: 0.89 High: 8.86	0.89	0.89
Misclassification proportion *False-negative rate for outcome recording*	-	$unif\{0.15, 0.35\}$	$unif\{0.45, 0.55\}$	0	0	Low: 0.10 Moderate: 0.25 High: 0.50	0

[1]CM incidence was determined randomly using a $unif\{0, 1\}$ sampling distribution and setting clinical mastitis incidence to 25 cases per 100 cow-years at risk.
[2]False-negative rates were randomly assigned to each farm using the described sampling distributions.

Table 5-4. Mean (IQR) of predictor variables from the original data, and the simulated data created by sampling and perturbing the original data. Means were first calculated for each herd or cow-lactation over the entire observation period, and the reported mean (IQR) represents the unweighted aggregation over all herds or cow-lactations between December 2018 and August 2019.

Variable	**Original dataset**	**Simulated dataset**
No. herds	37	100
No. cow-lactations	4,453	10,000
Herd level		
No. cows milking per herd	130.70 (61.85 – 114.02)	66.03 (61.58 – 71.82)
No. cows milking per AMS unit	46.74 (36.43 – 54.65)	48.09 (42.67 – 55.15)
Environmental temperature (°C)	2.42 (0.98 – 3.29)	3.00 (1.21 – 4.43)
Animal level		
Parity	2.44 (1.0 – 3.0)	2.30 (1.0 – 3.0)
DIM	156.86 (63.50 – 246.50)	160.53 (70.32 – 249.01)
Total milk yield (kg/d)	34.73 (27.02 – 42.41)	34.79 (26.51 – 43.07)
SCC[1] (1000 cells/mL)	204.62 (70.89 – 193.80)	251.24 (154.19 – 277.53)
Milking interval (h)	9.97 (7.76 – 11.44)	9.56 (7.52 – 11.01)
Observed days	122.27 (57.0 – 181.0)	123.02 (55.0 – 188.0)

[1] Somatic cell count

Table 5-5. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of farm-level random effects, and two different splits of training and valida tion data. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. σ = standard deviation of farm-level random effects where the random effects follow a skewed normal distribution $\mathcal{N}(0, \sigma, -1.5)$.

Farm-level random effects	Training-validation split	No. epochs	Dataset	Sensitivity	Specificity	Accuracy
Low $\sigma = 0$	First 75% days training; last 25% validation	25	Training	0.839 (0.821 - 0.857)	0.780 (0.779 - 0.781)	0.780 (0.779 - 0.782)
			Validation	0.855 (0.828 - 0.879)	0.880 (0.879 - 0.881)	0.880 (0.879 - 0.881)
			Test	0.813 (0.785 - 0.839)	0.851 (0.850 - 0.852)	0.851 (0.850 - 0.852)
	75% farms training; 25% validation	122	Training	0.938 (0.926 - 0.949)	0.889 (0.888 - 0.890)	0.889 (0.888 - 0.89)
			Validation	0.902 (0.879 - 0.923)	0.886 (0.885 - 0.887)	0.886 (0.885 - 0.887)
			Test	0.88 (0.856 - 0.901)	0.897 (0.896 - 0.898)	0.897 (0.896 - 0.898)
Moderate $\sigma = 0.56$	First 75% days training; last 25% validation	25	Training	0.795 (0.774 - 0.815)	0.807 (0.806 - 0.809)	0.807 (0.806 - 0.808)
			Validation	0.789 (0.758 - 0.818)	0.903 (0.903 - 0.904)	0.903 (0.903 - 0.904)
			Test	0.783 (0.751 - 0.813)	0.877 (0.876 - 0.878)	0.877 (0.876 - 0.878)
	75% farms training; 25% validation	102	Training	0.919 (0.904 - 0.932)	0.908 (0.907 - 0.908)	0.908 (0.907 - 0.908)
			Validation	0.825 (0.797 - 0.852)	0.920 (0.919 - 0.921)	0.920 (0.919 - 0.921)
			Test	0.865 (0.837 - 0.889)	0.917 (0.916 - 0.917)	0.916 (0.916 - 0.917)
High $\sigma = 5.60$	First 75% days training; last 25% validation	25	Training	0.776 (0.747 - 0.802)	0.786 (0.784 - 0.787)	0.786 (0.784 - 0.787)
			Validation	0.844 (0.796 - 0.884)	0.887 (0.887 - 0.888)	0.887 (0.887 - 0.888)
			Test	0.788 (0.763 - 0.813)	0.854 (0.853 - 0.855)	0.853 (0.852 - 0.855)
	75% farms training; 25% validation	25	Training	0.869 (0.844 - 0.892)	0.864 (0.863 - 0.865)	0.864 (0.863 - 0.865)
			Validation	0.842 (0.803 - 0.876)	0.865 (0.864 - 0.866)	0.865 (0.864 - 0.866)
			Test	0.864 (0.842 - 0.884)	0.854 (0.853 - 0.855)	0.854 (0.853 - 0.855)

Table 5-6. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of cow-level random effects, and two different splits of training and validation data. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. σ = standard deviation of cow-level random effects where the random effects follow a skewed normal distribution $\mathcal{N}(0, \sigma, -1.5)$.

Cow-level random effects	Training-validation split	No. epochs	Dataset	Sensitivity	Specificity	Accuracy
Low $\sigma = 0$	First 75% days training; last 25% validation	25	Training	0.815 (0.796 - 0.832)	0.776 (0.775 - 0.777)	0.776 (0.775 - 0.778)
			Validation	0.868 (0.842 - 0.890)	0.870 (0.869 - 0.871)	0.870 (0.869 - 0.871)
			Test	0.836 (0.811 - 0.858)	0.847 (0.846 - 0.848)	0.847 (0.846 - 0.848)
	75% farms training; 25% validation	106	Training	0.893 (0.877 - 0.907)	0.881 (0.880 - 0.882)	0.881 (0.880 - 0.882)
			Validation	0.902 (0.881 - 0.920)	0.890 (0.889 - 0.891)	0.890 (0.889 - 0.891)
			Test	0.894 (0.873 - 0.912)	0.888 (0.887 - 0.889)	0.888 (0.888 - 0.889)
Moderate $\sigma = 0.89$	First 75% days training; last 25% validation	25	Training	0.803 (0.783 - 0.822)	0.789 (0.788 - 0.791)	0.789 (0.788 - 0.791)
			Validation	0.862 (0.834 - 0.887)	0.893 (0.892 - 0.894)	0.893 (0.892 - 0.894)
			Test	0.772 (0.740 - 0.801)	0.859 (0.858 - 0.860)	0.859 (0.858 - 0.860)
	75% farms training; 25% validation	138	Training	0.942 (0.929 - 0.953)	0.896 (0.895 - 0.896)	0.896 (0.895 - 0.896)
			Validation	0.936 (0.915 - 0.953)	0.886 (0.885 - 0.887)	0.886 (0.885 - 0.887)
			Test	0.913 (0.890 - 0.932)	0.888 (0.887 - 0.889)	0.888 (0.887 - 0.889)
High $\sigma = 8.86$	First 75% days training; last 25% validation	25	Training	0.847 (0.826 - 0.866)	0.755 (0.754 - 0.756)	0.755 (0.754 - 0.756)
			Validation	0.815 (0.787 - 0.840)	0.868 (0.867 - 0.869)	0.868 (0.867 - 0.869)
			Test	0.679 (0.637 - 0.719)	0.816 (0.815 - 0.817)	0.816 (0.814 - 0.817)
	75% farms training; 25% validation	144	Training	0.896 (0.879 - 0.911)	0.898 (0.898 - 0.899)	0.898 (0.898 - 0.899)
			Validation	0.848 (0.818 - 0.874)	0.905 (0.904 - 0.905)	0.904 (0.903 - 0.905)
			Test	0.755 (0.716 - 0.792)	0.896 (0.895 - 0.897)	0.896 (0.895 - 0.897)

Table 5-7. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets with varying levels of misclassification, where randomly selected mastitis cases were recoded to healthy in either the training/validation set or the test set. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals. FNR = false-negative rate (proportion of clinical mastitis cases incorrectly recorded as healthy).

Misclassification rate	Dataset with misclassification	No. epochs	Dataset	Sensitivity	Specificity	Accuracy
Low $FNR = 0.10$	Training/Validation	25	Training	0.884 (0.865 - 0.901)	0.869 (0.868 - 0.870)	0.869 (0.868 - 0.870)
			Validation	0.887 (0.859 - 0.911)	0.872 (0.871 - 0.873)	0.872 (0.871 - 0.873)
			Test	0.887 (0.863 - 0.907)	0.868 (0.867 - 0.869)	0.868 (0.867 - 0.869)
	Test	25	Training	0.871 (0.852 - 0.888)	0.881 (0.881 - 0.882)	0.881 (0.881 - 0.882)
			Validation	0.858 (0.830 - 0.884)	0.884 (0.883 - 0.886)	0.884 (0.883 - 0.885)
			Test	0.872 (0.848 - 0.894)	0.880 (0.879 - 0.881)	0.880 (0.879 - 0.881)
Moderate $FNR = 0.25$	Training/Validation	25	Training	0.864 (0.841 - 0.884)	0.866 (0.865 - 0.867)	0.866 (0.865 - 0.867)
			Validation	0.823 (0.788 - 0.855)	0.869 (0.868 - 0.870)	0.869 (0.868 - 0.870)
			Test	0.857 (0.831 - 0.880)	0.869 (0.868 - 0.870)	0.869 (0.868 - 0.870)
	Test	25	Training	0.864 (0.845 - 0.882)	0.889 (0.888 - 0.889)	0.889 (0.888 - 0.889)
			Validation	0.866 (0.838 - 0.890)	0.891 (0.890 - 0.892)	0.891 (0.890 - 0.892)
			Test	0.868 (0.843 - 0.890)	0.886 (0.885 - 0.887)	0.886 (0.885 - 0.887)
High $FNR = 0.50$	Training/Validation	25	Training	0.882 (0.856 - 0.905)	0.852 (0.851 - 0.853)	0.852 (0.851 - 0.853)
			Validation	0.879 (0.838 - 0.912)	0.858 (0.856 - 0.859)	0.858 (0.856 - 0.859)
			Test	0.872 (0.848 - 0.894)	0.855 (0.854 - 0.857)	0.855 (0.854 - 0.857)
	Test	25	Training	0.874 (0.856 - 0.891)	0.877 (0.876 - 0.877)	0.877 (0.876 - 0.877)
			Validation	0.860 (0.831 - 0.885)	0.881 (0.880 - 0.882)	0.881 (0.879 - 0.882)
			Test	0.874 (0.849 - 0.895)	0.876 (0.875 - 0.877)	0.876 (0.875 - 0.877)

Table 5-8. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in datasets when a highly predictive variable (SCC) is excluded from either the training/validation or test data by replacing all values with 0. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals.

Dataset with missing variable	No. epochs	Dataset	Sensitivity	Specificity	Accuracy
Training/validation	107	Training	0.909 (0.893 - 0.924)	0.902 (0.902 - 0.903)	0.902 (0.902 - 0.903)
		Validation	0.843 (0.818 - 0.866)	0.908 (0.908 - 0.909)	0.908 (0.907 - 0.909)
		Test	0.887 (0.862 - 0.909)	0.901 (0.900 - 0.902)	0.901 (0.900 - 0.902)
Test	98	Training	0.911 (0.894 - 0.925)	0.907 (0.907 - 0.908)	0.907 (0.907 - 0.908)
		Validation	0.841 (0.815 - 0.864)	0.913 (0.912 - 0.914)	0.913 (0.912 - 0.913)
		Test	0.891 (0.867 - 0.912)	0.905 (0.904 - 0.906)	0.905 (0.904 - 0.906)

Table 5-9. Performance metrics (95% CI) and number of training epochs for recurrent neural networks trained to detect the occurrence of clinical mastitis in 3 different baseline scenarios: 1) completely random association; 2) realistic, moderate noise and 3) realistic, high noise. 95% CIs are computed using the Clopper-Pearson method for exact binomial intervals.

Baseline scenario	No. epochs	Dataset	Sensitivity	Specificity	Accuracy
1	36	Training	1.000	0.000	0.003
		Validation	1.000	0.000	0.003
		Test	1.000	0.000	0.003
2	66	Training	0.893 (0.874 - 0.91)	0.905 (0.904 - 0.906)	0.905 (0.904 - 0.906)
		Validation	0.798 (0.76 - 0.832)	0.900 (0.899 - 0.901)	0.900 (0.899 - 0.901)
		Test	0.862 (0.830 - 0.890)	0.895 (0.894 - 0.896)	0.895 (0.894 - 0.896)
3	25	Training	0.841 (0.801 - 0.876)	0.817 (0.816 - 0.818)	0.817 (0.816 - 0.818)
		Validation	0.842 (0.780 - 0.891)	0.813 (0.812 - 0.814)	0.813 (0.812 - 0.814)
		Test	0.766 (0.707 - 0.818)	0.789 (0.788 - 0.791)	0.789 (0.788 - 0.791)

CHAPTER 6: Cytokines and chemokines in pediatric appendicitis: A multiplex analysis of inflammatory protein mediators

6.1. Abstract

Objectives. We aimed to demonstrate the potential of precision medicine to describe the inflammatory landscape present in children with suspected appendicitis. Our primary objective was to determine levels of seven inflammatory protein mediators previously associated with intra-abdominal inflammation (C-reactive protein—CRP, procalcitonin—PCT, interleukin-6 (IL), IL-8, IL-10, monocyte chemoattractant protein-1—MCP-1, and serum amyloid A—SAA) in a cohort of children with suspected appendicitis. Subsequently, using a multiplex proteomics approach, we examined an expansive array of novel candidate cytokine and chemokines within this population. *Methods*. We performed a secondary analysis of targeted proteomics data from Alberta Sepsis Network studies. Plasma mediator levels, analyzed by Luminex multiplex assays, were evaluated in children aged 5-17 years with non-appendicitis abdominal pain (NAAP), acute appendicitis (AA), and non-appendicitis sepsis (NAS). We used multivariate regression analysis to evaluate the seven target proteins, followed by decision tree and heat mapping analyses for all proteins evaluated. *Results*. 185 children were included: 83 with NAAP, 79 AA, and 23 NAS. Plasma levels of IL-6, CRP, MCP-1, PCT, and SAA were significantly different in children with AA compared to those with NAAP ($P < 0.001$). Expansive proteomic analysis demonstrated 6 patterns in inflammatory mediator profiles based on severity of illness. A decision tree incorporating the proteins CRP, ferritin, SAA, regulated on activation normal T-cell expressed and secreted (RANTES), monokine induced by gamma interferon (MIG), and PCT demonstrated

excellent specificity (0.920) and negative predictive value (0.882) for children with appendicitis. *Conclusions*. Multiplex proteomic analyses described the inflammatory landscape of children presenting to the ED with suspected appendicitis. We have demonstrated the feasibility of this approach to identify potential novel candidate cytokines/chemokine patterns associated with a specific illness (appendicitis) amongst those with a broad ED presentation (abdominal pain). This approach can be modelled for future research initiatives in pediatric emergency medicine.

6.2. Introduction

Appendicitis results in both local and systemic inflammatory changes, which often clinically manifest with right lower quadrant (RLQ) abdominal pain, fever, nausea/vomiting, and anorexia (Bundy et al., 2007) and, left untreated, can progress over the course of the illness to peritonitis, abscess formation, sepsis, and death (Blomqvist et al., 2001; Andersson and Andersson 2011; Andersson 2013). Not surprisingly, clinicians take advantage of this inflammatory landscape by including laboratory markers as part of the standard workup of children presenting to the Emergency Department (ED) with abdominal pain and suspected appendicitis; most commonly, this includes white blood cell count (WBC), neutrophil count (NC), C-reactive protein (CRP), and/or procalcitonin (PCT) (Thompson et al., 2015). While elevated levels of such markers certainly help to *support* a clinical suspicion, their individual test characteristics (sensitivity, specificity, and predictive values) are suboptimal for use as diagnostic tests.

Attempts to identify novel appendicitis-specific biomarkers have significantly increased over the last decade. Interleukins (IL) 6 (Zviedre et al., 2016; Kharbanda et al., 2011; Anielski et al., 2010; Groselj-Grenc et al., 2007; Sack et al., 2006; Yildirim et al., 2006) and 10 (Zviedre et al., 2016; Yildirim et al., 2006; Hachim and Ahmed 2006) have been the subject of multiple recent studies, as has serum amyloid A (SAA) (Andersson et al., 2014; Sit et al., 2014). While offering some promise, the overall accuracy of these tests remains to be determined. Furthermore, the majority of attempts to identify appendicitis-specific biomarkers have focused on individual proteins. Given the diverse etiological causes of abdominal pain in children, it is unlikely that a single biomarker will definitively identify those children with true appendicitis from those with alternate causes of intra-abdominal inflammation (mesenteric adenitis, viral gastroenteritis,

inflammatory bowel disease, etc.); it is more likely that a *combination* of protein mediators will separate different etiologies, using multiple data elements similar to an inflammatory "fingerprint."

In this study, we demonstrate the potential of precision medicine to describe the inflammatory landscape present in children with appendicitis. Our primary objective was to compare levels of individual inflammatory protein mediators previously associated with intra-abdominal inflammation (CRP (Kharbanda et al., 2011; Anielski et al., 2010; Groselj-Grenc et al., 2007; Sack et al., 2006; Yildirim et al., 2006; Acharya et al., 2017; Allister et al., 2011; Paajanen et al., 2002; Dalal et al., 2005; Benito et al., 2016; Gavela et al., 2012; Kwan and Nager 2010), PCT (Benito et al., 2016; Gavela et al., 2012; Kwan and Nager 2010; Kafetzis et al., 2005; Chakhunashvili et al., 2005; Yu et al., 2013; Blab et al., 2004), interleukin-6 (IL-6) (Zviedre et al., 2016; Kharbanda et al., 2011; Anielski et al., 2010; Groselj-Grenc et al., 2007; Sack et al., 2006; Yildirim et al., 2006), IL-8 (Zviedre et al., 2016; Kharbanda et al., 2011; Paajanen et al., 2002; Yoon et al., 2002), IL-10 (Zviedre et al., 2016; Yildirim et al., 2006; Hachim and Ahmed 2006), and monocyte chemoattractant protein-1 (MCP-1) (Zviedre et al., 2016; Andersson et al., 2014), SAA (Andersson et al., 2014; Sit et al., 2014; Schellekens et al., 2013; Lycopoulou et al., 2005; Abbas et al., 2014)) in a cohort of children with suspected appendicitis. Furthermore, using a targeted multiplex proteomics approach, we examined an expansive array of novel candidate cytokine and chemokines within this population. Using suspected appendicitis as a high-volume, high-stakes model, we aim to show how precision medicine profiling could be applied across a number of pediatric emergency medicine (PEM) presentations to shape the next generation of PEM research initiatives.

6.3. Materials and Methods

6.3.1. Study Design

We completed an observational multicohort study. Data for the current analysis are a subset of the data collected as part of a series of studies through the Alberta Sepsis Network assessing inflammatory protein mediators in children with infection. These studies were approved by the Health Research Ethics Board of the University of Alberta and the Conjoint Health Research Ethics Board of the University of Calgary (REB13-0586; REB15-1045; Pro00008797). Informed consent or assent was obtained from the children and/or their caregivers. In those circumstances where ongoing resuscitative measures were underway, delayed consent was obtained at the earliest possible opportunity.

6.3.2. Study setting and population

Study participants were enrolled at the Alberta Children's Hospital, the tertiary pediatric health centre serving southern Alberta, eastern British Columbia, and western Saskatchewan (catchment 1.8 million).

Between 2009 and 2015, 3 cohorts of children were prospectively enrolled according to the inclusion and exclusion criteria described below.

1) Suspected appendicitis: children presenting to the ED with abdominal pain in whom the managing physician suspected a diagnosis of appendicitis defined by either (a) performance of an ultrasound (US) evaluation of the appendix or (b) consultation with the pediatric surgical team for suspected appendicitis. Children were excluded if they had previous appendectomy, required active resuscitation in the ED, were discharged directly to the PICU,

were pregnant, had abdominal pain for more than 5 days, had a history of illness resulting in immune suppression, were previously enrolled in the study, or had an imaging study of the appendix performed at an external healthcare centre (Shommu et al., 2018a; Shommu et al., 2018b).

2) Sepsis: children presenting to the ED with sepsis, defined as SIRS caused by a suspected or proven bacterial or fungal infection and who had antibiotic/antifungal medications and blood culture ordered, but did not require PICU care (Mickiewicz et al., 2015a).
3) Severe sepsis: children admitted to the pediatric intensive care unit (PICU) for sepsis as evidenced by the systemic inflammatory response syndrome (SIRS) caused by a suspected or proven bacterial or fungal infection orders for antibiotics/antifungal medication and an arterial and/or central venous line were required. Children were excluded if they were not expected to survive ≥24 hours, had palliative goals of care (no intubation or vasoactive infusions), or had severe sepsis for ≥48 hours (defined as sepsis with cardiovascular dysfunction, acute respiratory distress syndrome, or two other organ dysfunctions) (Mickiewicz et al., 2015a)

For the purposes of the current analysis, we limited inclusion to those children aged 5 through 17 years, as these are the most typical ages of appendicitis presentations. We excluded children from these studies that were enrolled external to the Alberta Children's Hospital. Those children that were co-enrolled in more than one of the cohorts were analyzed once.

6.3.3. Care of the patient

Clinical care of the child was left to the discretion of the managing ED and surgical teams according to our local appendicitis and/or sepsis pathways.

6.3.4 Inflammatory protein mediator profiling

The laboratory technicians performing the analyses were blinded to patient allocation. Inflammatory mediators were measured according to standardized, validated processes as previously described (Shommu et al., 2018a; Shommu et al., 2018b; Roberts et al., 2013). Three human cytokine and chemokine assay kits (Bio-Plex Pro Human Cytokine 21-Plex Assay, Bio-Plex Pro Human Cytokine 27-Plex Assay, Bio-Plex Pro Human Acute Phase 5- + 4-Plex Panel Complete Kit), obtained from Bio-Rad Laboratories Inc. (Hercules, CA, USA), were used to detect 57 inflammatory mediators in the plasma samples (Supplementary Table 6-1). Assays were run according to manufacturer-provided protocols. Briefly, analyte capture beads were sonicated to disrupt and eliminate any aggregates. Once the beads had been sonicated, 50 μL of the bead mix was aliquoted into each assay well of a 96-well microtiter plate. Beads were washed twice with supplied wash buffer using a magnetic plate base. Standards, controls, or samples were added to each well and incubated for 30 min at room temperature on a horizontal shaking platform. Bead-analyte complexes were washed three times followed by the addition of 25 μL of biotinylated detection antibody to each well. Plates were incubated while shaking for 30 min at room temperature. Bead-analyte complexes washed three times and 50 μL of phycoerythrin- (PE-) conjugated streptavidin was added to each well. Plates were incubated while shaking for 10 min at room temperature and washed three times, after which beads were resuspended in 125 mL of assay buffer and analyzed on a Luminex 200 apparatus (Applied Cytometry Systems, Sheffield, UK). Acquisition and analysis of the samples were driven by Bio-Plex Manager 6.0 software (Bio-Rad Laboratories Inc.). If the coefficient of variance between two replicates was greater than 20%, the data was considered as a missing value. Individual data points that were below the limit of detection were set to the lowest observed value for that

protein. Similarly, individual data points that were above the limit of detection were set to the highest observed value for that protein. Some mediators measured were consistently outside of the dynamic range of the assay. Any mediator with >40% of samples analyzed falling outside of the detectable limits were excluded from the overall analysis.

