Optimising the use of new data streams for making epidemiological inferences in veterinary epidemiology

A thesis presented in partial fulfilment of the
requirements for the degree of

PhD
in
Veterinary Epidemiology

at Massey University, Manawatu, New Zealand

Arata Hidano

2018

Abstract

Many ‘big data’ streams have recently become available in animal health disciplines. While these data may be able to provide valuable epidemiological information, researchers are at risk of making erroneous inferences if limitations in these data are overlooked. This thesis focused on understanding the better use of two data streams—livestock movement records and genetic sequence data.

The first study analysed national dairy cattle movement data in New Zealand to explore whether regionalisation of the country based on bovine tuberculosis risk influenced trade decisions. The results suggested that the observed livestock movement patterns could be explained by the majority of, but not all, farmers avoiding purchasing cattle from high disease risk areas. The second study took an alternative approach—qualitative interviews—to understanding farmers’ livestock purchasing practices. This study suggested that farmers are not necessarily concerned with disease status of source farms and that it may be the reliance on stock agents to facilitate trade that creates the observed livestock movement patterns in New Zealand. The findings from this study also implied that various demographic and production characteristics of animals may influence farmers’ livestock selling practices, which were quantitatively verified in the third study analysing livestock movement data and animal production data. These studies not only showed that analyses based solely on ‘big data’ can be misleading but also provided useful information necessary to predict future livestock movement patterns. The final study evaluated the performance of various genetic sequence sampling strategies in making phylodynamic inferences. We showed that using all available genetic samples can be not only computationally expensive, but also may lead to erroneous inferences. The results also suggested that strategies for sampling genetic sequences for phylodynamic analyses may need to be tailored based on epidemiological characteristics of each epidemic.

Acknowledgments

I would like to take advantage of this opportunity to express my very personal belief. It has been increasingly recognised that many postgraduate students suffer from poor mental health. Fortunately, I have never had any serious mental breakdown in my last four years. In fact, I have had an amazing time, doing what I like for most of the time. There are several reasons that I can think have contributed to my mental health, but I have one important message for you if you are doing PhD, reading my thesis for some interesting reasons, and being somewhat interested in potentially improving the quality of everyday PhD life.

It's quite simple. Please stand up from your chair, go out, and do some exercises. Any exercise will do—hard or soft, short or long, run or walk, weight training or swimming—whatever you find it fun because you need to continue doing it. Do not be afraid of leaving office and doing something else! Your 30min or one hour exercise per day will never delay your PhD completion. A PhD period may last four or five years, but our life is likely to go on for longer. Finishing PhD is important but being a healthy human is much more important. Our life is much more than about PhD.

I may sound so pushy and arrogant, but please bear with me a bit more. The positive impact an exercise brings on our health, not only physical but immensely on mental, is huge. A good sweat releases endorphins in your brain, which gives you a strength of overcoming a hard time to revise your manuscript with endless comments and red lines, finish a demanding lab work, or cope with sometimes sensitive human relationships inherent to being a PhD student. Please consider your mental health seriously, please! It's nothing to feel shamed, we as a science community, should talk about our mental health more openly. I strongly believe healthy science can be more likely to be generated from physically and mentally healthy scientists. And PhD days are the right time to start taking care of our health.

As I said, the last four years of my PhD journey were incredible one. It was a life changing, broadened my views, and what is more, very enjoyable. My former main supervisor Professor Tim Carpenter told us from time to time that his PhD time was one of the best time in his life, pursuing what he enjoyed with much freedom. I always kept this word in me and tried to achieve it. And I achieved it. I have been so fortunate to have funders and a PhD project that allowed me to do so. I spent a lot of time reading something that did not seem to be directly relevant to my PhD. In the hindsight, they were all useful and they contributed to shaping the final form of my PhD project, which had needed to change many times for different reasons. Who can define what is relevant and irrelevant clearly? I know that in reality there are many factors that prohibit us from doing something that look irrelevant—money, family, and most importantly support from your mentors and colleagues. I am therefore so much grateful to my great mentors.

I came Massey to work with Professor Mark Stevenson. I respect his authenticity and a sincere attitude towards science and epidemiology. He is a true educator, and his passion is contagious. Even though he left to Melbourne just prior to my PhD start, what I have learnt from him is and will continue to be my backbone as an epidemiologist. Mark, I cannot thank you enough for guiding me and being excited about exploring new methods, with spending so much of your time. I will never forget those hard days trying to match messy data from Bhutan which had slightly different dialect names in different datasets for the same town. Then you wanted me more to estimate the number of populations in each town using information that I had to manually input from somewhere else! It was a really draining work and I spent nearly half of my first two years of PhD working on non-PhD works. But that set my standard for research. I am proud of what we have done for these projects, and any other works I have done since then.