6.3.5 Measures

The primary outcome evaluated was the presence of appendicitis (yes/no), defined as (a) the presence of an inflamed appendix on pathological evaluation or (b) management of appendicitis via percutaneous drain + antimicrobials for an appendiceal abscess. Those children who were evaluated for suspected appendicitis but did not undergo surgical intervention were considered not to have appendicitis provided they did not return for appendectomy within 2 weeks of the index ED presentation. At the time of the studies, nonoperative management of appendicitis was not offered at the study site.

Secondary outcomes included severity of disease. Simple appendicitis included those children with a diagnosis of appendicitis (as above) without evidence of perforation. Perforated appendicitis was identified by evidence of inflamed appendix along with any presence of perforation on pathologic examination or percutaneous management for an appendiceal abscess. Septic appendicitis was defined as those children with appendicitis admitted to the PICU specifically as a result of the infectious process.

6.3.6 Statistical analyses

All statistical analyses were conducted in R v.3.4.2 (R Core Team, 2017), and a P-value <0.05 was considered statistically significant.

6.3.6.1 Regression models

Two different groupings of patients were used in the analysis of seven biomarkers chosen a priori, to compare biomarker concentrations between patients with and without appendicitis, and to compare between patients of different severities of appendicitis. For the first analysis, all patients with appendicitis, perforated appendix, and appendicitis causing sepsis were grouped together into the "Appendicitis" category and compared against patients without appendicitis and those with non-appendicitis sepsis. For the second analysis, patients were separated into four categories: patients without appendicitis, patients with appendicitis, patients with severe appendicitis (perforated appendix and appendicitis causing sepsis grouped together due to sample size in each group), and patients with non-appendicitis sepsis.

Biomarker concentrations were transformed using the natural logarithm prior to analysis to normalize the data, and differences in mean log concentrations of the seven biomarkers chosen a priori were assessed using multivariate normal regression models. Multivariate regression models take into account within-patient correlations in the concentrations of different biomarkers and adjust the coefficient and standard error estimates to reflect that. In addition to patient category, age and sex of the patient were also included in the model, to adjust for any age or sex-specific differences in concentrations and provide population-averaged estimates. The same type of model was used for the appendicitis versus no-appendicitis analysis as with the appendicitis severity analysis. After fitting the models, pairwise comparisons between patient categories (averaged over age and sex) were conducted using least-square means predictions and the Bonferroni correction for multiple comparisons through the R package "emmeans" (Lenth 2018).

6.3.6.2 Decision tree

Prior to analysis using a decision tree model, proportions of missing observations were determined for all biomarkers, and any biomarker with 15% or more missing values were excluded from the analysis. For all the included biomarkers, missing values were replaced with the mean concentration across all patient categories to minimize bias. The patient categories used for these predictions were the same as the appendicitis severity regression models.

The decision tree was fit using the R package "rpart" (Therneau 2018) on the entire dataset at once as opposed to having separate training and validation datasets. An initial tree using all biomarkers was fit, after which a plot of the error rate as a function of the complexity parameter was created. The complexity parameter represents the number of splits in the decision tree and reflects the number of biomarkers used in the prediction of the outcome. Based on the results of this plot, the complexity parameter at which error rate stops decreasing was identified, and the tree was pruned to this level by removing uninformative biomarkers. The resulting decision tree includes only biomarkers which were identified to be informative and most important for reducing misclassification errors.

6.4. Results

There were 317 potentially eligible participants in the parent cohorts, of which 83 were excluded. A further 49 subjects did not have adequate sample for testing, leaving a final population of 185 children (Figure 6-1). Overall, 79 children were included in the appendicitis cohort, 83 in the non-appendicitis cohort, and 23 in the non-appendicitis sepsis cohort. Patient characteristics are outlined in Table 6-1. Of the 57 inflammatory protein mediators evaluated, 3

(SCGF-B, fibrinogen, haptoglobin) consistently demonstrated values outside of reference range and were excluded from further analysis.

6.4.1. Appendicitis vs non-appendicitis abdominal pain vs non-appendicitis sepsis

Plasma levels of IL-6, CRP, MCP-1, PCT, and SAA were significantly different ($p < 0.001$) in children with appendicitis compared to those with non-appendicitis abdominal pain. Similarly, significant differences in IL-6, IL-8, and PCT were demonstrated when comparing children with appendicitis and those with non-appendicitis sepsis. IL-6 and PCT were the only markers demonstrating significant differences between all 3 distinct populations (Table 6-2, Figure 6-2). Pairwise Pearson's correlation coefficients and multivariate regression model coefficients are listed in Appendices 6-2 and 6-3.

6.4.2. Appendicitis severity assessment

IL-6, IL-8, and PCT demonstrated the most significant difference ($P < 0.001$) between simple and complex appendicitis. SAA did not demonstrate significant differences across severity of illness (Table 6-2, Figure 6-3). Pairwise Pearson's correlation coefficients and multivariate regression model coefficients are listed in Appendices 6-4 and 6-5.

6.4.3. Expansive inflammatory-related protein mediator profiling (heat mapping)

Levels of acute phase reactants (7), chemokines (11), regulatory growth hormones (15), inflammatory (17), and anti-inflammatory (4) mediators were determined (mean and standard deviation for all 54 proteins found in Appendix 6-6) and compared across severity of illness;

upon visual inspection, six main response patterns were observed (Figure 6-4). Appendix 6-1 describes each of these inflammatory proteins.

1) Progressively higher plasma levels with increasing severity of illness: ferritin, G-CSF, IL-6, IL-15, MIP-1β, IL-18, MCP-1, and PCT
2) Progressive suppression of plasma levels with increasing illness severity: RANTES, IL-1α, TNF-β, and TRAIL
3) Multiphasic pattern, with initial suppression followed by progressive elevation: CTACK, MIG, IL-8, IL-2Rα, IFN-α2, M-CSF, VEGF, and IP-10
4) Multiphasic pattern, with initial elevation followed by progressive suppression: RANTES, SAA, and PDGF-BB
5) Clear distinction between appendicitis (simple, perforated) and sepsis (regardless of underlying condition): PCT, CTACK, GROα, G-CSF, IL-1α, IL-6, TRAIL, and INF-α2
6) Clear distinction in all appendicitis when compared to non-appendicitis sepsis: IL-5, MIP-1α, IL-3, IL-6, MCP-3, and IL-7

6.4.4 Protein mediator decision tree analysis

54 protein mediators were evaluated for decision tree analysis. The final decision tree was composed of 6 biomarkers including CRP, ferritin, SAA, RANTES, MIG, and PCT (Figure 6-5). Operating characteristics demonstrate high levels of specificity and negative predictive values for each patient category (Table 6-3).

6.4.5 Assessment of current common appendicitis evaluations

Characteristics of tests commonly used in the evaluation of pediatric appendicitis (WBC, neutrophils, and PAS) are shown in Appendix 6-7.

6.5. Discussion

Current advances in, and the availability of, precision medicine technologies offer the potential to transform our understanding of the underlying physiologic responses of children with abdominal pain, providing a gateway to novel diagnostic and risk-stratification strategies. In this study, we have described the inflammatory landscape of children presenting to the ED with suspected appendicitis. Specifically, using a conventional approach based on evaluating individual bio-markers, we have demonstrated statistically significant differences in 5 previously described plasma markers (IL-6, CRP, MCP-1, PCT, and SAA) in children with appendicitis compared to those with NAAP. Furthermore, the results of our decision-tree analysis and expansive proteomic heat mapping approaches determined several potential future biomarker candidates that had not previously been identified. The identification of these inflammatory protein mediators could reveal a "fingerprint" for appendicitis and, together with future bio-technology partnership, result in a point-of-care clinical tool for timely and accurate diagnosis which is readily available/accessible across the spectrum of health care settings.

Despite being the most common atraumatic surgical emergency in the pediatric population, appendicitis continues to challenge clinicians managing children with abdominal pain, from primary care providers in rural settings to tertiary pediatric emergency subspecialists. Current

diagnostic and risk stratification strategies (Thompson et al., 2015) include combinations of common clinical scoring systems (Pediatric Appendicitis Score (Samuel 2002), Alvarado Score (Alvarado 1986), Lintula Score (Lintula et al., 2009), Appy1 (Huckins et al., 2017), etc.), laboratory investigations (WBC, neutrophils +/- CPR) (Bundy et al., 2007), and imaging studies (ultrasound) (Gongidi and Bellah 2017). While often useful in the first-line work-up of a child with suspected appendicitis, these strategies remain limited because (a) children often present with atypical symptoms (Becker et al., 2007; Mallick 2008), (b) clinical scoring systems and conventional laboratory investigations have suboptimal test characteristics (sensitivities and specificities in the 70-85% range) (Bundy et al., 2007; Khanafer et al., 2016; Thompson 2012; Kulik et al., 2013), and (c) ultrasound is known to have high rates of incomplete visualization (Ross et al., 2014; Keller et al., 2017; Ramarajan et al., 2014; Mittal et al., 2013), may be painful for the child (compression on an already tender abdomen), and in females requires a full bladder (for optimal evaluation of pelvic organs) (Ross et al., 2016). Second-line/advanced imaging techniques are limited by exposure of developing organs to unacceptable levels of ionizing radiation (computed tomography) (Brenner et al., 2001) or have limited accessibility outside tertiary settings (magnetic resonance imaging). Sadly, a missed or misdiagnosis of appendicitis (Naiditch et al., 2013; Galai et al., 2017; Lee et al., 2012) can lead to adverse patient outcomes; it remains amongst the highest concerns of caregivers/parents, ranks amongst the highest pediatric emergency presentations leading to litigation (Raine 2011; Brown et al., 2010), and can result in unnecessary exposure to surgical and anesthetic intervention (negative appendectomies).

In evaluating a set of 7 previously identified inflammatory protein markers, our results demonstrated statistically significant differences in IL-6, CRP, MCP-1, PCT, and SAA. These results are consistent with prior studies (Zviedre et al., 2016; Kharbanda et al., 2011; Anielski et

al., 2010; Andersson et al., 2014; Sit et al., 2014; Acharya et al., 2017; Kharbanda et al., 2012). Despite these significant differences, no single mediator demonstrates satisfactory sensitivity or specificity for the clinical identification of appendicitis. This has been a consistent problem in the application of biomarker analysis to complex disease. More recent studies have demonstrated the value in multiplex analysis of biomarkers, assessing an overall disease or inflammatory "fingerprint" over the measurement of a single mediator (Shommu et al., 2018a; Shommu et al., 2018b; Mickiewicz et al., 2015a). Assessment of multiple biomarkers is better able to filter out the noise originating from multiple disease etiologies (many mediators are specific to the initiating source of inflammation, independent of the actual clinical course of disease) allowing studies to focus on the handful of mediators responsible for the actual disease/tissue pathology. Importantly, inflammatory mediators do not work in isolation but rather function as an overall milieu. For example, a patient with an increase in one specific proinflammatory cytokine may have similar disease progression as a second patient who has normal levels of the proinflammatory mediator but reduced levels of an anti-inflammatory cytokine. By measuring single mediators, these overall "fingerprints" or bioprofiles are missed. Additionally, analysis of single mediators is inherently biased, focusing on the most obvious or logical targets and ignoring other mediators that may be much more biologically important but are simply not commonly studied or on the surface have no apparent mechanistic link to the condition being studied. Through the use of a large, multiplexed panel, one is better able to identify these nonobvious targets that may in fact be more informative than the common a priori selected mediators. Finally, the surge or suppression of individual biomarkers may have a temporal relationship to disease evolution that would allow a multimarker panel to identify changes across time.

The principle drawback associated with the analysis of a large array of biomediators is the need to separate the key core markers that predict disease or progression from the noise of the other background mediators. To accomplish this, we applied a machine learning approach to the overall data set to generate a decision tree. The purpose of the decision tree was to identify important markers and have a way to graphically display between-group differences rather than to develop an accurate predictive tool. The decision tree analysis identified 6 key biomarkers, as using any additional biomarkers would not have resulted in an appreciable increase in classification accuracy. Only two out of the 6 biomarkers were included in the more in-depth analysis (CRP and PCT) while the role of the other 4 (Ferritin, SAA, RANTES, and MIG) in the progression of appendicitis may be more unclear. Despite achieving the best possible fit for this current dataset, the test characteristics of the decision tree preclude its use in clinical settings; in particular, misclassification of patients with non-appendicitis sepsis and severe appendicitis as healthy would have the most severe consequences. Although this specific decision tree would require revision and validation before implementation into clinical practice as a decision tool, it can serve as a hypothesis-generating mechanism for investigating the roles of different biomarkers in pathogenesis, graphically display possible classification schemes, and demonstrate the potential importance of machine learning methods which considers multiple markers and features for disease diagnosis.

Machine learning (ML) has already successfully been applied to identify patients with chronic and long-term conditions like cancer (Kourou et al., 2015) or in detecting sepsis based on continuous heart rate and blood pressure monitoring in critical care patients (Shashikumar et al., 2017; Mao et al., 2018). In these cases, however, data is often available for long periods of time (cancer) or is continuously collected (critical care patients), whereas in emergency medicine

collecting high volumes of sequential data is not practical or possible in short periods of time. Success with using ML methods in identifying patients with preclinical Alzheimer's and clinical cases of Alzheimer's (Casanova et al., 2018a; Casanova et al., 2018b; Casanova et al., 2016; Laske et al., 2011) based on blood metabolite and protein panels could suggest the use of similar panel-based decision tools in emergency medicine. Indeed, ML methodologies are making their way into the ED (Stewart et al., 2018); they have been shown to be as accurate—or better than—certain clinical outcome prediction models for ED triage (Levin et al., 2018), predicting adverse cardiovascular outcomes (VanHouten et al., 2014; Liu et al., 2016), and identifying traumatic injury requiring life-saving intervention (Liu and Salinas 2017; Sefrioui et al., 2017).

Historically, it has not been practical to conduct such a broad spectrum biomediator analysis. Measurement and characterization of more than 50 mediators in a clinical study using single analyte assessments is not feasible due to cost, time, and sample volume requirements. However, the application of multiplexed approaches, such as we have demonstrated in the current study, offers significant advantages. It is possible to generate data on scores of mediators in a cost- and time-efficient manner with very reasonable sample volumes. For the current study, less than 200 μL of blood from each patient was needed; this research strategy provides significant advantages in studying even the youngest neonates in the future. Although the current approach examines a very broad array of inflammatory mediators, many of which appear to be noninformative for this study, in future studies, similar multiplex technology can be developed to focus only on the biomarkers found to identify appendicitis and stratify disease severity. This approach would allow transition from a discovery-based overall assessment of the inflammatory "landscape" to a focused, bedside diagnostic study that can accurately and specifically identify patients that would best benefit from a given treatment. Importantly, the current assessment of

protein mediators can be partnered in future studies with additional panels of biomarkers (damage-associated molecular patterns [DAMPs], metabolites, transcriptomes, etc.). Although multiplex approaches provide better sensitivity and specificity than single analyte assessments, previous work has demonstrated that the integration of multiple biomarker platforms has the capacity to further enhance the accuracy of these diagnostic tests (Mickiewicz et al., 2015b).

6.6. Limitations

One significant limitation of the current study is the turnaround time required between sample collection and determination of biomarker levels. Much of this limitation is related to the assessment of a broad array of biomarkers requiring sequential analysis of multiple assay plates. Future studies aimed at the assessment of a narrow range of useful markers will greatly streamline the assessment, allowing results to be available to the clinician within one to two hours. Further refinement and engagement with industry can enable the development of bedside, dipstick-based tests that could provide results within minutes using a single drop of blood. These features are highly relevant in busy, high-volume, high-stakes ED environments. Not only does such an approach have substantial appeal with respect to diagnostic turnaround time, but simple bedside tests do not require specialized equipment or staff expertise allowing them to be used in the smallest and most remote health care settings. Diagnosis of patients within a rural or remote setting can rapidly inform a local health care professional if transfer of the patient to the nearest surgical centre is required, reducing delays, lowering the risk of severe complication, and improving overall patient outcomes.

The dataset in our current study was limited in power for training a complex ML classification algorithm with 4 different outcome levels. In developing this kind of ML decision tool, it is important to have a diverse set of data for training models; Casanova et al. (2016) have demonstrated how limited datasets can impact repeatability and reliability, which is vital in the high-stakes ED environment. External data for validating model performance is required.

6.7. Conclusions

While assessment of previously identified inflammatory plasma mediators demonstrate statistical differences in children with appendicitis when compared to those with non-appendicitis abdominal pain, analysis of individual mediators does not have sufficiently acceptable test characteristics to be used to rule in or out appendicitis. However, a precision medicine multiplex approach to evaluating the inflammatory protein mediator landscape identifies novel patterns of candidate biomarkers that could be used to identify a fingerprint of disease. Together with industry partners, point-of-care diagnostic technologies could be developed. This discovery-to-translation approach can be used across multiple acute pediatric presentations and can be modelled for future research initiatives in pediatric emergency medicine.

6.8. References

Abbas, M.H., M.N. Choudhry, N. Hamza, B. Ali, A.A. Amin, and B.J. Ammori. 2014. Admission levels of Serum Amyloid A and Procalcitonin are more predictive of the diagnosis of acute appendicitis compared with C-reactive protein. Surg. Laparosc. Endosc. Percutaneous Tech. 24:488–494. doi: 10.1097/SLE.0000000000000067.

Acharya, A., S.R. Markar, M. Ni, and G.B. Hanna. 2017. Biomarkers of acute appendicitis: systematic review and cost-benefit trade-off analysis. Surg. Endosc. other Interv. Tech. 31:1022–1031. doi:10.1007/s00464-016-5109-1.

Allister, L., R. Bachur, J. Glickman, and B. Horwitz. 2011. Serum Markers in Acute Appendicitis. J. Surg. Res. 168:70–75. doi:10.1016/j.jss.2009.10.029.

Alvarado, A. 1986. A practical score for the early diagnosis of acute appendicitis. Pediatr. Emerg. Care. 2: 206-207. doi: 10.1016/s0196-0644(86)80993-3.

Andersson, M.N., and R.E. Andersson. 2011. Causes of Short-Term Mortality After Appendectomy A Population-Based Case-Controlled Study. Ann. Surg. 254:103–107. doi:10.1097/SLA.0b013e31821ad9c4.

Andersson, M., M. Ruber, C. Ekerfelt, H.B. Hallgren, G. Olaison, and R.E. Andersson. 2014. Can new inflammatory markers improve the diagnosis of acute appendicitis? World J. Surg. 38:2777–2783. doi:10.1007/s00268-014-2708-7.

Andersson, R.E. 2013. Short and long-term mortality after appendectomy in Sweden 1987 to 2006. Influence of appendectomy diagnosis, sex, age, co-morbidity, surgical method, hospital volume, and time period. A National Population-Based Cohort Study. World J. Surg. 37:974–981. doi:10.1007/s00268-012-1856-x.

Anielski, R., B. Kusnierz-Cabala, and K. Szafraniec. 2010. An evaluation of the utility of additional tests in the preoperative diagnostics of acute appendicitis. Langenbecks Arch. Surg. 395:1061–1068. doi:10.1007/s00423-009-0565-x.

Becker, T., A. Kharbanda, and R. Bachur. 2007. Atypical clinical features of pediatric appendicitis. Acad. Emerg. Med. 14:124–129. doi:10.1197/j.aem.2006.08.009.

Benito, J., Y. Acedo, L. Medrano, E. Barcena, R. Perez Garay, and E. Arana Arri. 2016. Usefulness of new and traditional serum biomarkers in children with suspected appendicitis. Am. J. Emerg. Med. 34:871–876. doi:10.1016/j.ajem.2016.02.011.

Blab, E., U. Kohlhuber, S. Tillawi, M. Schweitzer, G. Stangl, E. Ogris, and A. Rokitansky. 2004. Advancements in the diagnosis of acute appendicitis in children and adolescents. Eur. J. Pediatr. Surg. 14:404–409. doi:10.1055/s-2004-821152.

Blomqvist, P.G., R.E.B. Andersson, F. Granath, M.P. Lambe, and A.R. Ekbom. 2001. Mortality after appendectomy in Sweden, 1987-1996. Ann. Surg. 233:455–460. doi:10.1097/00000658-200104000-00001.

Brenner, D.J., C.D. Elliston, E.J. Hall, and W.E. Berdon. 2001. Estimated risks of radiation-induced fatal cancer from pediatric CT. Am. J. Roentgenol. 176:289–296. doi:10.2214/ajr.176.2.1760289.

Brown, T.W., M.L. McCarthy, G.D. Kelen, and F. Levy. 2010. An epidemiologic study of closed emergency department malpractice claims in a national database of physician malpractice insurers. Acad. Emerg. Med. 17:553–560. doi:10.1111/j.1553-2712.2010.00729.x.

Bundy, D.G., J.S. Byerley, E.A. Liles, E.M. Perrin, J. Katznelson, and H.E. Rice. 2007. Does this child have appendicitis? J. Am. Med. Assoc. 298:438–451. doi:10.1001/jama.298.4.438.

Casanova, R., R.T. Barnard, S.A. Gaussoin, S. Saldana, K.M. Hayden, J.E. Manson, R.B. Wallace, S.R. Rapp, S.M. Resnick, M.A. Espeland, J.-C. Chen, and W.-M.S.G. Alzheimer's. 2018a. Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases. Neuroimage 183:401–411. doi:10.1016/j.neuroimage.2018.08.040.

Casanova, R., S. Saldana, M.W. Lutz, B.L. Plassman, M. Kuchibhatla, and K.M. Hayden. 2018b. Investigating predictors of cognitive decline using machine learning. Journals Gerontol. Ser. B - Psychol. Sci. Soc. Sci. 75:733–742. doi:10.1093/geronb/gby054.

Casanova, R., S. Varma, B. Simpson, M. Kim, Y. An, S. Saldana, C. Riveros, P. Moscato, M. Griswold, D. Sonntag, J. Wahrheit, K. Klavins, P. V Jonsson, G. Eiriksdottir, T. Aspelund, L.J. Launer, V. Gudnason, C.L. Quigley, and M. Thambisetty. 2016. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. ALZHEIMERS \& Dement. 12:815–822. doi:10.1016/j.jalz.2015.12.008.

Chakhunashvili, L., A. Inasaridze, S. Svanidze, J. Samkharadze, and I. Chkhaidze. 2005. Procalcitonin as the biomarker of inflammation in diagnostics of pediatric appendicular peritonitis and for the prognosis of early postoperative complications. Georgian Med. News 78–81.

Dalal, I., E. Somekh, A. Bilker-Reich, M. Boaz, A. Gorenstein, and F. Serour. 2005. Serum and peritoneal inflammatory mediators in children with suspected acute appendicitis. Arch. Surg. 140:169–173. doi:10.1001/archsurg.140.2.169.

Galai, T., O.Z. Beloosesky, D. Scolnik, A. Rimon, and M. Glatstein. 2017. Misdiagnosis of acute appendicitis in children attending the emergency department: the experience of a large,

tertiary care pediatric hospital. Eur. J. Pediatr. Surg. 27:138–141. doi:10.1055/s-0035-1570757.

Gavela, T., B. Cabeza, A. Serrano, and J. Casado-Flores. 2012. C-reactive protein and procalcitonin are predictors of the severity of acute appendicitis in children. Pediatr. Emerg. Care 28:416–419. doi:10.1097/PEC.0b013e318252d875.

Gongidi, P., and R.D. Bellah. 2017. Ultrasound of the pediatric appendix. Pediatr. Radiol. 47:1091–1100. doi:10.1007/s00247-017-3928-4.

Groselj-Grenc, M., S. Repse, D. Vidmar, and M. Derganc. 2007. Clinical and laboratory methods in diagnosis of acute appendicitis in children. Croat. Med. J. 48:353–361.

Hachim, M.Y., and A.-R.H. Ahmed. 2006. The role of the cytokines and cell-adhesion molecules on the immunopathology of acute appendicitis. Saudi Med. J. 27:1815–1821.

Huckins, D.S., K. Copeland, W. Self, C. Vance, P. Hendry, K. Borg, and J. Gogain. 2017. Diagnostic performance of a biomarker panel as a negative predictor for acute appendicitis in adult ED patients with abdominal pain. Am. J. Emerg. Med. 35:418–424. doi:10.1016/j.ajem.2016.11.027.

Kafetzis, D.A., I.M. Velissariou, P. Nikolaides, M. Sklavos, M. Maktabi, G. Spyridis, D.D. Kafetzis, and E. Androulakakis. 2005. Procalcitonin as a predictor of severe appendicitis in children. Eur. J. Clin. Microbiol. Infect. Dis. 24:484–487. doi:10.1007/s10096-005-1360-4.

Keller, C., N.E. Wang, D.L. Imler, S.S. Vasanawala, M. Bruzoni, and J. V Quinn. 2017. Predictors of nondiagnostic ultrasound for appendicitis. J. Emerg. Med. 52:318–323. doi:10.1016/j.jemermed.2016.07.101.

Khanafer, I., D.-A. Martin, T.P. Mitra, R. Eccles, M.E. Brindle, A. Nettel-Aguirre, and G.C. Thompson. 2016. Test characteristics of common appendicitis scores with and without

laboratory investigations: a prospective observational study. BMC Pediatr. 16. doi:10.1186/s12887-016-0687-6.

Kharbanda, A.B., Y. Cosme, K. Liu, S.L. Spitalnik, and P.S. Dayan. 2011. Discriminative accuracy of novel and traditional biomarkers in children with suspected appendicitis adjusted for duration of abdominal pain. Acad. Emerg. Med. 18:568–574. doi:10.1111/j.1553-2712.2011.01095.x.

Kharbanda, A.B., A.J. Rai, Y. Cosme, K. Liu, and P.S. Dayan. 2012. Novel serum and urine markers for pediatric appendicitis. Acad. Emerg. Med. 19:56–62. doi:10.1111/j.1553-2712.2011.01251.x.

Kourou, K., T.P. Exarchos, K.P. Exarchos, M. V Karamouzis, and D.I. Fotiadis. 2015. Machine learning applications in cancer prognosis and prediction. Comput. Struct. Biotechnol. J. 13:8–17. doi:10.1016/j.csbj.2014.11.005.

Kulik, D.M., E.M. Uleryk, and J.L. Maguire. 2013. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. J. Clin. Epidemiol. 66:95–104. doi:10.1016/j.jclinepi.2012.09.004.

Kwan, K.Y., and A.L. Nager. 2010. Diagnosing pediatric appendicitis: usefulness of laboratory markers. Am. J. Emerg. Med. 28:1009–1015. doi:10.1016/j.ajem.2009.06.004.

Laske, C., T. Leyhe, E. Stransky, N. Hoffmann, A.J. Fallgatter, and J. Dietzsch. 2011. Identification of a blood-based biomarker panel for classification of Alzheimer's disease. Int. J. Neuropsychopharmacol. 14:1147–1155. doi:10.1017/S1461145711000459.

Lee, J., D.B. Tashjian, and K.P. Moriarty. 2012. Missed opportunities in the treatment of pediatric appendicitis. Pediatr. Surg. Int. 28:697–701. doi:10.1007/s00383-012-3108-4.

Lenth, R. 2018. emmeans: estimated marginal means, aka least-squares means. R package version 1.54. URL: https://CRAN.R-project.org/package=emmeans

Levin, S., M. Toerper, E. Hamrock, J.S. Hinson, S. Barnes, H. Gardner, A. Dugas, B. Linton, T. Kirsch, and G. Kelen. 2018. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann. Emerg. Med. 71:565–574. doi:10.1016/j.annemergmed.2017.08.005.

Lintula, H., H. Kokki, R. Kettunen, and M. Eskelinen. 2009. Appendicitis score for children with suspected appendicitis. A randomized clinical trial. Langenbecks Arch. Surg. 394:999–1004. doi:10.1007/s00423-008-0425-0.

Liu, N.T., and J. Salinas. 2017. Machine learning for predicting outcomes in trauma. Shock. 48:504–510. doi:10.1097/SHK.0000000000000898.

Liu, Y., B.M. Scirica, C.M. Stultz, and J. V Guttag. 2016. Beatquency domain and machine learning improve prediction of cardiovascular death after acute coronary syndrome. Sci. Rep. 6. doi:10.1038/srep34540.

Lycopoulou, L., C. Mamoulakis, E. Hantzi, D. Demetriadis, S. Antypas, M. Giannaki, C. Bakoula, G. Chrousos, and I. Papassotiriou. 2005. Serum amyloid A protein levels as a possible aid in the diagnosis of acute appendicitis in children. Clin. Chem. Lab. Med. 43:49–53. doi:10.1515/CCLM.2005.007.

Mallick, M.S. 2008. Appendicitis in pre-school children: a continuing clinical challenge. A retrospective study. Int. J. Surg. 6:371–373. doi:10.1016/j.ijsu.2008.06.003.

Mao, Q., M. Jay, J.L. Hoffman, J. Calvert, C. Barton, D. Shimabukuro, L. Shieh, U. Chettipally, G. Fletcher, Y. Kerem, Y. Zhou, and R. Das. 2018. Multicentre validation of a sepsis

prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 8. doi:10.1136/bmjopen-2017-017833.

Mickiewicz, B., G.C. Thompson, J. Blackwood, C.N. Jenne, B.W. Winston, H.J. Vogel, A.R. Joffe, and A.S. Network. 2015a. Development of metabolic and inflammatory mediator biomarker phenotyping for early diagnosis and triage of pediatric sepsis. Crit. Care 19. doi:10.1186/s13054-015-1026-2.f

Mickiewicz, B., P. Tam, C.N. Jenne, C. Leger, J. Wong, B.W. Winston, C. Doig, P. Kubes, H.J. Vogel, and A.S. Network. 2015b. Integration of metabolic and inflammatory mediator profiles as a potential prognostic approach for septic shock in the intensive care unit. Crit. Care 19. doi:10.1186/s13054-014-0729-0.

Mittal, M.K., P.S. Dayan, C.G. Macias, R.G. Bachur, J. Bennett, N.C. Dudley, L. Bajaj, K. Sinclair, M.D. Stevenson, A.B. Kharbanda, and P.E.M. Collaborative. 2013. Performance of ultrasound in the diagnosis of appendicitis in children in a multicenter cohort. Acad. Emerg. Med. 20:697–702. doi:10.1111/acem.12161.

Naiditch, J.A., T.B. Lautz, S. Daley, M.C. Pierce, and M. Reynolds. 2013. The implications of missed opportunities to diagnose appendicitis in children. Acad. Emerg. Med. 20:592–596. doi:10.1111/acem.12144.

Paajanen, H., A. Mansikka, M. Laato, R. Ristamaki, K. Pulkki, and S. Kostiainen. 2002. Novel serum inflammatory markers in acute appendicitis. Scand. J. Clin. Lab. Invest. 62:579–584. doi:10.1080/003655102764654312.

Raine, J.E. 2011. An analysis of successful litigation claims in children in England. Arch. Dis. Child. 96:838–840. doi:10.1136/adc.2011.212555.

Ramarajan, N., R. Krishnamoorthi, L. Gharahbaghian, E. Pirrotta, R.A. Barth, and N.E. Wang. 2014. Clinical correlation needed: What do emergency physicians do after an equivocal ultrasound for pediatric acute appendicitis? J. Clin. Ultrasound 42:385–394. doi:10.1002/jcu.22153.

Roberts, D.J., C.N. Jenne, C.G. Ball, C. Tiruta, C. Leger, Z. Xiao, P.D. Faris, P.B. McBeth, C.J. Doig, C.R. Skinner, S.G. Ruddell, P. Kubes, and A.W. Kirkpatrick. 2013. Efficacy and safety of active negative pressure peritoneal therapy for reducing the systemic inflammatory response after damage control laparotomy (the Intra-peritoneal Vacuum Trial): study protocol for a randomized controlled trial. Trials 14. doi:10.1186/1745-6215-14-141.

Ross, M.J., H. Liu, S.J. Netherton, R. Eccles, P.-W. Chen, G. Boag, E. Morrison, and G.C. Thompson. 2014. Outcomes of children with suspected appendicitis and incompletely visualized appendix on ultrasound. Acad. Emerg. Med. 21:538–542. doi:10.1111/acem.12377.

Ross, M., S. Selby, N. Poonai, H. Liu, S. Minoosepehr, G. Boag, R. Eccles, and G. Thompson. 2016. The effect of a full bladder on proportions of diagnostic ultrasound studies in children with suspected appendicitis. Can. J. Emerg. Med. 18:414–419. doi:10.1017/cem.2016.23.

Sack, U., B. Biereder, T. Elouahidi, K. Bauer, T. Keller, and R.-B. Trobs. 2006. Diagnostic value of blood inflammatory markers for detection of acute appendicitis in children. BMC Surg. 6:15. doi: 10.1186/1471-2482-6-15.

Samuel, M. 2002. Pediatric appendicitis score. J. Pediatr. Surg. 37:877–881. doi:10.1053/jpsu.2002.32893.

Schellekens, D.H.S.M., K.W.E. Hulsewe, B.A.C. van Acker, A.A. van Bijnen, T.M.H. de Jaegere, S.H. Sastrowijoto, W.A. Buurman, and J.P.M. Derikx. 2013. Evaluation of the

diagnostic accuracy of plasma markers for early diagnosis in patients suspected for acute appendicitis. Acad. Emerg. Med. 20:703–710. doi:10.1111/acem.12160.

Sefrioui, I., R. Amadini, J. Mauro, A. El Fallahi, and M. Gabbrielli. 2017. Survival prediction of trauma patients: a study on US National Trauma Data Bank. Eur. J. Trauma Emerg. Surg. 43:805–822. doi:10.1007/s00068-016-0757-3.

Shashikumar, S.P., M.D. Stanley, I. Sadiq, Q. Li, A. Holder, G.D. Clifford, and S. Nemati. 2017. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. J. Electrocardiol. 50:739–743. doi:10.1016/j.jelectrocard.2017.08.013.

Shommu, N.S., C.N. Jenne, J. Blackwood, A.R. Joffe, D.-A. Martin, G.C. Thompson, and H.J. Vogel. 2018. Metabolomic and inflammatory mediator based biomarker profiling as a potential novel method to aid pediatric appendicitis identification. PLoS One 13. doi:10.1371/journal.pone.0193563.

Shommu, N.S., C.N. Jenne, J. Blackwood, D.-A. Martin, A.R. Joffe, R. Eccles, M. Brindle, I. Khanafer, H.J. Vogel, and G.C. Thompson. 2018. The use of metabolomics and inflammatory mediator profiling provides a novel approach to identifying pediatric appendicitis in the emergency department. Sci. Rep. 8. doi:10.1038/s41598-018-22338-1.

Stewart, J., P. Sprivulis, and G. Dwivedi. 2018. Artificial intelligence and machine learning in emergency medicine. Emerg. Med. Australas. 30:870–874. doi:10.1111/1742-6723.13145.

Therneau, T. and B. Atkinson. 2018. rpart: recursive partitioning and regression trees. R package version 4.1-15. URL: https://CRAN.R-project.org/package=rpart.

Thompson, G.C. 2012. Clinical scoring systems in the management of suspected appendicitis in children. In Appendicitis – A Collection of Essays from Around the World. URL:

http://www.intechopen.com/articles/show/title/clinical-scoring-systems-in-the-management-of-suspected-appendicitis-in-children

Thompson, G.C., S. Schuh, J. Gravel, S. Reid, E. Fitzpatrick, T. Turner, M. Bhatt, D. Beer, G. Blair, R. Eccles, S. Jones, J. Kilgar, N. Liston, J. Martin, B. Hagel, A. Nettel-Aguirre, and P.E.R. Canada. 2015. Variation in the diagnosis and management of appendicitis at Canadian pediatric hospitals. Acad. Emerg. Med. 22:811–822. doi:10.1111/acem.12709.

VanHouten, J.P., J.M. Starmer, N.M. Lorenzi, D.J. Maron, and T.A. Lasko. 2014. Machine learning for risk prediction of acute coronary syndrome. Proc. Annu. AMIA Symp. 2014:1940–1949.

Yildirim, O., C. Solak, B. Kocer, B. Unal, M. Karabeyoglu, B. Bozkurt, S. Aksaray, and O. Cengiz. 2006. The role of serum inflammatory markers in acute appendicitis and their success in preventing negative laparotomy. J. Investig. Surg. 19:345–352. doi:10.1080/08941930600985686.

Yoon, D.Y., J. Chu, C. Chandler, S. Hiyama, J.E. Thompson, and O.J. Hines. 2002. Human cytokine levels in nonperforated versus perforated appendicitis: Molecular serum markers for extent of disease? Am. Surg. 68:1033–1037.

Yu, C.-W., L.-I. Juan, M.-H. Wu, C.-J. Shen, J.-Y. Wu, and C.-C. Lee. 2013. Systematic review and meta-analysis of the diagnostic accuracy of procalcitonin, C-reactive protein and white blood cell count for suspected acute appendicitis. Br. J. Surg. 100:322–329. doi:10.1002/bjs.9008.

Zviedre, A., A. Engelis, P. Tretjakovs, A. Jurka, I. Zile, and A. Petersons. 2016. Role of serum cytokines in acute appendicitis and acute mesenteric lymphadenitis among children. Med. Lith. 52:291–297. doi:10.1016/j.medici.2016.10.002.

Table 6-1. Characteristics of children in control, appendicitis and sepsis cohorts.

	Non-appendicitis abdominal pain (n=83)	Appendicitis (n=48)	Perforated appendicitis (n=17)	Appendicitis-related sepsis (n=14)	Non-appendicitis sepsis (n=23)
Females, n (%)	52 (62.7)	20 (41.7)	10 (58.8)	6 (42.9)	11 (47.8)
Age, mean (SD)	11.7 (3.6)	11.5 (3.2)	11.2 (3.6)	11.7 (3.4)	11.2 (4.1)
PAS, median (IQR)	5 (4)	7 (3)	8.5 (3)	N/A	N/A
Fever, n (%)	50 (60.2)	17 (35.4)	3 (17.7)	3 (21.4)	2 (8.7)
WBC, mean (SD)	10.2 (5.2)	15.6 (4.8)	17.9 (4.8)	18.1 (9.1)	15.2 (10.4)

Table 6-2. Inflammatory protein mediators and their test characteristics in children with suspected appendicitis. Area under receiver operator characteristic curve and other test characteristics calculations are based on classification of patients with non-appendicitis abdominal pain and appendicitis (non-appendicitis sepsis group is excluded).

	Non-appendicitis abdominal pain (n=83)	**Appendicitis (n = 79)**	**Non-appendicitis sepsis (n=23)**	**AUROC**	**Cut-off (%correct)**	**Sensitivity: Specificity**	**LR+:LR-**
CRP, mean (SD)[1]	7.6 (14.1)	46.6 (51.8)	70.5 (44.3)	0.8472	7.9 (78.1)	80.5:75.6	3.31:0.26
IL-6, mean (SD)[2]	28.0 (44.3)	1113.8 (6165.1)	4834.8 (20025.2)	0.8101	68.9 (69.4)	48.1:90.1	4.87:0.58
IL-8, mean (SD)[2]	24.4 (20.0)	164.5 (914.5)	385.8 (922.5)	0.5425	33.2 (59.4)	34.2:83.5	2.08:0.79
IL-10, mean (SD)[2]	19.9 (23.7)	94.9(297.0)	271.4 (864.9)	0.6294	30.1 (62.26	35.1:87.8	2.87:0.74
MCP-1, mean (SD[2]	88.3 (102.7)	361.5 (1129.2)	605.0 (1786.3)	0.6813	133.8 (68.4)	45.5:90.1	4.60:0.61
PCT, mean (SD)[3]	6.1 (2.3)	10.0 (9.2)	33.1 (52.0)	0.7382	6.5 (69.2)	57.7:80.8	3.00:0.52
SAA, mean (SD)[1]	122.7 (254.5)	248.3 (377.0)	92.5 (83.5)	0.7514	73.2 (73.6)	93.6:53.1	2.05:0.12

CRP – c-reactive protein; IL- interleukin; MCP – monocyte chemoattractant protein; PCT – procalcitonin; SAA – serum amyloid A; AUROC – area under receiver operating characteristic; LR – likelihood ratio.
[1]Concentrations measured in mg/L
[2]Concentrations measured in pg/mL
[3]Concentrations measured in ng/mL

Table 6-3. Operating characteristics (95% CI) of patient categorization based on the decision tree.

Patient category	Sensitivity	Specificity	PPV	NPV
Non-appendicitis abdominal pain	0.916 (0.834-0.965)	0.843 (0.758-0.908)	0.826 (0.733-0.897)	0.925 (0.851-0.969)
Simple appendicitis	0.646 (0.495-0.778)	0.920 (0.862-0.960)	0.738 (0.580-0.861)	0.882 (0.818-0.930)
Severe appendicitis	0.581 (0.391-0.755)	0.968 (0.926-0.989)	0.783 (0.563-0.925)	0.920 (0.867-0.957)
Non-appendicitis sepsis	0.826 (0.612-0.950)	0.945 (0.898-0.974)	0.679 (0.476-0.841)	0.975 (0.936-0.993)
Non-appendicitis abdominal pain	0.916 (0.834-0.965)	0.860 (0.776-0.921)	0.844 (0.753-0.912)	0.925 (0.851-0.969)
Appendicitis	0.722 (0.609-0.817)	0.925 (0.857-0.967)	0.877 (0.772-0.945)	0.817 (0.736-0.881)
Non-appendicitis sepsis	0.826 (0.612-0.950)	0.944 (0.897-0.974)	0.679 (0.476-0.841)	0.975 (0.936-0.993)

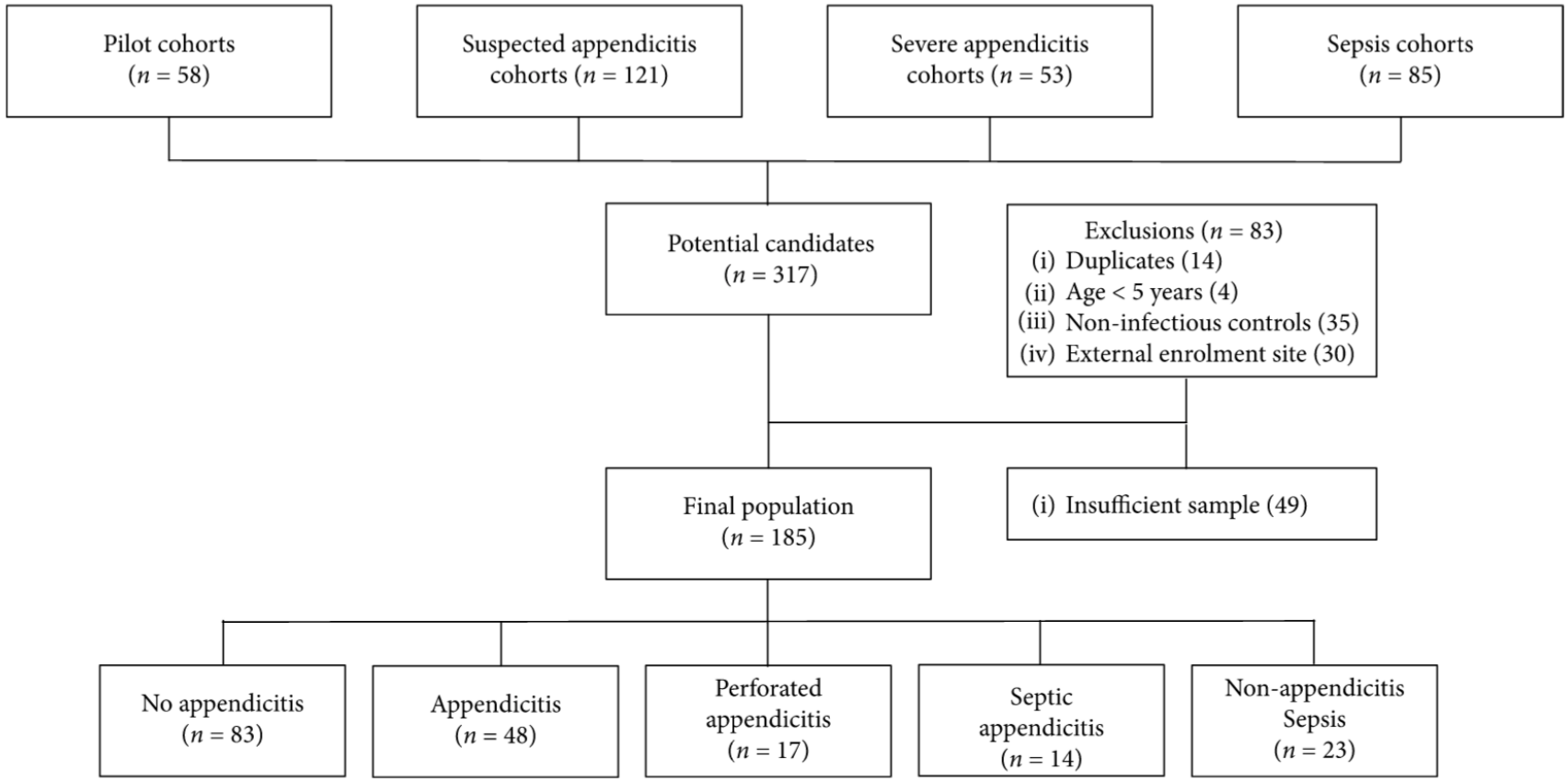


Figure 6-1. Selection of study candidates and final illness categories.

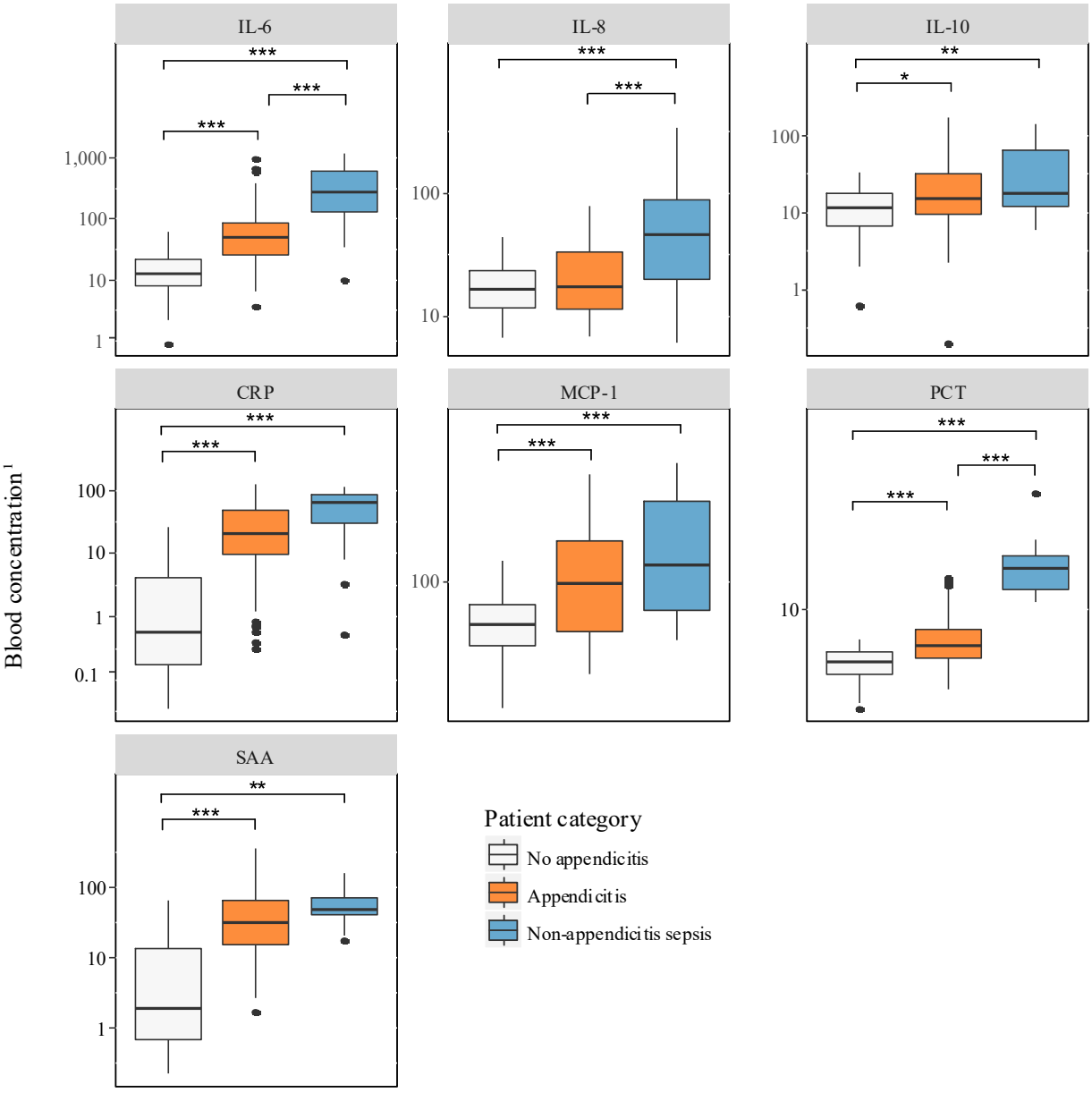


Figure 6-2. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category. Outliers (values > 95th percentile) are not presented but were included in statistical comparisons using multivariate normal regression followed by least-square means post-hoc pairwise contrasts using the Bonferroni correction.
[1]Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for IL-6, IL-8, IL-10, and MCP-1
* *P*-value < 0.05;** *P*-value < 0.01; *** *P*-value < 0.001.

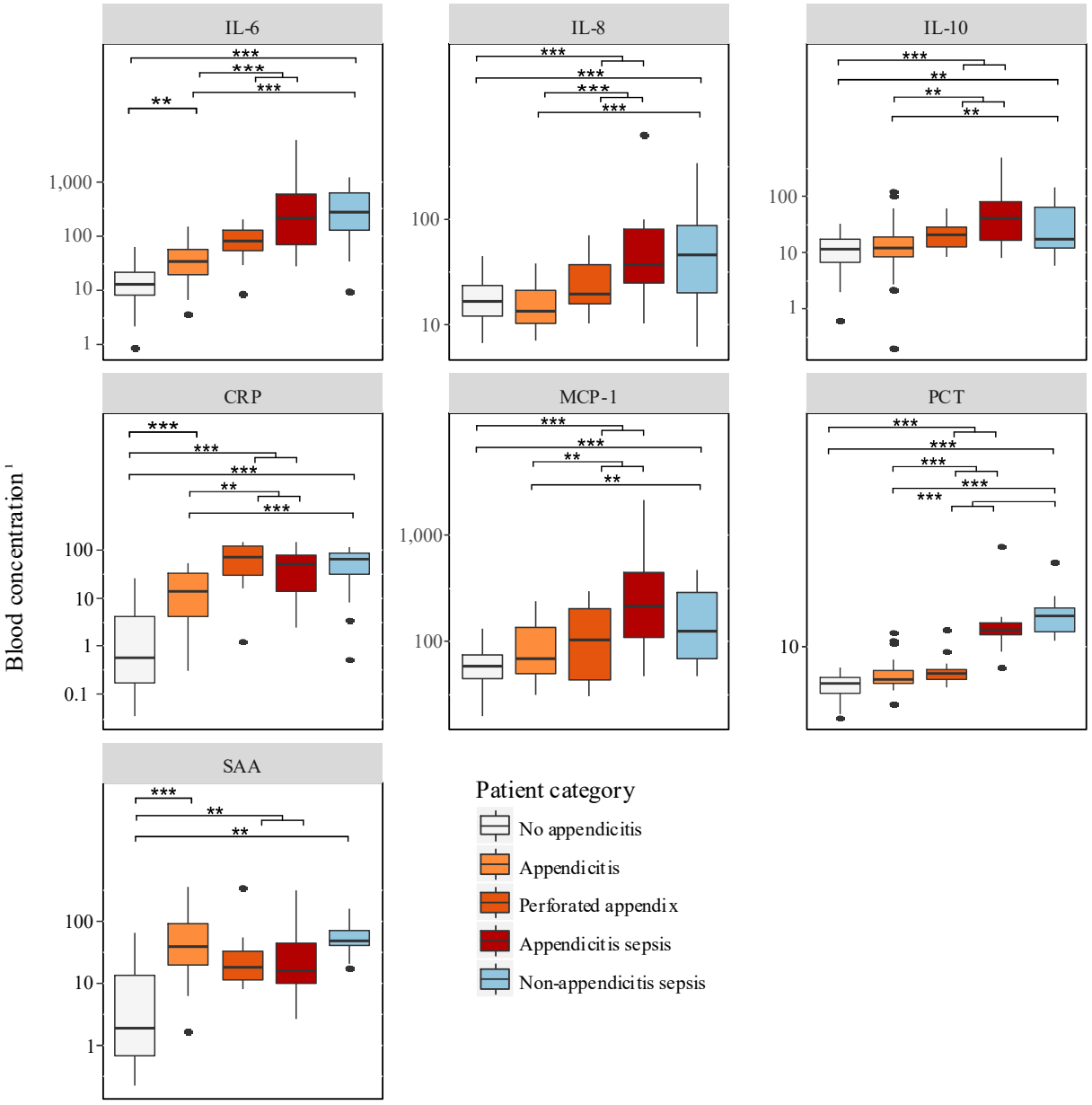


Figure 6-3. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category and appendicitis severity. Outliers (values > 95th percentile) are not presented in the plots but were included in statistical comparisons conducted using multivariate normal regression followed by least-square means post-hoc pairwise contrasts using the Bonferroni correction.
[1]Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for IL-6, IL-8, IL-10, and MCP-1
* *P*-value < 0.05; ** *P*-value < 0.01; *** *P*-value < 0.001.

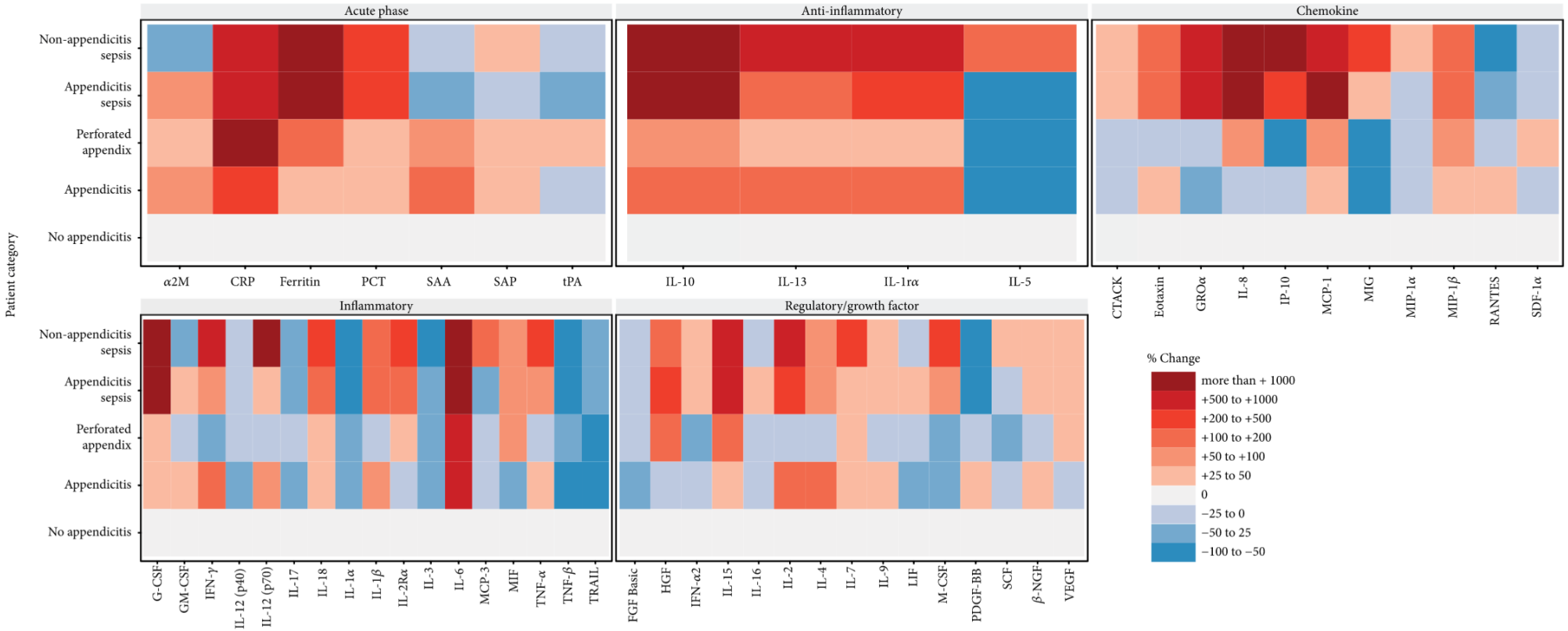


Figure 6-4. Heatmap showing relative change over mean concentration in patients without appendicitis of 54 selected cytokines grouped by patient category, appendicitis severity and class of cytokine based on 196 pediatric patients.

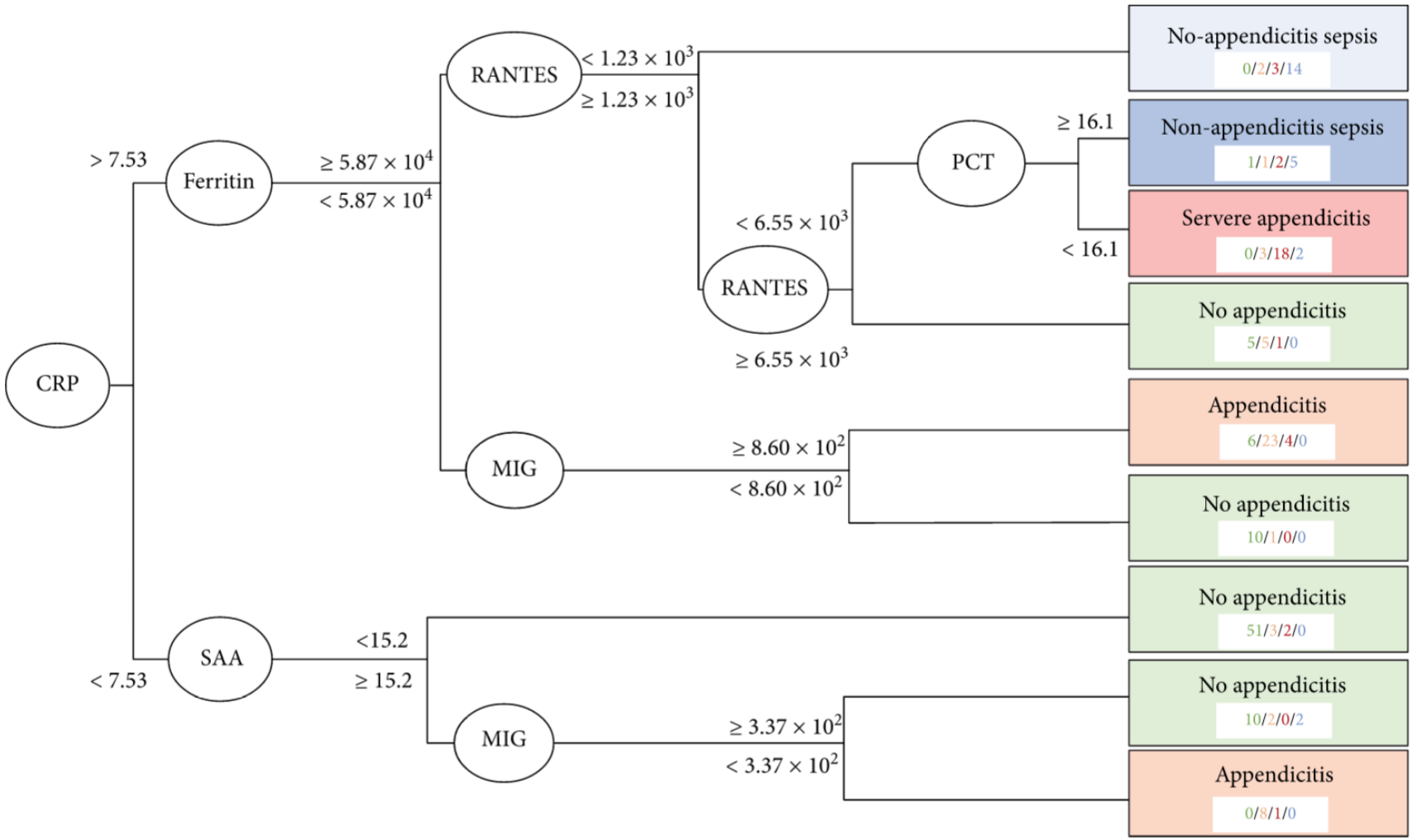


Figure 6-5. Decision tree for classifying patients into different categories or severities of appendicitis based on concentrations of 4 informative cytokines. Thresholds presented are blood concentrations of the cytokines Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for all others. Numbers below category refer to how many individuals in each terminal node have been classified into each category: green = no appendicitis; orange = appendicitis; red = severe appendicitis; blue = non-appendicitis sepsis.

6.9. Appendices

Appendix 6-1. Luminex Multiplex Assays of Inflammatory Protein Mediators.

Assay Kits (Bio-Rad Laboratories)	**Inflammatory Protein Mediators**
Bio-Plex Pro™ Human Cytokine 21-plex (#MG0005KMII)	Interleukin (IL)-1α, IL-2Rα, IL-3, IL-12 (p40), IL-16, IL18, cutaneous T-cell attracting chemokine (CTACK), growth related oncogene (Gro)-α, hepatocyte growth factor (HGF), interferon (IFN)-α2, leukemia inhibitory factor (LIF), monocyte chemotactic protein (MCP)-3, monocyte colony stimulating factor (M-CSF), macrophage migration inhibitory factor (MIF), monokine induced by gamma interferon (MIG), β-nerve growth factor (NGF), stem cell factor (SCF), stem cell growth factor (SCGF)-β, stromal cell-derived factor (SDF)-1α, tumor necrosis factor (TNF)-β, TNF-related apoptosis-inducing ligand (TRAIL).
Bio-Plex Pro™ Human Cytokine 27-plex (#M500KCAF0Y)	Fibroblast growth factor basic (FGF2), eotaxin, granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), IFN-γ, IL-1β, IL-1ra, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12 (p70), IL-13, IL-15, IL-17, Interferon gamma-induced protein (IP)-10, monocyte chemotactic protein (MCP)-1 , macrophage inflammatory protein (MIP)-1α and -1β, platelet-derived growth factor (PDGF)-BB, regulated on activation normal T-cell expressed and secreted (RANTES), tumour necrosis factor (TNF)-α, vascular endothelial growth factor (VEGF).
Bio-Plex Pro™ Human Acute Phase 5 + 4-plex Panel Complete (#171A4S07M)	Ferritin, fibrinogen, procalcitonin (PCT), serum amyloid A (SAA), tissue plasminogen activator (TPA), α-2-macroglobin (α2M), C-reactive protein (CRP), haptoglobin, serum amyloid P (SAP).

Appendix 6-2. Pairwise Pearson's correlation coefficient computed between cytokine concentrations of pediatric patients stratified by outcome category. Top-half of correlation matrix is omitted as it mirrors values in the lower half.

Patient Category (n)	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
No appendicitis (83)	IL-6	1.000						
	IL-8	0.373	1.000					
	IL-10	0.465	0.434	1.000				
	CRP	0.415	0.223	0.057	1.000			
	MCP-1	0.385	0.334	0.343	0.165	1.000		
	PCT	0.414	-0.010	0.184	0.429	0.202	1.000	
	SAA	0.407	0.199	0.129	0.855	0.164	0.430	1.000
	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
Appendicitis (79)	IL-6	1.000						
	IL-8	0.817	1.000					
	IL-10	0.590	0.595	1.000				
	CRP	0.371	0.312	0.162	1.000			
	MCP-1	0.746	0.766	0.459	0.266	1.000		
	PCT	0.746	0.745	0.418	0.277	0.746	1.000	
	SAA	-0.255	-0.233	-0.242	0.010	-0.001	-0.143	1.000
	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
Non-appendicitis sepsis (23)	IL-6	1.000						
	IL-8	0.703	1.000					
	IL-10	0.729	0.603	1.000				
	CRP	-0.208	0.000	-0.115	1.000			
	MCP-1	0.743	0.810	0.737	-0.022	1.000		
	PCT	0.688	0.605	0.711	0.041	0.712	1.000	
	SAA	0.112	0.146	0.318	-0.100	0.140	0.078	1.000

Appendix 6-3. Pairwise Pearson's correlation coefficient computed between cytokine concentrations of pediatric patients stratified by outcome category and severity of appendicitis. Top-half of correlation matrix is omitted as it mirrors values in the lower half.

Patient Category (n)	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
No appendicitis (83)	IL-6	1.000						
	IL-8	0.373	1.000					
	IL-10	0.465	0.434	1.000				
	CRP	0.415	0.223	0.057	1.000			
	MCP-1	0.385	0.334	0.343	0.165	1.000		
	PCT	0.414	-0.010	0.184	0.429	0.202	1.000	
	SAA	0.407	0.199	0.129	0.855	0.164	0.430	1.000
	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
Appendicitis (48)	IL-6	1.000						
	IL-8	0.695	1.000					
	IL-10	0.347	0.338	1.000				
	CRP	0.255	0.406	-0.021	1.000			
	MCP-1	0.598	0.457	-0.019	0.297	1.000		
	PCT	0.482	0.292	-0.122	0.234	0.594	1.000	
	SAA	0.037	0.133	-0.041	0.255	0.318	0.288	1.000
	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
Severe appendicitis (31)	IL-6	1.000						
	IL-8	0.841	1.000					
	IL-10	0.702	0.710	1.000				
	CRP	0.187	0.016	0.042	1.000			
	MCP-1	0.773	0.821	0.722	0.014	1.000		
	PCT	0.815	0.826	0.687	0.036	0.754	1.000	
	SAA	-0.370	-0.326	-0.346	-0.028	-0.067	-0.306	1.000
	Cytokine	IL-6	IL-8	IL-10	CRP	MCP-1	PCT	SAA
Non-appendicitis sepsis (23)	IL-6	1.000						
	IL-8	0.703	1.000					
	IL-10	0.729	0.603	1.000				
	CRP	-0.208	0.000	-0.115	1.000			
	MCP-1	0.743	0.810	0.737	-0.022	1.000		
	PCT	0.688	0.605	0.711	0.041	0.712	1.000	
	SAA	0.112	0.146	0.318	-0.100	0.140	0.078	1.000

Appendix 6-4. Model coefficients for the multivariate normal regression fit assessing differences between cytokine concentrations in pediatric patients (n=185) of different categories. Includes an adjustment for age and sex of patient.

Cytokine	Coefficient	Estimate	Standard Error	*P*-value
IL-6	Intercept†	2.720	0.207	< 0.001
	Appendicitis	1.673	0.256	< 0.001
	Non-appendicitis sepsis	3.363	0.385	< 0.001
	Age	-0.065	0.033	0.050
	Sex	0.137	0.240	0.569
IL-8	Intercept†	2.965	0.153	< 0.001
	Appendicitis	0.321	0.188	0.091
	Non-appendicitis sepsis	1.451	0.284	< 0.001
	Age	-0.033	0.024	0.172
	Sex	0.055	0.177	0.758
IL-10	Intercept†	2.611	0.186	< 0.001
	Appendicitis	0.547	0.230	0.019
	Non-appendicitis sepsis	1.250	0.346	< 0.001
	Age	-0.059	0.030	0.049
	Sex	-0.067	0.216	0.756
CRP	Intercept†	7.321	0.246	< 0.001
	Appendicitis	2.851	0.303	< 0.001
	Non-appendicitis sepsis	3.555	0.456	< 0.001
	Age	-0.115	0.039	0.004
	Sex	-0.243	0.284	0.394
MCP-1	Intercept†	4.125	0.137	< 0.001
	Appendicitis	0.760	0.170	< 0.001
	Non-appendicitis sepsis	1.290	0.255	< 0.001
	Age	0.009	0.022	0.675
	Sex	0.030	0.159	0.850
PCT	Intercept†	8.619	0.070	< 0.001
	Appendicitis	0.383	0.087	< 0.001
	Non-appendicitis sepsis	1.326	0.131	< 0.001
	Age	0.002	0.011	0.860
	Sex	0.107	0.082	0.190
SAA	Intercept†	9.048	0.296	< 0.001
	Appendicitis	2.048	0.366	< 0.001
	Non-appendicitis sepsis	1.956	0.550	< 0.001
	Age	-0.040	0.047	0.397
	Sex	0.171	0.343	0.619

†Intercept refers to reference group against which the coefficients are being compared: females, of the mean age (11.16yr) with no appendicitis.

Appendix 6-5. Model coefficients for the multivariate normal regression fit assessing differences between cytokine concentrations in pediatric patients (n=185) of different categories and severity of appendicitis. Includes an adjustment for age and sex of patient.

Cytokine	Coefficient	Estimate	Standard Error	*P*-value
IL-6	Intercept†	2.691	0.194	< 0.001
	Appendicitis	0.981	0.281	< 0.001
	Severe appendicitis	2.557	0.305	< 0.001
	Non-appendicitis sepsis	3.359	0.360	< 0.001
	Age	-0.063	0.031	0.044
	Sex	0.206	0.225	0.362
IL-8	Intercept†	2.945	0.145	< 0.001
	Appendicitis	-0.132	0.211	0.532
	Severe appendicitis	0.898	0.228	< 0.001
	Non-appendicitis sepsis	1.449	0.270	< 0.001
	Age	-0.032	0.023	0.169
	Sex	0.100	0.168	0.555
IL-10	Intercept†	2.591	0.180	< 0.001
	Appendicitis	0.072	0.261	0.782
	Severe appendicitis	1.152	0.282	< 0.001
	Non-appendicitis sepsis	1.247	0.334	< 0.001
	Age	-0.057	0.029	0.047
	Sex	-0.020	0.209	0.925
CRP	Intercept†	7.295	0.238	< 0.001
	Appendicitis	2.242	0.344	< 0.001
	Severe appendicitis	3.627	0.373	< 0.001
	Non-appendicitis sepsis	3.552	0.441	< 0.001
	Age	-0.113	0.038	0.003
	Sex	-0.182	0.275	0.508
MCP-1	Intercept†	4.110	0.133	< 0.001
	Appendicitis	0.413	0.192	0.033
	Severe appendicitis	1.203	0.208	< 0.001
	Non-appendicitis sepsis	1.288	0.246	< 0.001
	Age	0.010	0.021	0.626
	Sex	0.065	0.154	0.675
PCT	Intercept†	8.611	0.067	< 0.001
	Appendicitis	0.185	0.098	0.060
	Severe appendicitis	0.637	0.106	< 0.001
	Non-appendicitis sepsis	1.325	0.125	< 0.001
	Age	0.003	0.011	0.807
	Sex	0.127	0.078	0.106
SAA	Intercept†	9.065	0.294	< 0.001
	Appendicitis	2.443	0.427	< 0.001
	Severe appendicitis	1.543	0.462	0.001
	Non-appendicitis sepsis	1.958	0.547	< 0.001
	Age	-0.041	0.047	0.378
	Sex	0.132	0.341	0.700

†Intercept refers to reference group against which the coefficients are being compared: females, of the mean age (11.16yr) without appendicitis.

Appendix 6-6. Mean and standard deviation of 54 protein mediators in children with suspected appendicitis. All concentrations measured in pg/mL unless otherwise marked.

Category	Protein	Non-appendicitis Abdominal Pain (n=84)	Appendicitis (n=50)	Perforated Appendicitis (n=17)	Appendicitis-related Sepsis (n=14)	Non-appendicitis Sepsis (n=31)
Inflammatory	IL-1β	3.06 (11.49)	4.76 (11.49)	2.56 (1.79)	7.34 (12.36)	8.71 (27.24)
	IL-1rα	28.04 (568.32)	197.43 (568.32)	222.5 (393.41)	5337.7 (14259.47)	4834.78 (20025.24)
	IL-2	31.03 (88.38)	49.34 (88.38)	26.98 (33.72)	43 (51.16)	406.08 (1142.31)
	IL-4	91.25 (57.85)	65.49 (57.85)	82.1 (97.58)	61.74 (57.19)	51.47 (96.42)
	IL-5	227.46 (297.25)	262.06 (297.25)	296.43 (311.33)	11003.57 (30278.86)	5683.09 (17321.13)
	IL-6	56.7 (79.56)	75 (79.56)	54.75 (25.6)	61.62 (67.69)	39.48 (65.92)
	IL-7	102.96 (673.69)	231.54 (673.69)	64.9 (56.05)	156.45 (193.6)	624.83 (2392.47)
	IL-8	37.94 (135.31)	56.11 (135.31)	35.28 (45.45)	59.48 (66.69)	149.3 (498.36)
	IL-9	15.53 (9.84)	10.73 (9.84)	11.59 (11.65)	2.99 (4.36)	1.08 (1.99)
	IL-10	224.47 (156.02)	194.67 (156.02)	256.22 (175.56)	623.3 (624.74)	827.08 (1007.93)
	IL-12(p70)	1063.25 (706.37)	701.92 (706.37)	729.83 (587.9)	657.35 (1015.87)	329.39 (647.94)
	IL-13	1251.38 (735.85)	854.78 (735.85)	1038.81 (1099.58)	950.69 (1426.8)	1021.15 (2148.51)
	IL-15	83.92 (64.36)	84.13 (64.36)	87.19 (39.23)	193.86 (131.86)	392.61 (363.08)
	IL-17	96.35 (123.9)	86.57 (123.9)	75.86 (120.25)	49.47 (77.62)	209.94 (698.44)
	Eotaxin	281.31 (185.22)	179.89 (185.22)	497.93 (857.34)	478.17 (400.51)	462.48 (382.13)
	FGF Basic	14.53 (8.66)	6.22 (8.66)	7.61 (9.3)	4.11 (5.98)	2.14 (3.08)
	G-CSF	167.48 (57.92)	80.78 (57.92)	61.7 (74.01)	93.13 (90)	113.04 (186.99)
Anti-Inflammatory	GM-CSF	336.94 (1826.48)	839.12 (1826.48)	404.26 (481.62)	1867.09 (2389.17)	3001.06 (8695.13)
	IFN-γ	7.94 (3.23)	2.13 (3.23)	1.49 (1.46)	1.48 (0.91)	18.16 (56.39)
	IP-10	19.91 (121.03)	49.54 (121.03)	36.36 (36.46)	314.94 (632)	271.41 (864.88)
	MCP-1	5.27 (23.52)	10.62 (23.52)	6.11 (10.2)	11.61 (10.3)	37.12 (123.21)
Regulatory	MIP-1α	12.7 (65.61)	31.24 (65.61)	12.58 (19.45)	47.67 (102.83)	94.73 (231.41)
	PDGF-BB	1.61 (8.69)	3.4 (8.69)	1.58 (1.24)	2.7 (2.08)	2.92 (3.97)
	MIP-1β	10.39 (27.34)	14.56 (27.34)	12.24 (14.45)	12.65 (12.49)	54.84 (176.83)
	RANTES	29.95 (55.03)	36.14 (55.03)	28.78 (30.53)	31.49 (17.84)	42.14 (52.96)
	TNF-α	4.73 (10.88)	7.05 (10.88)	13.19 (16.15)	37.3 (56.77)	29.62 (56.01)
	VEGF	68.85 (27.55)	48.64 (27.55)	59.3 (42.84)	58.73 (32.19)	54.81 (47.41)
	IL-1α	268.51 (299.95)	279.49 (299.95)	233.41 (224.9)	111.33 (62.52)	120.42 (118.88)
	IL-2Rα	52.61 (24.63)	49.84 (24.63)	55.12 (36.56)	75.87 (50.75)	76.28 (114.11)
	IL-3	475.03 (224.47)	411.14 (224.47)	450.5 (275.41)	511.33 (369.03)	457.29 (325.58)
	IL-12(p40)	356.97 (274.26)	354.86 (274.26)	789.58 (743.33)	1607.22 (1404.51)	928.98 (664.94)
	IL-16	79.37 (69.76)	71.13 (69.76)	57.59 (40.47)	83.69 (39.06)	97.06 (77.22)
	IL-18	30.49 (27.41)	22.48 (27.41)	25.52 (28.38)	30.5 (21.57)	23.9 (53.24)
	CTACK	35.4 (20.02)	23.73 (20.02)	24.8 (20.52)	67.39 (70.74)	124.6 (337.73)
	GROα	3.9 (12.21)	5.52 (12.21)	3.33 (3.09)	5.63 (5.27)	5.48 (2.8)
	HGF	165.64 (84.54)	124.81 (84.54)	114.5 (58.27)	159.4 (118.39)	210.09 (181.72)
Chemokine	IFN-α2	24.37 (16.87)	21.41 (16.87)	43.64 (48.9)	770.92 (2081.97)	385.83 (922.46)
	LIF	40.49 (34.11)	46.5 (34.11)	32.01 (30.25)	100.36 (94.67)	116.13 (225.43)
	MCP-3	553.53 (864.12)	484.35 (864.12)	206.67 (165.27)	2891.13 (5579.93)	6803.96 (23573.25)
	M-CSF	88.32 (115.79)	129.72 (115.79)	164.76 (130.98)	1361.91 (2463.48)	604.99 (1786.25)
	MIF	4.68 (1.75)	3.98 (1.75)	4.29 (2.36)	4.64 (1.7)	4.89 (5.17)

Appendix 6-6. Continued.

	MIG	85.78 (64.83)	105.09 (64.83)	148.83 (174.25)	185.38 (118.52)	244.06 (558.94)
	β-NGF	5441.47 (3523.72)	5715.89 (3523.72)	4976.66 (3378.26)	3339.55 (3587.5)	1781.21 (1824.66)
	SCF	1058.86 (327.13)	904.19 (327.13)	977.03 (473.31)	1191.78 (666.65)	1329.83 (964.45)
	SDF-1α	92.66 (49.2)	64.3 (49.2)	88.23 (77.32)	867.44 (2477.29)	573.5 (1616.44)
	TNF-β	1958.52 (2540.74)	764.73 (2540.74)	602.98 (1251.25)	2299.7 (3371.95)	6584.13 (18715.15)
	TRAIL	321.12 (86.87)	272.81 (86.87)	329.56 (171.26)	256.62 (114.3)	260.62 (213.53)
Acute Phase	PCT[1]	6.09 (30.60)	7.35 (30.60)	8.12 (4.03)	20.98 (17.00)	33.06 (51.93)
	Ferritin	35519.11 (30061.62)	52622.23 (30061.62)	79630.09 (63234.64)	547857.19 (1022728.32)	444619.05 (979265.21)
	tPA	4550.68 (1837.21)	4252.04 (1837.21)	4697.83 (1337.32)	2347.36 (1086.55)	3615.36 (1475.96)
	SAA[2]	122.73 (435.79)	312.36 (435.79)	200.68 (296.19)	90.96 (132.13)	92.52 (83.47)
	α2M[1]	1680543.36 (3497764.31)	2926176.71 (3497764.31)	1953556.09 (1621340.26)	3034418.33 (3917537.66)	1101392.78 (422610.36)
	CRP[2]	7.62 (21.80)	23.41 (21.80)	91.82 (65.34)	72.97 (62.31)	70.58 (44.28)
	SAP[1]	25574.33 (16462.68)	26301.23 (16462.68)	27670.08 (12649.25)	24577.78 (14834.48)	32841.91 (13573.2)

[1]Concentrations measured in ng/mL
[2]Concentrations measured in mg/L

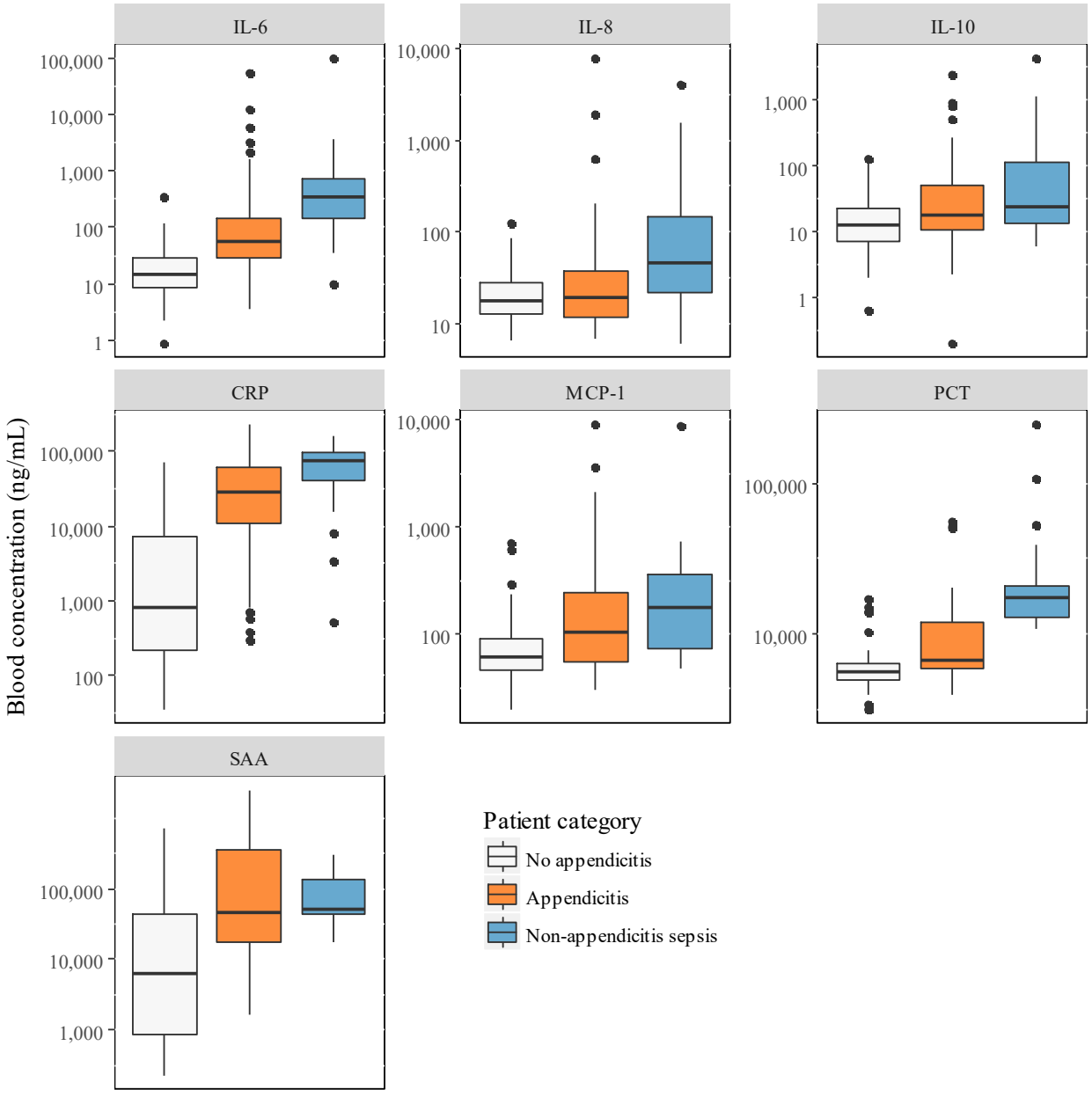
Appendix 6-7. Test characteristics of current "gold standard" evaluations in children with suspected appendicitis.

	Cut-off value	Sensitivity (%)	Specificity (%)	Accuracy (%)
WBC ($\times {10^3}/{\mu L}$)	>10	88.9	58.4	72.1
	>12	81.0	71.4	75.7
	>15	61.9	80.5	72.1
Neutrophils ($\times {10^3}/{\mu L}$)	>8	85.7	74.3	78.6
	>10	78.6	80.0	79.5
PAS	≥ 7	73.5	71.7	72.3

WBC – white blood cell count; PAS – pediatric appendicitis score

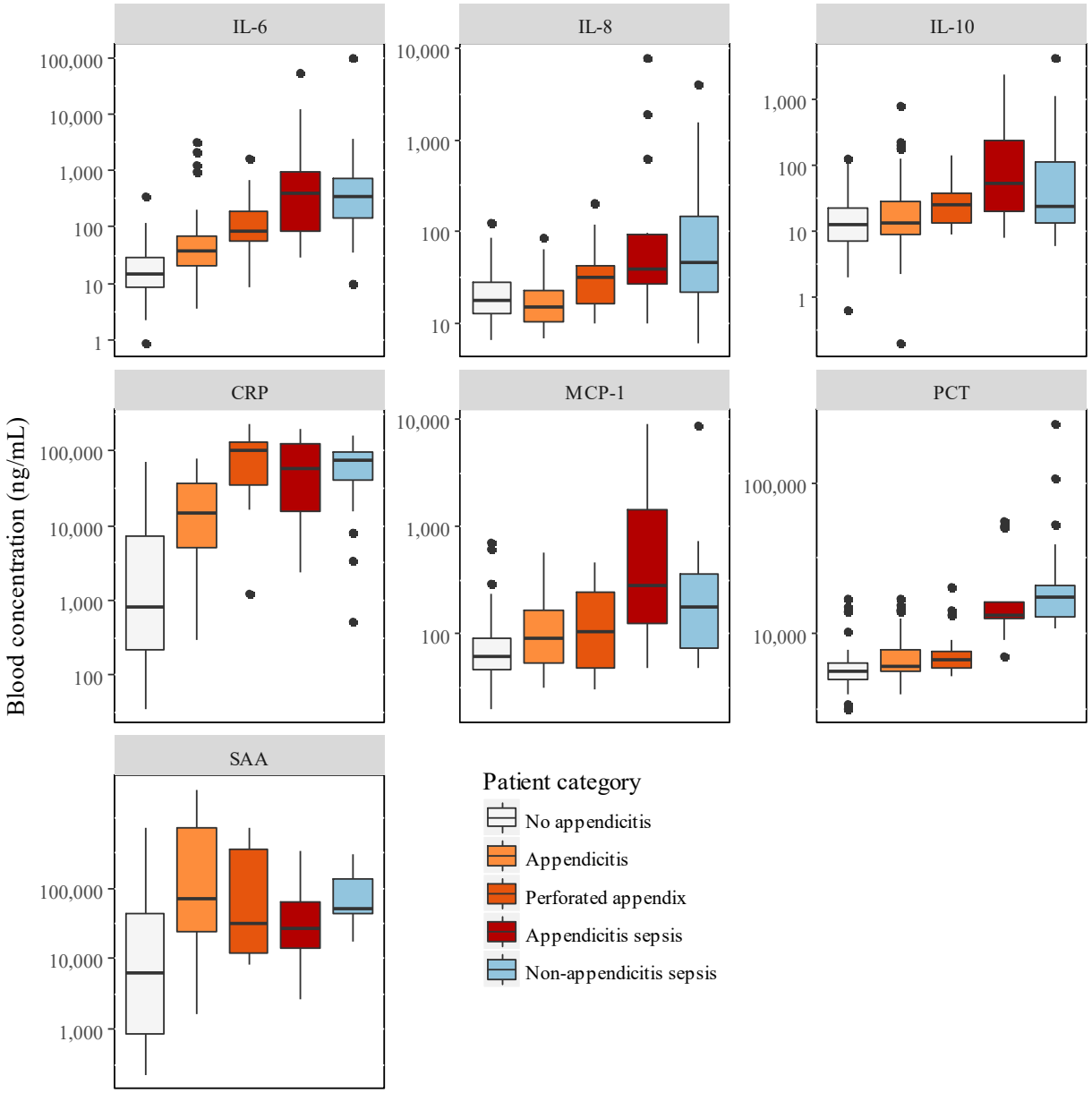
Appendix 6-8. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category, with outliers (values > 95th percentile) included.

[1]Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for IL-6, IL-8, IL-10, and MCP-1
* P-value < 0.05; ** P-value < 0.01; *** P-value < 0.001



Appendix 6-9. Boxplots of 7 selected cytokine concentrations in pediatric patients (n=196) grouped by category and appendicitis severity, with outliers (values > 95th percentile) included.

[1]Concentrations are in ng/mL for PCT; mg/L for CRP, and SAA; pg/mL for IL-6, IL-8, IL-10, and MCP-1
* P-value < 0.05; ** P-value < 0.01; *** P-value < 0.001



CHAPTER 7: Dietary patterns, food groups and nutrients in Crohn's disease: associations with gut and systemic inflammation

7.1. Abstract

This study examined associations between dietary intake and gut and systemic inflammation assessed by fecal calprotectin ≤ or > 100 μg/mg (FCP), C-reactive protein ≤ or > 5 mg/L (CRP) and serum cytokine profiles in Crohn's disease (CD) patients in clinical remission. A 3-month observational study was conducted at the University of Calgary in Calgary, Alberta, Canada between 2016 and 2018 in 66 outpatients with CD in clinical remission. FCP was obtained from stool samples at baseline and 3-months and serum CRP and serum cytokines were assessed at 3-months only (n = 41). Dietary intakes were collected using 3-day food records at baseline and 3-months and categorized as: PREDIMED Mediterranean diet scores (pMDS) total and individual components, the dietary inflammatory index (DII), food groups, and common micro- and macro-nutrients. Statistical models were developed to identify relationships between dietary factors and FCP, CRP and cytokine levels. Daily intake of leafy green vegetables was associated with FCP ≤ 100 μg/mg ($p < 0.05$). Increasing omega 6:3 ratio was associated with CRP ≤ 5 mg/L ($p = 0.02$). Different cytokines were significantly associated with various dietary variables. Future studies in patients with greater disease activity should be undertaken to explore these relationships.

7.2. Introduction

Inflammatory bowel disease (IBD) is believed to arise from a shared interaction between genetic and environmental influences (Opstelten et al., 2019), resulting in an imbalance between pro- and anti-inflammatory signalling. There has been a steady rise in incident cases of IBD over the past several decades with the highest prevalence rates seen in Europe and North America of approximately 0.3% of the population (Ng et al., 2018). Its recent increase in incidence across newly industrialized countries suggests environmental factors induce or modify disease expression. This has occurred in parallel with global dietary shifts towards a more "Western" style of eating. Common medical therapies include corticosteroids, immunomodulatory and biologic agents, and surgery. However, these therapies are not universally effective, the disease may recur and drugs may be associated with serious side effects. Recently, guidelines have included dietary therapies as a treatment option (Levine et al., 2020).

Western diets are characterized by high intakes of total, polyunsaturated fatty acids (PUFA) and saturated fats, red and processed meats, commercial baked goods, sweetened beverages and food additives such as carrageenan, maltodextrin and emulsifiers together with low intakes of phytochemicals, fibre, fruits, vegetables and fish (Ananthakrishnan et al., 2013; Nickerson et al., 2015; Bhattacharyya et al., 2017; Levine et al., 2019; Taylor et al., 2018). This dietary pattern can alter the gut microbiota and impair intestinal barrier function, thus resulting in a cascade of reactions that may trigger an enhanced inflammatory response. In contrast, prospective cohort and intervention studies have identified certain dietary patterns and individual diet components that may reduce gut inflammation and improve symptoms in IBD (Levine et al., 2019; Konijeti et al., 2017; Chiba et al., 2010; Suskind et al., 2018; Shivappa et al., 2016). While knowledge of

dietary relationships with the gut microbiome in IBD is evolving, the effects of dietary patterns and components on systemic markers of immune function is unexplored.

Given this context, it is justified to investigate the effect of dietary patterns and individual dietary components on surrogate biomarkers of Crohn's disease (CD) activity and systemic markers of inflammation. Fecal calprotectin (FCP) is a useful surrogate marker of gastrointestinal inflammation and levels less than 100 μg/g are consistently associated with disease remission, whereas levels > 250 μg/g predict disease relapse (D'Haens et al., 2012). C-Reactive Protein (CRP) is a marker of systemic inflammation and is associated with therapeutic response in IBD, and CRP levels correlate modestly with endoscopic disease activity (Kiss et al., 2012). Furthermore, serum cytokines have been associated with clinical remission and mucosal healing. With the above-described rationale for these disease specific biomarkers, our study objective was to examine associations between gut and systemic inflammation assessed FCP ≤ or > 100 μg/mg, CRP ≤ or > 5 mg/L with dietary intake in patients with CD in remission. Exploratory analyses between serum cytokine profiles and dietary components were undertaken to characterize the relationship between diet and systemic markers of immune status.

7.3. Materials and Methods

7.3.1. Study design

A 3-month observational study of patients with CD was conducted at the gastrointestinal outpatient clinic at the University of Calgary, Alberta, Canada between 2016 and 2018.

7.3.2. Inclusion/exclusion criteria

Inclusion criteria were: (a) ≥ 18 years of age; (b) documented diagnosis of luminal ileal and/or ileocolonic CD based on conventional criteria; (c) steroid-free clinical remission defined by a Harvey Bradshaw Index (HBI) < 5 for at least three months; d) achieved induction of remission either through anti-tumor necrosis factor (anti-TNF) agents or corticosteroids, and maintenance of remission with anti-TNF agents either with or without immunomodulators; and (e) able to provide informed consent, and read and write in English. Patients were excluded if they had (a) > 1 bowel resection; (b) evidence of fistulizing or stricturing disease CD; (c) upper gastrointestinal disease, or; (d) an ostomy.

7.3.3. Data collection

Patient-related characteristics (i.e., age, sex, height, weight, body mass index: BMI) and medical history (i.e., treatment with anti-tumor necrosis factor biologics and immunomodulators at baseline and previous small bowel resection) were captured during the first study visit (n = 66). Stool was obtained at baseline and 3-months for FCP, whereas serum was collected for CRP and cytokine levels at 3 months only in patients who presented for timely follow-up collection (n = 41); patient demographics are presented in Table 7-1. Several inflammatory (Tumour Necrosis Factor alpha (TNF-α), interferon gamma (IFN-γ), interleukins 4, 6, 8, 17a, and 12/23p40 (i.e., IL-4, IL-6, IL-8, IL-17a, IL-12/23p40) and regulatory (IL-10 and IL-27) cytokines (all pg/mL) were examined. Cytokines were evaluated using electro-chemiluminescent, multi-plex ELISA kits by MesoScale Discovery (Rockville, MD, USA) as per the manufacturer's protocol.

Dietary intake was determined using 3-day weighed food records at baseline and 3-months. Study participants met with the study coordinator to receive training on how to record an accurate 3-day weighed food record and were instructed to document all food and drinks consumed and preparation method on two representative non-consecutive weekdays, and one weekend day. The study dietitian reviewed the food record with the patient to identify any missing information.

Dietary intake data was analyzed using ESHA Food Analysis Software Version 11.3 X, ESHA Food Processor Nutrition Analysis Software (Salem, OR, USA), and individual macro- and micro-nutrients amounts were calculated. Dietary intakes were categorized using the following:

1) Mediterranean Diet Score (pMDS; Martinez-Gonzalez et al., 2012), modified to exclude red wine consumption (Taylor et al., 2018), was calculated for each patient to measure how closely their overall dietary pattern resembled a Mediterranean diet.
2) The dietary inflammatory index (DII) developed by Shivappa et al. (2014) was calculated to score the inflammatory potential of individual dietary patterns.
3) Canada's Food Guide to Healthy Eating (CFGHE) food group serving sizes was used to examine servings of fruits, vegetables, leafy green vegetables, red and processed meats, legumes, fish, and nuts and seeds (Health Canada 2011). Definitions of serving sizes for each food group have been previously reported (Taylor et al., 2018).

7.3.4. Statistical analyses

All data management and statistical analyses were conducted in R version 3.5.3 (R Core Team 2018), and a p-value < 0.05 was considered statistically significant. Outcomes of interest

included FCP, dichotomized at 100 μg/mg to discriminate between patients in active or inactive intestinal inflammation (Mumolo et al., 2018); CRP dichotomized at 5 mg/L (Vermeire et al., 2004), and various inflammatory and/or regulatory cytokines (Table 7-1). Dietary variables explored in this analysis included total pMDS and its sub-components (Taylor et al., 2018); the DII, CFGHE food group servings; macro- and micro-nutrient intake and miscellaneous aspects of dietary intake (Tables 7-2, 7-3). Other patient characteristics including age in years at baseline, body mass index, previous bowel surgery, height, sex, taking anti-TNF medication or immunosuppressants, and weight were assessed as potential confounders of the associations of interest.

The association between FCP and diet was assessed using a mixed-effects logistic regression model to account for within-patient correlations between baseline and follow-up, and was implemented using the R package "lme4" (Bates et al., 2015). This allowed the model to account for the within-patient correlation affecting the strength of the associations as it incorporated a random-effects term in the model to capture this correlation. Linear regression models were used to determine the association between diet and cytokine measurements using the R "stats" package (R Core Team, 2019), while logistic regression models were used to determine the association between CRP and diet.

Descriptive statistics and visualization were used to identify and explore outliers. The "ggplot2" library was used for all exploratory visualizations and plotting. Regression models were then built in a mixed stepwise manner, beginning with a forward-stepwise selection based on univariable associations with the outcomes ($P < 0.15$), followed by a backwards-elimination while assessing confounding (removal of the variable causes a change in the remaining coefficient estimates of ± 30%), and retained only the variables with $P < 0.10$. Interactions

between variables were not assessed due to the small sample sizes resulting from the stratification necessary to test the corresponding interactions.

7.3.5. Ethical approval

The study was approved by the Calgary Health Research Ethics Board (REB15-1805) and informed consent was provided by all patients. All the research was carried out in accordance with relevant guidelines and regulations.

7.4. Results

7.4.1. Demographics

Table 7-1 describes the demographics and health characteristics for patients at baseline and 3-months. At baseline, participants were generally middle aged (Median = 51 years, IQR = 37–58) overweight (Median = 27.3 kg/m^2, IQR = 23.5–29.7) and 49% were male. All patients were in clinical remission (HBI < 5) and 64% of patients were in biological remission defined as FCP ≤ 100 μg/mg. At baseline, 53 patients (80%) were receiving anti-TNF biologics, with 29 patients (44%) receiving immunomodulators and no patients receiving corticosteroids.

7.4.2. Dietary pattern and macro- and micro-nutrients

Patients' DII scores were generally pro-inflammatory (Mean = 1.6, SD = 1.8) and pMDS scores showed only 4–5 healthy behaviours being performed (Mean = 4.6, SD = 1.5) out of a possible 13. Means and proportions of dietary patterns, individual food groups and macro- and micro-nutrients used in the analyses are described in Tables 7-2 and 7-3.

7.4.3. Fecal calprotectin

Increasing daily servings of leafy green vegetables were associated with FCP ≤ 100 μg/mg ($p < 0.05$; Table 7-4, Figure 7-1).

7.4.4. C-reactive protein

When assessing the odds of having a low CRP at 3-months (defined as a serum concentration ≤ 5 mg/L), males had higher odds of having a low CRP than females (Table 7-4). In addition, a higher omega 6:3 fatty acids ratio increased the odds CRP would be ≤ 5 mg/L (Figure 7-2). The mean omega 6:omega-3 (n-6:n-3) PUFA ratio in the low CRP group was 8:1, compared to the high CRP group with a ratio of 4:1.

7.4.5. Serum cytokines

Serum cytokine relationships with dietary variables are presented in Table 7-4. Different cytokines were significantly associated with various dietary variables. Considering dietary pattern, a negative association between energy adjusted DII and IL-12/23 p40 ($P = 0.04$), and positive association between pMDS and IL-17A ($P = 0.03$) was observed, pMDS and DII scores were not associated with the remaining serum cytokines.

7.5. Discussion

We present the first prospective study to characterize the relationship between pMDS, DII, CFGHE food groups, micro- and macro-nutrients with serum and fecal biomarkers of disease activity and peripheral cytokine profiles in patients with CD in clinical remission. Daily servings

of leafy green vegetables was associated with FCP ≤ 100ug/mg, and an n-6:n-3 PUFA ratio of 8:1 was associated with CRP ≤ 5 mg/L. We did not identify relationships for FCP or CRP with dietary patterns measured by pMDS and energy-adjusted DII.

Epidemiological studies have associated fruit, vegetable and fish intake with lower risk of IBD incidence (Racine et al., 2016; Papada et al., 2019). Leafy greens are a rich source of carotenoids which may help improve intestinal barrier function and reverse lipopolysaccharides induced intestinal barrier damage through enhancing the expression of tight junction proteins (He et al., 2019). A plant-based diet, rich in dietary fiber, appears to be beneficial for human health by promoting the development of a diverse microbial ecosystem (Eom et al., 2018); in this study however, total dietary fiber was not associated with FCP or CRP. Exclusive enteral nutrition (EEN), a low-fibre product, is used effectively as primary therapy for CD in pediatric patients with luminal CD (Ruemmele et al., 2014). EEN may deplete the nutrients (e.g., fibre) necessary for bacterial growth, or create a luminal microenvironment where the functionality of inflammatory microbes is suppressed, inhibiting the gut mediated immune response in patients with CD (Svolos et al., 2019). Therefore, the mechanisms underlying the protective benefits of leafy greens may extend beyond their fiber content.

Higher n-6:n-3 PUFA ratios were associated with increased odds of patients having a CRP ≤ 5 mg/L. In typical Western diets this ratio can be as high as 20:1 (Simopoulos 2002), and disproportionate n-6:n-3 intake ratios may lead toward a proinflammatory response which may affect the onset of IBD (Scaioli et al., 2017). Supplementation with n-3 PUFA to alter the n6:n3 ratio towards anti-inflammatory effects has shown mixed results in CD and genetic polymorphisms in n-3 metabolism could be one possible mechanism to explain conflicting results (Scaioli et al., 2017). A study in zebrafish compared various levels of n-6:n-3 ratios in

fish consuming a standard diet of 12% total calories from lipids (Powell et al., 2015). Similar to our study, this study identified an 8:1 ratio as optimal compared to a 5:1 and a 2:1 ratio to promote greater gains in lean body mass and less total body fat. In addition, zebrafish eating the n-6:n-3 PUFA 8:1 ratio diet had lower levels of CRP and amyloid, markers of inflammation identified on liver biopsies. The authors concluded the ratio of n6:n3 may influence the inflammatory response less than the total percent of calories consumed from lipids. In our study, total intake of fat grams did not have any significant association with FCP, CRP or inflammatory cytokines.

Having identified dietary associations with FCP and CRP, we proceeded to examine potential immune mediated mechanisms to explain these results. Both expected and unexpected relationships between peripheral cytokines and dietary components and patterns were identified.

Unexpectedly, increasing scores for the pMDS (more anti-inflammatory dietary pattern) were associated with increasing levels of IL-17A. Similarly, lower energy adjusted DII scores (more anti-inflammatory) were associated with increased IL12/23p40. In 320 healthy Iranian women, an inverse relationship with IL-17A and adherence to a Mediterranean diet was identified (Sakhaei et al., 2018) while another study demonstrated increasing consumption of fruit juices reduced IL-17A (Peluso et al., 2012). Although IL-17A has generally been classified as a pro-inflammatory cytokine, it may regulate components of the intestinal barrier by reinforcing tight junction proteins in epithelial cells, and it induces the expression of mucin-associated genes and the production of β-defensins in the colon (Lee et al., 2015; Hueber et al., 2012). IL-17A may have protective effects and its role in gut homeostasis may not be completely understood.

In previous work the DII score was associated with increased levels of TNF-α and IFN-γ and not associated with IL12/23p40 as seen in this study (Shivappa et al., 2017); however, this was in

healthy adolescents and not in a diseased cohort of patients. In another study in adults, the DII was positively associated with IL-6 and TNF-α (Shivappa et al., 2015). The DII was not significantly associated to these cytokines in our study; however, our study collected data using 3-day food records compared to a food frequency questionnaire, did not control for the same confounders as the previous work, and the study populations were different.

We identified a significant relationship between IL-10 and β-carotene intake. IL-10 is an immunoregulatory cytokine that plays a crucial role in orchestrating intestinal immune homeostasis (Kuhn et al., 1993). However, these data need to be interpreted with caution due to the relatively small sample size and the possible disconnect between serum cytokine levels and gut inflammation. β-carotene is widely distributed in green leafy, yellow and orange coloured vegetables and fruits, and is one of the most abundant carotenoids. β-carotene has been reported to suppress LPS-induced inflammatory responses by inhibiting nuclear factor kappa B expression, suggesting an anti-inflammatory role (Bai et al., 2005).

We demonstrated an inverse relationship between red meat intake and IL-27 serum levels, and an inverse relationship between leafy green vegetables and IL-27. It has not been definitively determined whether IL-27 ameliorates or promotes intestinal inflammation, as contradictory roles for IL-27 have been reported in animal models of IBD. Studies have demonstrated the ability of IL-27 to lessen colonic inflammation by increasing the production of IL-10 in animal models of colitis (Dann et al., 2014; Hanson et al., 2014). These somewhat conflicting results may speak to the dual effects of IL-27 on intestinal inflammation, however mechanistic interpretations cannot be concluded at this time.

Study strengths include dietary data collection and FCP measurements at two time points using dietitian reviewed 3-day food records, and rigorous statistical modeling. Study limitations

to the regression-based model analysis were related largely to sample size. We selected inclusion and exclusion criteria that would allow us to have a clinical population that was representative of the typical patient in remission without being too restrictive or broad. All patients presenting to the outpatient clinic were informed of the study and over the 2-year period we were able to achieve an initial n = 66. This sample size allowed us sufficient power to examine our question but certainly suggests that future larger studies are warranted to further confirm the interesting findings of our study. As the number of variables and dietary factors observed were greater than the number of individuals, a risk factor analysis was underpowered to begin with. This small sample size made it impossible to assess interactions between predictor variables in the models, meaning that some key relationships between variables may not have been identified, as effect modification was not tested. As the purpose of this work was exploratory to provide preliminary associations for further investigation, reducing the number of significant associations detected by adjusting the significance level may result in potential relationships not being further investigated. In addition, while serum cytokine levels do not necessarily reflect mucosal inflammation, the identified relationships may add context and foundation for future dietary mechanistic studies. Raw *P*-values have been included in the results tables so they can be interpreted independently of a pre-defined significance cutoff.

In conclusion, we have identified relationships between dietary components and FCP and CRP in patients with CD in clinical remission, and provided preliminary associations between dietary components, patterns and nutrients with serum cytokines. The lack of robust associations between the DII and pMDS with FCP and CRP were unanticipated; however, the remitted patient population may limit these associations, and future studies building on this work should be undertaken in patients with greater disease activity.

7.6. References

Ananthakrishnan, A.N., H. Khalili, G.G. Konijeti, L.M. Higuchi, P. De Silva, J.R. Korzenik, C.S. Fuchs, W.C. Willett, J.M. Richter, and A.T. Chan. 2013. A prospective study of long-term intake of dietary fiber and risk of Crohn's Disease and ulcerative colitis. Gastroenterology 145:970–977. doi:10.1053/j.gastro.2013.07.050.

Angel Martinez-Gonzalez, M., A. Garcia-Arellano, E. Toledo, J. Salas-Salvado, P. Buil-Cosiales, D. Corella, M. Isabel Covas, H. Schroeder, F. Aros, E. Gomez-Gracia, M. Fiol, V. Ruiz-Gutierrez, J. Lapetra, R. Maria Lamuela-Raventos, L. Serra-Majem, X. Pinto, M. Angel Munoz, J. Waernberg, E. Ros, R. Estruch, and P.S. Investigators. 2012. A 14-item Mediterranean Diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. PLoS One 7. doi:10.1371/journal.pone.0043134.

Bai, S.K., S.J. Lee, H.J. Na, K.S. Ha, J.A. Han, H. Lee, Y.G. Kwon, C.K. Chung, and Y.M. Kim. 2005. beta-Carotene inhibits inflammatory gene expression in lipopolysaccharide-stimulated macrophages by suppressing redox-based NF-kappa B activation. Exp. Mol. Med. 37:323–334. doi:10.1038/emm.2005.42.

Bates, D., M. Maechler, B.M. Bolker, and S.C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67:1–48. doi:10.18637/jss.v067.i01.

Bhattacharyya, S., T. Shumard, H. Xie, A. Dodda, K.A. Varady, L. Feferman, A.G. Halline, J.L. Goldstein, S.B. Hanauer, and J.K. Tobacman. 2017. A randomized trial of the effects of the no-carrageenan diet on ulcerative colitis disease activity. Nutr. Heal. Aging 4:181–192. doi:10.3233/NHA-170023.

Chiba, M., T. Abe, H. Tsuda, T. Sugawara, S. Tsuda, H. Tozawa, K. Fujiwara, and H. Imai. 2010. Lifestyle-related disease in Crohn's disease: Relapse prevention by a semi-vegetarian diet. World J. Gastroenterol. 16:2484–2495. doi:10.3748/wjg.v16.i20.2484.

D'Haens, G., M. Ferrante, S. Vermeire, F. Baert, M. Noman, L. Moortgat, P. Geens, D. Iwens, I. Aerden, G. Van Assche, G. Van Olmen, and P. Rutgeerts. 2012. Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease. Inflamm. Bowel Dis. 18:2218–2224. doi:10.1002/ibd.22917.

Dann, S.M., C. Le, B.K. Choudhury, H. Liu, O. Saldarriaga, E.M. Hanson, Y. Cong, and L. Eckmann. 2014. Attenuation of intestinal inflammation in Interleukin-10-deficient mice infected with *Citrobacter rodentium*. Infect. Immun. 82:1949–1958. doi:10.1128/IAI.00066-14.

Eom, T., Y.S. Kim, C.H. Choi, M.J. Sadowsky, and T. Unno. 2018. Current understanding of microbiota- and dietary-therapies for treating inflammatory bowel disease. J. Microbiol. 56:189–198. doi:10.1007/s12275-018-8049-8.

Hanson, M.L., J.A. Hixon, W. Li, B.K. Felber, M.R. Anver, C.A. Stewart, B.M. Janelsins, S.K. Datta, W. Shen, M.H. McLean, and S.K. Durum. 2014. Oral delivery of IL-27 recombinant bacteria attenuates immune colitis in mice. Gastroenterology 146:210-221. doi:10.1053/j.gastro.2013.09.060.

He, C., J. Deng, X. Hu, S. Zhou, J. Wu, D. Xiao, K.O. Darko, Y. Huang, T. Tao, M. Peng, Z. Wang, and X. Yang. 2019. Vitamin A inhibits the action of LPS on the intestinal epithelial barrier function and tight junction proteins. Food Funct. 10:1235–1242. doi:10.1039/c8fo01123k.

Health Canada. 2011. Eating well with Canada's food guide. Health Canada, Ottawa, ON, Canada. URL: https://food-guide.canada.ca/en/.

Hueber, W., B.E. Sands, S. Lewitzky, M. Vandemeulebroecke, W. Reinisch, P.D.R. Higgins, J. Wehkamp, B.G. Feagan, M.D. Yao, M. Karczewski, J. Karczewski, N. Pezous, S. Bek, G. Bruin, B. Mellgard, C. Berger, M. Londei, A.P. Bertolino, G. Tougas, S.P.L. Travis, and S.C.D.S. Grp. 2012. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. Gut 61:1693–1700. doi:10.1136/gutjnl-2011-301668.

Kiss, L.S., M. Papp, B.D. Lovasz, Z. Vegh, P.A. Golovics, E. Janka, E. Varga, M. Szathmari, and P.L. Lakatos. 2012. High-sensitivity C-reactive protein for identification of disease phenotype, active disease, and clinical relapses in Crohn's disease: A marker for patient classification? Inflamm. Bowel Dis. 18:1647–1654. doi:10.1002/ibd.21933.

Konijeti, G.G., N. Kim, J.D. Lewis, S. Groven, A. Chandrasekaran, S. Grandhe, C. Diamant, E. Singh, G. Oliveira, X. Wang, B. Molparia, and A. Torkamani. 2017. Efficacy of the Autoimmune Protocol Diet for Inflammatory Bowel Disease. Inflamm. Bowel Dis. 23:2054–2060. doi:10.1097/MIB.0000000000001221.

Kuhn, R., J. Lohler, D. Rennick, K. Rajewsky, and W. Muller. 1993. Interleukin-10-deficient mice develop chronic enterocolitis. Cell 75:263–274. doi:10.1016/0092-8674(93)80068-P.

Lee, J.S., C.M. Tato, B. Joyce-Shaikh, F. Gulan, C. Cayatte, Y. Chen, W.M. Blumenschein, M. Judo, G. Ayanoglu, T.K. McClanahan, X. Li, and D.J. Cua. 2015. Interleukin-23-Independent IL-17 production regulates intestinal epithelial permeability. Immunity 43:727–738. doi:10.1016/j.immuni.2015.09.003.

Levine, A., J.M. Rhodes, J.O. Lindsay, M.T. Abreu, M.A. Kamm, P.R. Gibson, C. Gasche, M.S. Silverberg, U. Mahadevan, R.S. Boneh, E. Wine, O.M. Damas, G. Syme, G.L. Trakman, C.K. Yao, S. Stockhamer, M.B. Hammami, L.C. Garces, G. Rogler, I.E. Koutroubakis, A.N. Ananthakrishnan, L. McKeever, and J.D. Lewis. 2020. Dietary guidance from the International Organization for the Study of Inflammatory Bowel Diseases. Clin. Gastroenterol. Hepatol. 18:1381–1392. doi:10.1016/j.cgh.2020.01.046.

Levine, A., E. Wine, A. Assa, R.S. Boneh, R. Shaoul, M. Kori, S. Cohen, S. Peleg, H. Shamaly, A. On, P. Millman, L. Abramas, T. Ziv-Baran, S. Grant, G. Abitbol, K.A. Dunn, J.P. Bielawski, and J. Van Limbergen. 2019. Crohn's Disease Exclusion Diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. Gastroenterology 157:440-450. doi:10.1053/j.gastro.2019.04.021.

Mumolo, M.G., L. Bertani, L. Ceccarelli, G. Laino, G. Di Fluri, E. Albano, G. Tapete, and F. Costa. 2018. From bench to bedside: fecal calprotectin in inflammatory bowel diseases clinical setting. World J. Gastroenterol. 24:3681–3694. doi:10.3748/wjg.v24.i33.3681.

Ng, S.C., H.Y. Shi, N. Hamidi, F.E. Underwood, W. Tang, E.I. Benchimol, R. Panaccione, S. Ghosh, J.C.Y. Wu, F.K.L. Chan, J.J.Y. Sung, and G.G. Kaplan. 2017. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 390:2769–2778. doi:10.1016/S0140-6736(17)32448-0.

Nickerson, K.P., R. Chanin, and C. McDonald. 2015. Deregulation of intestinal anti-microbial defense by the dietary additive, maltodextrin. Gut Microbes 6:78–83. doi:10.1080/19490976.2015.1005477.

Opstelten, J.L., J.H.M. de Vries, A. Wools, P.D. Siersema, B. Oldenburg, and B.J.M. Witteman. 2019. Dietary intake of patients with inflammatory bowel disease: A comparison with

individuals from a general population and associations with relapse. Clin. Nutr. 38:1892–1898. doi:10.1016/j.clnu.2018.06.983.

Papada, E., C. Amerikanou, A. Forbes, and A.C. Kaliora. 2020. Adherence to Mediterranean diet in Crohn's disease. Eur. J. Nutr. 59:1115–1121. doi:10.1007/s00394-019-01972-z.

Peluso, I., A. Raguzzini, D. V Villano, E. Cesqui, E. Toti, G. Catasta, and M. Serafini. 2012. High fat meal increase of IL-17 is prevented by ingestion of fruit juice drink in healthy overweight subjects. Curr. Pharm. Des. 18:85–90. doi:10.2174/138161212798919020.

Powell, M.L., M.A. Pegues, A.J. Szalai, V.K. Ghanta, L.R. D'Abramo, and S.A. Watts. 2015. Effects of the dietary omega 3:omega 6 fatty acid ratio on body fat and inflammation in Zebrafish (*Danio rerio*). Comp. Med. 65:289–294.

Racine, A., F. Carbonnel, S.S.M. Chan, A.R. Hart, H.B. Bueno-de-Mesquita, B. Oldenburg, F.D.M. van Schaik, A. Tjonneland, A. Olsen, C.C. Dahm, T. Key, R. Luben, K.-T. Khaw, E. Riboli, O. Grip, S. Lindgren, G. Hallmans, P. Karling, F. Clavel-Chapelon, M.M. Bergman, H. Boeing, R. Kaaks, V.A. Katzke, D. Palli, G. Masala, P. Jantchou, and M.-C. Boutron-Ruault. 2016. Dietary patterns and risk of inflammatory bowel disease in Europe: Results from the EPIC study. Inflamm. Bowel Dis. 22:345–354. doi:10.1097/MIB.0000000000000638.

Ruemmele, F.M., G. Veres, K.L. Kolho, A. Griffiths, A. Levine, J.C. Escher, J.A. Dias, A. Barabino, C.P. Braegger, J. Bronsky, S. Buderus, J. Martin-de-Carpi, L. De Ridder, U.L. Fagerberg, J.P. Hugot, J. Kierkus, S. Kolacek, S. Koletzkou, P. Lionetti, E. Miele, V.M. Navas Lopez, A. Paerregaard, R.K. Russell, D.E. Serban, R. Shaoul, P. Van Rheenen, G. Veereman, B. Weiss, D. Wilson, A. Dignass, A. Eliakim, H. Winter, and D. Turner. 2014.

Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. J. Crohns Colitis 8:1179–1207. doi:10.1016/j.crohns.2014.04.005.

Sakhaei, R., S. Shahvazi, H. Mozaffari-Khosravi, M. Samadi, N. Khatibi, A. Nadjarzadeh, F. Zare, and A. Salehi-Abargouei. 2018. The Dietary Approaches to Stop Hypertension (DASH)-Style Diet and an alternative Mediterranean Diet are differently associated with serum inflammatory markers in female adults. Food Nutr. Bull. 39:361–376. doi:10.1177/0379572118783950.

Scaioli, E., E. Liverani, and A. Belluzzi. 2017. The imbalance between n-6/n-3 polyunsaturated fatty acids and inflammatory bowel disease: a comprehensive review and future therapeutic perspectives. Int. J. Mol. Sci. 18. doi:10.3390/ijms18122619.

Shivappa, N., J.R. Hebert, A. Marcos, L.-E. Diaz, S. Gomez, E. Nova, N. Michels, A. Arouca, E. Gonzalez-Gil, G. Frederic, M. Gonzalez-Gross, M.J. Castillo, Y. Manios, M. Kersting, M.J. Gunter, S. De Henauw, K. Antonios, K. Widhalm, D. Molnar, L. Moreno, I. Huybrechts, and H.S. Grp. 2017. Association between dietary inflammatory index and inflammatory markers in the HELENA study. Mol. Nutr. Food Res. 61. doi:10.1002/mnfr.201600707.

Shivappa, N., J.R. Hebert, S. Rashvand, B. Rashidkhani, and A. Hekmatdoost. 2016. Inflammatory potential of diet and risk of ulcerative colitis in a case-control study from Iran. Nutr. Cancer - an Int. J. 68:404–409. doi:10.1080/01635581.2016.1152385.

Shivappa, N., J.R. Hebert, E.R. Rietzschel, M.L. De Buyzere, M. Langlois, E. Debruyne, A. Marcos, and I. Huybrechts. 2015. Associations between dietary inflammatory index and inflammatory markers in the Asklepios Study. Br. J. Nutr. 113:665–671. doi:10.1017/S000711451400395X.

Shivappa, N., S.E. Steck, T.G. Hurley, J.R. Hussey, and J.R. Hebert. 2014. Designing and developing a literature-derived, population-based dietary inflammatory index. Public Health Nutr. 17:1689–1696. doi:10.1017/S1368980013002115.

Simopoulos, A.P. 2002. The importance of the ratio of omega-6/omega-3 essential fatty acids. Biomed. Pharmacother. 56:365–379. doi:10.1016/S0753-3322(02)00253-6.

Suskind, D.L., S.A. Cohen, M.J. Brittnacher, G. Wahbeh, D. Lee, M.L. Shaffer, K. Braly, H.S. Hayden, J. Klein, B. Gold, M. Giefer, A. Stallworth, and S.I. Miller. 2018. Clinical and fecal microbial changes with diet therapy in active inflammatory bowel disease. J. Clin. Gastroenterol. 52:155–163. doi:10.1097/MCG.0000000000000772.

Svolos, V., R. Hansen, B. Nichols, C. Quince, U.Z. Ijaz, R.T. Papadopoulou, C.A. Edwards, D. Watson, A. Alghamdi, A. Brejnrod, C. Ansalone, H. Duncan, L. Gervais, R. Tayler, J. Salmond, D. Bolognini, R. Klopfleisch, D.R. Gaya, S. Milling, R.K. Russell, and K. Gerasimidis. 2019. Treatment of active Crohn's Disease with an ordinary food-based diet that replicates exclusive enteral nutrition. Gastroenterology 156:1354+. doi:10.1053/j.gastro.2018.12.002.

Taylor, L., A. Almutairdi, N. Shommu, R. Fedorak, S. Ghosh, R.A. Reimer, R. Panaccione, and M. Raman. 2018. Cross-sectional analysis of overall dietary intake and Mediterranean dietary pattern in patients with Crohn's disease. Nutrients 10. doi:10.3390/nu10111761.

Vermeire, S., G. Van Assche, and P. Rutgeerts. 2004. C-reactive protein as a marker for inflammatory bowel disease. Inflamm. Bowel Dis. Bowel Dis. 10:661–665. doi:10.1097/00054725-200409000-00026.

Table 7-1. Patient demographics and health characteristics at baseline and 3-month follow-up

Variables	**Baseline n = 66**	**Patients with cytokine measurement at time 2 n = 41**	***P*-value of difference n = 41**
Median age, years, (IQR)	51 (37-58)	53 (36-60)	-
Gender, males (%)	32 (49)	18 (44)	-
Median disease duration, years (IQR)	10.0 (4.5-17.8)	10.5 (4.9-19.0)	-
Median body mass index, kg/m^2, (IQR)	27.3 (23.5-29.7)	27.5 (24.8-29.5)	-
Current smoker, n (%)	3, (4.5)	1, (2.4)	-
Anti-TNF use, n (%)	53 (80)	29 (71)	-
Steroid use, n (%)	0	0	-
Immunomodulators use, n (%)	29 (44)	16 (39)	-
Surgical Hx, n (%)	18 (27)	8 (20)	-
Median dietary inflammatory index, energy adjusted, (IQR)	+1.9 (0.5-2.8)	+2.2 (1.1-3.2)	0.06[1]
Median Mediterranean diet score, (IQR)	4.0 (4.0-5.0)	4.0 (3.0-5.0)	0.08[1]
Fecal calprotectin, μg/mg Median (IQR)	73 (28-216)	73 (25-288)	0.79[1]
Biological remission, ≤ 100 μg/mg, (%)	42 (64)	22 (54)	0.18[2]
C-reactive protein mg/L Median (IQR)	n/a	1.47 (0.80-3.79)	n/a
≤ 5 mg/L		9 (22)	
TNF-α, pg/mL, median (IQR)	n/a	1.10 (0.77-1.44)	n/a
IFN-γ, pg/mL, median (IQR)	n/a	2.87 (1.90-5.10)	n/a
IL-4, pg/mL, median (IQR)	n/a	0.024 (0.02-0.04)	n/a
IL-6, pg/mL, median (IQR)	n/a	0.28 (0.17-0.39)	n/a
Il-8, pg/mL, median (IQR)	n/a	6.02 (5.43-7.62)	n/a
IL-17A, pg/mL, median (IQR)	n/a	1.70 (1.23-2.47)	n/a
IL-12/23p40, pg/mL, median (IQR)	n/a	75.12 (49.08-101.35)	n/a
IL-10, pg/mL, median (IQR)	n/a	0.16 (0.12-0.29)	n/a
IL-27, pg/mL, median (IQR)	n/a	370.66 (286.45-472.59)	n/a

[1]P-value computed using the Wilcoxon Signed-Rank test for paired data
[2]P-value computed using McNemar's Chi-squared test for paired data

Table 7-2. Description of dietary pattern variables in the mixed-stepwise selection process for the regression models.

Variables	Units/Levels	Median (IQR) at baseline (n=66)	Median (IQR) at follow-up (n=41)
Diet Scores			
Mediterranean diet score (pMDS)	-	4.00 (4.00-5.00)	4.00 (3.00-5.00)
Dietary inflammatory index (DII), energy adjusted	-	1.89 (0.46-2.78)	2.16 (1.10-3.27)
Food Groups			
Vegetables 1/2 cup, 1 medium or 1 cup leafy greens	Servings/d	1.30 (0.58-2.43)	1.42 (0.61-3.03)
Leafy green vegetables, 1 cup raw	Servings/d	0.30 (0.00-0.80)	0.25 (0.00-1.02)
Fruit, 1/2 cup or 1 medium and juice, 1/2 cup	Servings/d	1.60 (1.00-2.78)	1.33 (0.67-2.70)
Fruit, 1/2 cup or 1 medium, no juice	Servings/d	1.50 (0.78-2.63)	1.28 (0.67-2.38)
Legumes, ¾ cup	Servings/wk	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Nuts, ¼ cup	Servings/wk	0.25 (0.00-4.40)	1.17 (0.00-2.63)
Red and processed meat, 84 g (3oz)	Servings/d	1.15 (0.40-1.90)	0.33 (0.00-0.69)
Red meat, 84 g (3oz)	Servings/d	0.40 (0.00-1.13)	0.00 (0.00-0.47)
Processed meat, 84 g (3oz)	Servings/d	0.40 (0.08-0.93)	0.50 (0.00-0.43)
Fish and shellfish (84 g)	Servings/wk	0.00 (0.00-4.10)	2.33 (0.00-8.11)
Fish (84 g)	Servings/wk	0.00 (0.00-2.60)	1.05 (0.00-7.06)

Table 7-3. Description of macro- and micro-nutrients and single dietary components included in a mixed-stepwise selection process for regression models.

Variable	Type	Units/Levels	Median (IQR) at baseline (n=66)	Median (IQR) at follow-up (n=41)
Macronutrients				
Calories	Measured	kcal	2123.21 (1635.79-2401.25)	1838.41 (1611.78-2372.05)
Carbohydrates	Measured	g	248.49(192.85-287.67)	227.88 (186.52-274.27)
Fiber	Measured	g	20.09 (15.07-29.72)	19.72 (16.20-28.23)
Fat	Measured	g	80.09 (53.97-99.67)	66.00 (56.84-97.07)
Mono-unsaturated fatty acids	Measured	g	15.27 (10.11-24.77)	13.97 (7.27-20.67)
Omega-3	Measured	g	0.71 (0.48-1.15)	0.74 (0.37-1.01)
Omega-6	Measured	g	6.00 (3.81- 10.02)	4.94 (2.44-7.12)
Omega6:Omega3	Continuous; computed	-	8.74 (6.05-10.75)	7.31 (4.12-9.71)
Poly-unsaturated fatty acids	Measured	g	8.84 (5.69-12.52)	7.20 (4.11-9.14)
Saturated fat	Measured	g	25.13 (17.28-32.41)	23.96 (16.90-34.80)
Trans-fats	Measured	g	0.55 (0.18-1.20)	0.56 (0.20-0.81)
Cholesterol	Measured	mg	280.90 (204.36-400.24)	226.17 (141.91-376.47)
Protein	Measured	g	89.66 (70.16-108.73)	82.39 (61.97-102.00)
Single dietary components				
Vitamin A	Measured	µg	439.58 (264.85-711.58)	410.10 (231.01-702.14)
Beta-carotene	Measured	µg	712.29 (202.36-2962.85)	680.41 (194.79-2061.24))
Vitamin B1	Measured	mg	1.05 (0.68-1.46)	0.69 (0.47-1.08)
Vitamin B2	Measured	mg	1.33 (0.99-2.04)	1.09 (0.68-1.70)
Vitamin B3	Measured	mg	16.60 (11.83-22.80)	13.06 (8.97-14.51)
Vitamin B6	Measured	mg	86.85(66.21-121.99)	1.09 (0.69-1.80)
Folate, B9	Measured	µg	263.03 (158.77-353.83)	162.23 (117.27-251.73)
Vitamin B12	Measured	µg	2.54 (1.95-4.48)	2.58 (1.52-3.92)
Vitamin C	Measured	mg	66.17 (49.83-130.92)	71.44 (39.91-127.54)
Vitamin D	Measured	µg	2.01 (1.10-3.47)	2.40 (1.09-4.24)
Vitamin E	Measured	mg	3.76 (2.76-6.70)	2.74 (2.02-4.25)
Iron	Measured	mg	13.06 (10.40-16.09)	10.80 (9.42- 15.92)
Magnesium	Measured	mg	188.32 (141.63-279.11)	161.99 (110.13-233.85)
Potassium	Measured	mg	1894.46 (1494.45-2542.69)	1755.77 (1094.58-2282.07)
Selenium	Measured	µg	72.73 (45.85-104.94)	52.54 (37.50-74.47)
Sodium	Measured	mg	3316.86 (2503.30-4239.60)	2720.15 (2150.07-3497.78)
Zinc	Measured	mg	5.70 (4.48-9.83)	5.06 (3.04-6.62)
Alcohol/d	Measured	g	0.00 (0.00-6.47)	0.00 (0.00-5.21)
Caffeine	Measured	mg	94.68 (18.97-189.44)	76.98 (10.03-217.01)

Table 7-4. Final multivariate model results for fecal calprotectin, C-reactive protein, and pro- and anti-inflammatory cytokines.

Outcome	Variable	Coefficient (95% CI)	Standard error	*P*-value
Fecal calprotectin				
>100 μg/mg	Reference[1]	1.04 (-0.40 to 3.20)	0.79	0.19
	Leafy green vegetables/d	-1.14 (-2.64 to -0.12)	0.58	<0.01
	Vitamin B2 intake	- 0.90 (-2.48 to -0.03)	0.54	0.09
C-reactive protein				
>5 mg/L	Reference	1.95 (-0.14 to 4.61)	1.17	0.10
	Omega 6:3 ratio	-0.37 (-0.76 to -0.10)	0.16	0.02
	Sex	-2.91 (-6.20 to -0.68)	1.33	0.03
Pro-inflammatory cytokines (pg/ml)				
TNF-α	Reference	-0.46 (-1.45 to 0.53)	0.487	0.35
	Age	0.01 (-1.92x10^{-3} to 0.01)	0.01	0.10
	Body mass index	0.04 (3.15x10^{-3} to 0.08)	0.02	0.03
IFN-γ	Reference	1.17 (-2.33 to 4.66)	1.72	0.50
	Processed meat consumption	17.06 (9.83 to 24.29)	3.57	< 0.01
IL-4	Reference	4.71 x10^{-3} (-0.01 to 0.02)	0.01	0.40
	Alcohol consumption	-5.81 x10^{-4} (-9.89 x10^{-4} to -1.72 x10^{-4})	2.30 x10^{-4}	0.01
	Not taking anti-TNF	8.61 x10^{-3} (-3.95 x10^{-4} to 0.02)	4.42 x10^{-3}	0.06
	Iron intake	1.68 x10^{-3} (8.03 x10^{-4} to 2.55 x10^{-3})	4.12x10^{-4}	< 0.01
IL-6	Reference	0.02 (0.22 to 0.18)	0.10	0.85
	Fiber intake	0.01 (1.69 x10^{-3} to 0.02)	0.01	0.02
	Weekly fish consumption	-0.01 (-0.03 to 2.90 x10^{-3})	0.01	0.10
	Omega-3 fatty acids	0.37 (0.18 to 0.55)	0.09	< 0.01
	Poly-unsaturated fatty acids	-0.02 (-0.05 to 3.95 x10^{-3})	0.01	0.10
IL-8	Reference	2.05 (-0.41 to 4.51)	1.20	0.10
	Alcohol consumption	-0.09 (-0.15 to -0.02)	0.03	0.01
	Beta-carotene intake	4.11 x10^{-4} (1.74x10^{-5} to 8.05 x10^{-4})	2.23x10^{-4}	0.04
	Fiber intake	0.09 (0.01 to 0.17)	0.04	0.03
	Inadequate fruit pMDS	1.89 (0.41 to 3.37)	0.72	0.01
	Vitamin B3 Intake	0.07 (0.01 to 0.16)	0.04	0.07
IL-17A	Reference	1.80 (0.66 to 2.96)	0.57	<0.01
	Inadequate fruit pMDS	-0.77 (-0.15 to -0.09)	0.34	0.03
	Not taking anti-TNF medication	-0.70 (-1.30 to -0.09)	0.30	0.03
	Total pMDS	0.21 (0.03 to 0.39)	0.09	0.03
Il-12/23 p40	Reference	-86.31 (-200.16 to 27.54)	55.96	0.13
	Body mass index	7.26 (3.37 to 11.14)	1.91	<0.01
	Energy adjusted DII[3]	-10.34 (-20.07 to -0.61)	4.78	0.04
Anti-inflammatory/regulatory cytokines				
IL-10	Reference	0.11 (-0.04 to 0.25)	0.01	0.14
	Beta-carotene intake	1.06 x10^{-4} (4.65 x10^{-5} to 1.65 x10^{-4})	3.20x10^{-5}	<0.01
IL-27	Reference	45.70 (-135.52 to 226.91)	89.17	0.61
	Leafy green vegetables/day	-75.14 (-123.18 to -27.11)	23.6	<0.01
	Red meat/day	-78.23 (-135.11 to -21.36)	27.99	<0.01
	Weight (kg)	1.06 (0.57 to 1.56)	0.24	< 0.01

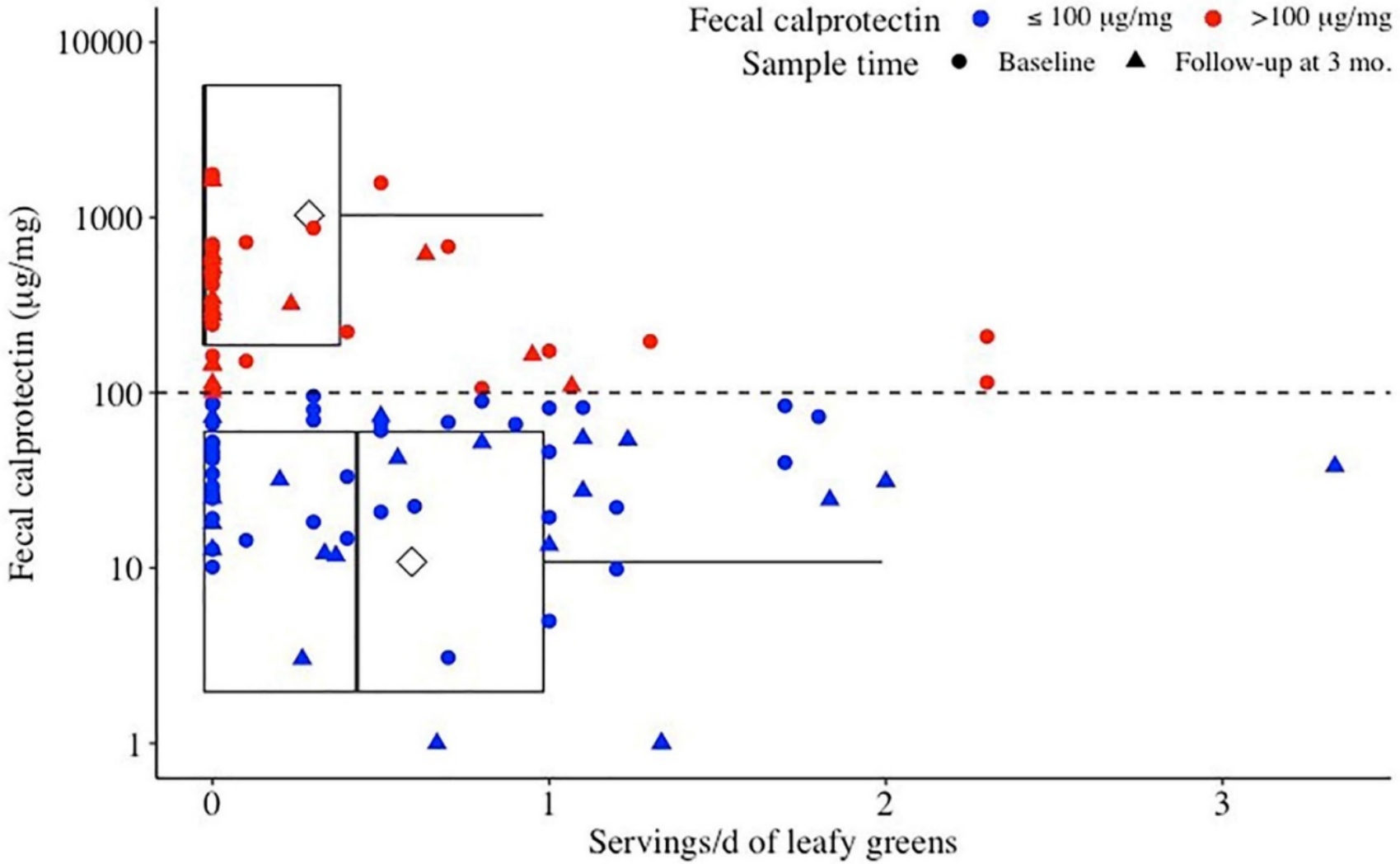


Figure 7-1. Fecal calprotectin ≤ 100 µg/mg is significantly related to leafy greens intake.

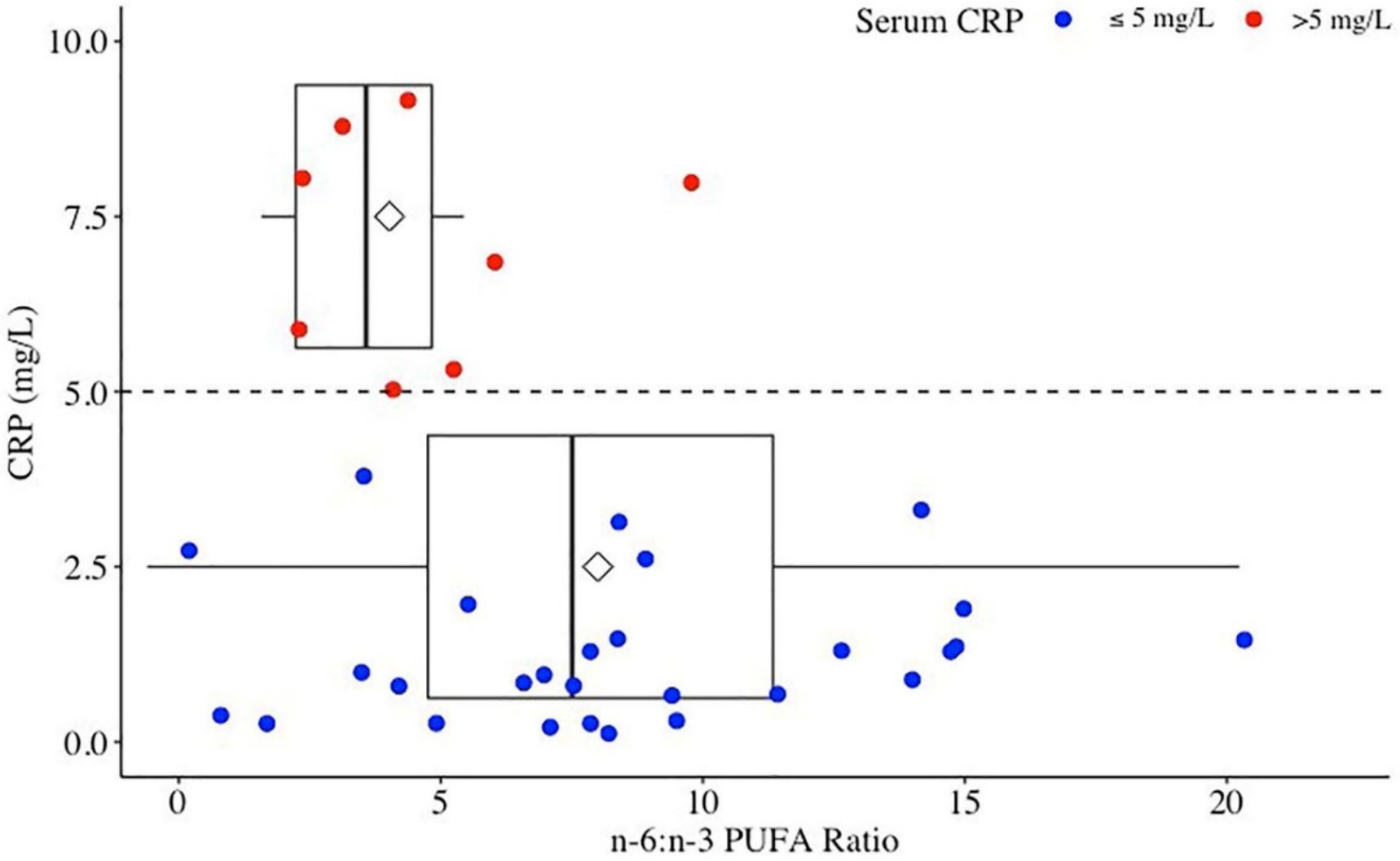


Figure 7-2. C-reactive protein ≤ 5 mg/L is significantly related to ratio of omega-6 to omega-3 polyunsaturated fat.

CHAPTER 8: Summarizing discussion

8.1. Concluding remarks

Rapid developments in data collection and storage technologies, along with increasing availability of high-performance computational resources have resulted in the potential for applying Big Data analytics in health research. Digitization of health and treatment records on dairy farms and in medical settings, high-frequency data from physiological sensors, and diagnostic tests that can quickly and accurately quantify large numbers of biologically important analytes are all data sources that can provide invaluable data in health research. However, to ensure validity and robustness of insights derived from these types of data, it is crucial that the most appropriate analytical methodologies be chosen, while also considering the limitations inherent to these data. Our results demonstrate how 1) dairy herd improvement (DHI) data, bacteriological culturing, and scientific databases can be used to facilitate effective management of mastitis in primiparous dairy cattle; 2) regularly collected data from automated milking systems (AMS) can be used to develop effective models for the detection of clinical mastitis in dairy cattle, while being robust to many of the challenges faced when combining data from varied sources; and 3) multiplex immunoassays can be used to understand disease severity, predict clinical outcomes and evaluate the effectiveness of diet when combined with other types of records.

8.1.1. Mastitis management in primiparous dairy cattle

Primiparous dairy cattle have typically been thought to be at low risk for diseases of the udder since they have not been exposed to the physiological stress of milk production and regular milking (De Vliegher et al., 2012). We demonstrated that this is not the case in Canada, as the incidence of clinical mastitis (CM) in primiparous cattle was significantly higher than in multiparous cows immediately after calving, and a higher proportion of primiparous cattle than multiparous cattle had intramammary bacterial infections during the first month after calving. In addition, the pathogens causing CM and intramammary infections (IMI) in primiparous cows were very different from those causing CM and IMI in multiparous cows, likely a reflection of the different environment the animals are exposed to prior to calving. Since the teat canal can open and be susceptible to infection up to 80d before calving (Krömker and Friedrich, 2009), pre-calving management of primiparous cattle will play an important role in the prevention of disease in the postpartum period.

Our systematic review and meta-analysis of studies assessing the effectiveness of pre-calving treatments on udder health after calving demonstrated that a number of different types of treatment are effective at improving udder health, but the effectiveness of treatments can vary considerably depending on the type of treatment, and the type of pathogen being targeted. Teat sealants were most effective at reducing the incidence and prevalence of IMIs by environmental pathogens, whereas antimicrobials, teat sealants, and vaccines were equally effective at reducing the occurrence of IMIs by contagious pathogens. Non-*aureus* staphylococci (NAS) are a diverse group of bacteria, and the prevalence of NAS IMIs is most effectively reduced by using a combination of different treatment types. Although antimicrobials were effective at reducing the prevalence of all types of IMIs, their use must be balanced with concerns about potentials for

antimicrobial resistance, and non-antimicrobials of equal effectiveness would be preferred. These results indicate that efficient management of mastitis in primiparous cattle should start before calving and should be informed by an understanding of the types of pathogens being targeted, which would require regular, pre-calving milk sampling for determination of the distribution of bacteria.

Both Chapters 2 and 3 highlighted the need for targeted data collection when designing disease control programs on farms. There is an immense volume of data collected on lactating animals on a dairy farm such as health and treatment records, milk production and components, and activity or behaviour for estrus detection, as these directly support the commercial operation of dairy farms, whereas typically only health and treatment records are kept for young stock. To take advantage of Big Data analytics most effectively for regular management on dairy farms, it would be necessary to make data collection and sampling from non-lactating animals more thorough. As technology adoption continues to grow in the dairy industry, and more technology is specifically developed for younger animals or becomes more accessible to allow use on non-lactating animals, targeted data collection will allow producers to develop and monitor their own preventive disease management programs.

8.1.2. Clinical mastitis detection using recurrent neural networks

The adoption of AMS on dairy farms in Canada is continuing to grow at a rapid pace (Barkema et al., 2015; Dairy barns by type, 2020), allowing producers to more easily monitor production, and improve milking labour management (Tse et al., 2018). Although AMS incorporate many different types of sensors, and the associated software generate alerts for potentially sick cows, only 23% of producers using AMS report relying on those alerts (Tse et

al., 2017). This likely indicates that the performance of the disease detection algorithms is not satisfactory, and either needs to be improved, or at least modified to better suit individual producers' management styles.

Studies that developed models for the detection of CM typically relied on a small number predictor variables, typically consisting of different milk characteristics (Hogeveen et al., 2010; Slob et al., 2021). In addition to milk yield and components, sensors accompanying AMS also collect information about animal behaviour and feeding which are also known to change during onset of CM (Siivonen et al., 2011; Sepúlveda-Varas et al., 2016; King et al., 2018). The results presented in Chapter 4 demonstrate that RNNs can combine observations from different sources on a dairy farm, to effectively detect the onset of CM in dairy cattle. Although variables associated with behaviour and feeding were not effective predictors for CM onset on their own, combining them with milk characteristics did improve model performance with both sensitivity and specificity > 0.80. These data are regularly collected by AMS, and so can readily be incorporated into predictive models without the need for additional hardware. In addition, unlike previous studies that downsample "healthy" observations to better balance the outcome (Ankinakatte et al., 2013; Khatun et al., 2018), we included all observations and applied a weighted loss function. This allowed us to evaluate how the model would perform in the real-world and assess its practical usefulness using the positive predictive value.

The challenge with developing RNNs is that they require large amounts of data to avoid overfitting and maximize generalizability. Although farm-specific models that can control for between-farm variability may perform better, it would not be feasible to collect enough data to train and validate the models and as a result, robust predictive models need to be developed using data from multiple farms. Our results from Chapter 5 demonstrate that RNNs are robust to

many of the issues that would typically be encountered when training models on data from multiple farms. Model performance was only when there were very high levels of unmeasured variability at the cow- and herd-level, which can be avoided by incorporating additional information about farms and cows into data from AMS. We also demonstrated that RNNs can be very robust to outcome misclassification (label noise), allowing from the use of data from multiple farms without needing to explicitly standardize how disease diagnoses are recorded. Including predictor variables from multiple sources may also benefit model performance, in case a highly predictive variable is not recorded on a farm, and model predictions rely more heavily on the other observed variables.

The research presented in both Chapters 4 and 5 has demonstrated how methodologies that have recently become more accessible due to reduced computation costs, can effectively leverage varied data types from different sources on a dairy farm to facilitate disease monitoring and animal health management.

8.1.3. Characterizing immune response using multiplex immunoassays

As diagnostic testing kits become cheaper and more easily accessible, they provide a good way to incorporate increasingly detailed, and subject-specific physiological information into research and health management. In Chapter 6, we described how levels of blood cytokines vary, and can be used to differentiate diagnoses in children presenting to the emergency department with very similar symptoms consistent with appendicitis. Although the predictive model was not validated due to limited sample size, it did highlight biomarkers that would be of interest when developing and validating such a model for clinical implementation. With a thoroughly validated model, the primary obstacle limiting its application in a clinical setting would be the time

required to get the samples analyzed, as critical care decisions need to be made quickly. As these diagnostic kits are further developed and refined to reduce turnaround time and the complexity in sample processing, their role in precision medicine will continue to grow to support decision-making by physicians.

When combined with weekly food records, results from multiplex immunoassays were used to identify potential risk factors for flare-ups and gut and systemic inflammation in patients with Crohn's disease. Of particular importance, we found that increasing consumption of leafy green vegetables was associated with a lower risk of flare-ups (defined by elevated fecal calprotectin), and that a high ratio of Omega 6:3 fatty acids in diet can be a good way to reduce inflammation overall. One challenge with this study was that the dataset was extremely wide (no. explanatory variables >> no. observations), so multivariable regression models were used to avoid multiple comparisons with the same data but resulted in lowered power. The *P*-values were not further corrected for the multiple outcomes, as not many were statistically significant anyway, but this limitation was noted, and carefully considered when interpreting the findings.

These studies highlight the importance of all the 3 V's of Big Data for analysis – although data from a variety of sources was collected, practical application of multiplex immunoassays in decision-making was limited by the velocity at which the data could be generated in Chapter 6 and the robustness and depth of insight generated in Chapter 7 was limited by the low volume of data collected.

8.1.4. Comparing the capabilities of different analytical approaches

In Chapters 2 and 6, linear models and generalized linear models were used for exploratory analyses to assess for overdispersion and narrow down a large list of potential risk factors. The

results of these analyses were not interpreted or discussed, and rather were used to assess the suitability of the collected data for more complex models that would answer the research questions of interest. In both chapters, the statistical model used to answer the primary research questions was a generalized linear mixed model, to account for the lack of independence between observations, and the underlying hierarchical structures. Chapter 5 also utilized linear mixed models to combine data from multiple, different populations (studies) to determine a pooled effect of different treatments while accounting for the between-population heterogeneity. Model assumptions and appropriateness of model type were thoroughly explored, and the results of the model were interpreted cautiously in the context of the limitations. In Chapters 4 and 5, neural networks were trained to maximize predictive performance by freely determining the relevant variable relationships without user input. Although this resulted in good performance, describing the variable relationships was more challenging, and the interpretation more abstract than for the models used in Chapters 2, 3, 6 and 7. In Chapter 6, we developed a decision tree model to predict disease severity, which is an example of a machine learning model with variable relationships that are easily and directly interpretable. The ease of interpretation comes at a cost however, as model generalizability and external validity is diminished due to the tendency of decision trees to overfit training data. Due to the small sample size that precluded the use of an external validation set, lack of confidence about model performance and generalizability resulted in the interpretation focusing primarily on the relationships between input variables.

8.2 Future directions

In summary, future research should aim to:

1) Motivate the regular collection of data from groups of animals on dairy farms other than just the milking herd

Although data about the milking herd are likely the most important for monitoring and optimizing milk production, data should also be routinely collected from other animals in the herd. An increasing number of dairy farms in developed countries uses automated milk feeders (e.g. Barkema et al., 2015). The automated feeders collect data that are similar to the data collected by AMS. However, the data on the calves are often not used (Costa et al., 2021). By expanding regularly collected data and sampling to incorporate these other animals, producers will be able to better monitor overall health of their herd, and focus on preventive mastitis management, rather than waiting until the animal joins the milking herd to address udder health. Additionally, by monitoring drinking volume and intervals, and growth, problems in young stock can be detected earlier (Cramer and Ollivett, 2020; Conboy et al., 2021). These additional data will also benefit researchers who aim to understand the epidemiology of diseases in other groups of animals since future studies will also be able to more effectively incorporate large amounts of routinely collected data.

2) Improve predictive models for CM onset by incorporating even more data from additional sources on the farm, to reduce unmeasured variability at the herd- and animal-levels, while improving performance and generalizability simultaneously

We demonstrated that unmeasured variability at the herd- and cow-levels can reduce model performance, so being as inclusive as possible with predictor selection is preferred when using

models like neural networks. Neural networks are better able to handle noise in data from including extraneous predictors than other machine learning models, while being robust to many of the statistical issues encountered when collecting data from multiple farms. Reduced costs for high-performance computing resources will make the use of these complex models more accessible, and their potential applications for health monitoring of dairy cattle should be explored with other diseases as well.

3) Incorporate all the V's of big data (volume, variety, and velocity) in studies using multiplex immunoassays to increase the value of insights that can be generated

Although clinical applications of these assays may still be limited by complex sample preparation and turnaround time, they have immense potential to develop more targeted and personalized approaches for the study and management of disease. When aiming to use these assays for clinical prediction models, sufficient sample size to ensure robust model validation must be ensured.

8.3 References

Ankinakatte, S., E. Norberg, P. Løvendahl, D. Edwards, and S. Højsgaard. 2013. Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. Comput. Electron. Agric. 99:1–6. doi:10.1016/j.compag.2013.08.024.

Barkema, H.W., M.A.G. von Keyserlingk, J.P. Kastelic, T.J.G.M. Lam, C. Luby, J.-P. Roy, S.J. LeBlanc, G.P. Keefe, and D.F. Kelton. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. J. Dairy Sci. 98:7426–45. doi:10.3168/jds.2015-9377.

Conboy, M.H., C.B. Winder, C. Medrano-Galarza, S.J. LeBlanc, D.B. Haley, J.H.C. Costa, M.A. Steele, and D.L. Renaud. 2021. Associations between feeding behaviors collected from an automated milk feeder and disease in group-housed dairy calves in Ontario: A cross-sectional study. J. Dairy Sci. 104:10183–10193. doi:10.3168/JDS.2021-20137.

Costa, J.H.C., M.C. Cantor, and H.W. Neave. 2021. Symposium review: Precision technologies for dairy calves and management applications. J. Dairy Sci. 104:1203–1219. doi:10.3168/JDS.2019-17885.

Cramer, C., and T.L. Ollivett. 2020. Behavior assessment and applications for BRD diagnosis: preweaned dairy calves. Anim. Heal. Res. Rev. 21:188–191. doi:10.1017/S1466252320000213.

Dairy barns by type. 2020. https://www.dairyinfo.gc.ca/eng/dairy-statistics-and-market-information/farm-statistics/dairy-barns-by-type/?id=1502467060775.

Hogeveen, H., C. Kamphuis, W. Steeneveld, and H. Mollenhorst. 2010. Sensors and clinical mastitis-the quest for the perfect alert. Sensors. doi:10.3390/s100907991.

Khatun, M., P.C.C. Thomson, K.L.L. Kerrisk, N.A.A. Lyons, C.E.F.E.F. Clark, J. Molfino, and

S.C.C. García. 2018. Development of a new clinical mastitis detection method for automatic milking systems. J. Dairy Sci. 101:9385–9395. doi:10.3168/jds.2017-14310.

King, M.T.M., S.J. LeBlanc, E.A. Pajor, T.C. Wright, and T.J. DeVries. 2018. Behavior and productivity of cows milked in automated systems before diagnosis of health disorders in early lactation. J. Dairy Sci. 101:4343–4356. doi:10.3168/jds.2017-13686.

Krömker, V., and J. Friedrich. 2009. Teat canal closure in non-lactating heifers and its association with udder health in the consecutive lactation. Vet. Microbiol. 134:100–105. doi:10.1016/j.vetmic.2008.09.002.

Sepúlveda-Varas, P., K.L. Proudfoot, D.M. Weary, and M.A.G. von Keyserlingk. 2016. Changes in behaviour of dairy cows with clinical mastitis. Appl. Anim. Behav. Sci. 175:8–13. doi:10.1016/j.applanim.2014.09.022.

Siivonen, J., S. Taponen, M. Hovinen, M. Pastell, B.J. Lensink, S. Pyörälä, and L. Hänninen. 2011. Impact of acute clinical mastitis on cow behaviour. Appl. Anim. Behav. Sci. 132:101–106. doi:10.1016/j.applanim.2011.04.005.

Slob, N., C. Catal, and A. Kassahun. 2021. Application of machine learning to improve dairy farm management: A systematic literature review. Prev. Vet. Med. 187:105237. doi:10.1016/j.prevetmed.2020.105237.

Tse, C., H.W. Barkema, T.J. DeVries, J. Rushen, and E.A. Pajor. 2017. Effect of transitioning to automatic milking systems on producers' perceptions of farm management and cow health in the Canadian dairy industry. J. Dairy Sci. 100:2404–2414. doi:10.3168/jds.2016-11521.

Tse, C., H.W. Barkema, T.J. DeVries, J. Rushen, and E.A. Pajor. 2018. Impact of automatic milking systems on dairy cattle producers' reports of milking labour management, milk production and milk quality. Animal 12:2649–2656. doi:10.1017/S1751731118000654.

De Vliegher, S., L.K. Fox, S. Piepers, S. McDougall, and H.W. Barkema. 2012. Mastitis in dairy heifers: nature of the disease, potential impact, prevention, and control. J. Dairy Sci. 95:1025–1040. doi: 10.3168/jds.2010-4074.

APPENDIX: Copyright Permissions

CHAPTER 2: Naqvi, S.A., J. De Buck, S. Dufour, and H.W. Barkema. 2018. Udder health in Canadian dairy heifers during early lactation. J. Dairy Sci. 101:3233–3247. https://doi.org/10.3168/jds.2017-13579.

CHAPTER 3: Naqvi, S.A., D.B. Nobrega, P.E. Ronksley, and H.W. Barkema. 2018. Invited review: Effectiveness of precalving treatment on postcalving udder health in nulliparous dairy heifers: A systematic review and meta-analysis. J. Dairy Sci. 101:4707–4728. https://doi.org/10.3168/jds.2017-14301.

Original Materials

Data (including graphs, figures, tables, and illustrations) must not have been published elsewhere except as abstracts, local or regional field day reports, extension letters, or non-peer-reviewed, noncopyrighted proceedings of conferences. Authors may post abstracts of manuscripts on the web at the time of submission; uploads to preprint servers (e.g., bioRxiv, Research Square) are not permitted. Manuscripts with a digital object identifier (DOI) assigned and publicly available are considered published and should not be submitted to JDS. Material submitted to JDS should not be submitted for publication to popular magazines, company advertisements, or organizational proceedings until the author has received notification of acceptance of the manuscript. Before manuscripts are submitted, authors should have them read critically by others well versed in English to facilitate review, and the senior author must have authorization to publish. All coauthors should approve the manuscript before its submission to the journal.

Reproducing Published Material in Your Article

Figures, tables, or other parts of a manuscript that have previously been published or are not the property of the authors must be properly acknowledged in the manuscript. The authors must obtain permission to republish these items from the person or institution holding the copyright, usually the publisher, and provide those permissions to the journal office.

Requests to reproduce material published in JDS must be made through Elsevier's Rights Department (permissionshelpdesk@elsevier.com), online via the Elsevier permissions page (https://www.elsevier.com/about/our-business/policies/copyright/permissions), or via the Copyright Clearance Center (https://www.copyright.com).

Copyright Agreement

A copyright agreement is collected online from each author, after a paper is accepted for publication. Authors who are not permitted to release copyright must complete the form and provide a statement of the reason for not releasing the copyright.

The Association grants to the authors the right of republication of their own material in any book, thesis, or dissertation of which they are authors or editors subject only to giving proper credit in the book to the original JDS publication and only for non-commercial use. In addition, authors may post abstracts of manuscripts on the web at the time of submission. Once an author receives notification of acceptance, the peer-reviewed, pre-typesetting manuscript can be posted to the author's website.

Articles for Deposit and Author Self-Archiving

Authors may deposit their peer-reviewed (pre-typeset, pre-editing) manuscript into a non-commercial institutional repository 12 months after publication of their paper in an issue of the journal. Manuscripts may not be uploaded to ResearchGate or SciHub. For more information on depositing and sharing your JDS manuscript or article, see Share your JDS Article.

Open Access (CC By-NC-ND)

To publish an article under an Open Access (OA) CC By-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/), authors should contact the editorial office when their paper is accepted for publication. A separate licensing form and fee agreement will be sent to the corresponding author of the paper for completion. For authors who choose to pay the OA fee instead of page charges, their paper will become freely available upon publication in an online issue.

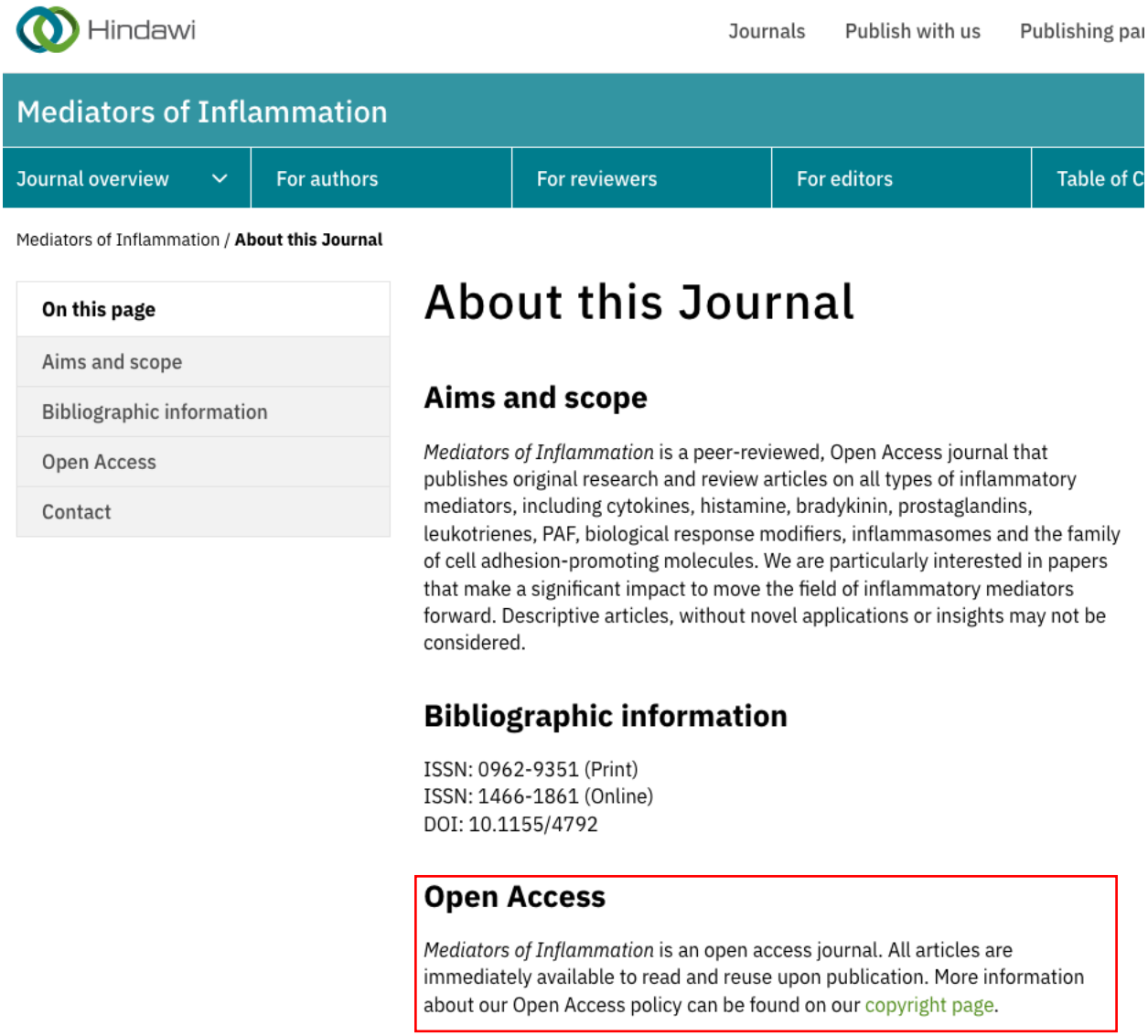
OA Articles for Deposit

Author(s) publishing articles under OA shall bear sole responsibility for meeting the specific posting requirements of their funders. Upon payment of the OA fee, authors may deposit the accepted (peer-reviewed pre-typeset only) manuscript in a non-commercial institutional repository. Deposited articles are embargoed for 12 months after publication in a journal issue before being released in the repository.

REVIEW OF MANUSCRIPTS

Upon submission to JDS, a manuscript is assigned to a section editor, who enlists reviewers to assist in the evaluation of the manuscript. The review process is confidential, which infers a bond of trust among the authors, editor, and reviewers. The editor is trustee of the manuscript until the review process is completed and ensures that the review process is fair, thorough, and confidential. Reviewers are asked not to share the contents of the manuscript with anyone, except that they may ask a colleague to assist with the review with approval of the editor. Communication with authors should only be through the editor. Reviewers should

CHAPTER 6: Naqvi, S.A., G.C. Thompson, A.R. Joffe, J. Blackwood, D.-A. Martin, M. Brindle, H.W. Barkema, and C.N. Jenne. 2019. Cytokines and chemokines in pediatric appendicitis: a multiplex analysis of inflammatory protein mediators. Mediators Inflamm. 2019. https://doi.org/10.1155/2019/2359681.



CHAPTER 7: Naqvi, S.A., L.M. Taylor, R. Panaccione, S. Ghosh, H.W. Barkema, N. Hotte, N. Shommu, S. Kaur, R.A. Reimer, K.L. Madsen, and M. Raman. 2021. Dietary patterns, food groups and nutrients in Crohn's disease: associations with gut and systemic inflammation. Sci. Rep. 11:1–9. https://doi.org/10.1038/s41598-020-80924-8.

License agreement and author copyright

Scientific Reports does not require authors to assign copyright of their published original research papers to the journal. Articles are published under a CC BY license (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license, users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor (read full legal code).

Visit our open research site for more information about Creative Commons licensing.