Professor Tim Carpenter has given me wisdom and a courage to stay mentally free. I have been truly inspired by his attitude to enjoy life and turn a hardship into a joy. He has been always a positive thinker. He taught me how to work best as a team. Now being back home at Davis, he still works tirelessly. Thank you Tim for always encouraging me to get out of a comfort zone and take a risk of getting into new situations—workshops for new area, conferences with unknown people, and trying to use new methods.

Dr Carolyn Gates, my last main supervisor, brought me a whole set of new skills and way of thinking. They were hands-on skills that I have been so longing to acquire in my PhD. I am so grateful to you to be always patient and helpful. Your way of thinking, a big picture thinking, is something I was totally lacking when I started my PhD. I learnt so much about making (but not making up) a good story for writing, putting your research in a bigger context, and thinking forward and for something good. I am perhaps not the easiest student and having me as the first PhD student may not have been the easiest task for you. I do appreciate what you have done for me for the last three years, Carolyn. All these three my great mentors always have allowed me to pursue my interests. This was a great treasure I have been given. Thank you so much.

I have had great collaborators and advisers through my PhD. Dr Gareth Enticott at Cardiff University has opened up a new world of human behaviour to me. He has been always passionate and never hesitated to share his great knowledge. It has been so fascinating to find what researches in other fields have done, and at the same time interesting to see the tremendous lack of communications between science fields, leading to repeating redundant studies. “Find what you can do and what you are good at”—this is the best advice I have had in my PhD, which came from Professor Javier Guitian at Royal Veterinary College, a friend of mine and an epidemiologist I respect. When I was not sure what I should pursue under the pile of too many literature in the archive, he reminded me of what veterinary epidemiologist can do for the world. Thanks Javier, your advice worked like a lighthouse in the dark night. Professor Rob Christley has inspired me in many ways—by me reading his paper, being exposed to his broad thinking, and how he has worked for our collaboration paper. Dr. Chris Jewell, thank you so much for spending your precious time for nothing. I am so regretful that we could not get funding that could allow us to work together officially. Nevertheless, you spent a whole two weeks for me. Your knowledge and skills were so vital for developing the simulation model used in my thesis. I hope we can work together in future for something in New Zealand! Professor Ian Dohoo, how can I describe how much your workshops have educated me as a veterinary epidemiologist? I learnt so much in every workshop you taught us. You always answered my questions with serious thoughts and gave me great suggestions.

I thank so much to Dr Andrew Conlan, Dr TJ McKinley, and all other tutors from Wellcome Trust workshop on Mathematical Models for Infectious Disease Dynamics. This workshop was no doubt one of the best workshops I have ever attended. This taught me so much about infectious disease modelling and without this I could never have developed the simulation model I used. Thank you for organising an excellent workshop and thanks for financially supporting my attendance too. I am also grateful to Professor Alexi Drummond, Dr Tim Vaughan, and all other tutors from Taming the BEAST in the South Pacific at Waiheke Island. This workshop gave me knowledge and skills that made me comfortable to work on the study in Chapter 6. Professor Dirk Pfeiffer, Dr Marisa Peyre, Dr Denis Haine, Dr Anna Alba Casals, Professor Crawford Revie, Dr Simon Fireston, and Professor Ian Gardener—they inspired me in many different ways, through friendship, authenticity, passion, clarity of teaching, broad views, wisdom, and tireless help.

Professor Cord Heuer, thank you for your friendship and providing us with enjoyable lunch time chat over the last four years. You said hard work is a virtue for German, similar to Japanese, but your hard work has truly inspired me. Your attitude to do ‘right things’ as a scientist was very important for me to know in the last stage of my PhD. I am looking forward to working with you more in soon future. Naomi, Art, Chris, Emilie, thank you for managing EpiCentre, which I imagine I do not fully know how much you have been putting efforts to maintain the project and this group. Christine, thank you so much for sorting out all financial things and keeping the peace at EpiCentre. Simon and Masood, thank you for your technical supports and creating that good time we all together played frisbee. Sarah, although short time, thank you for being my co-supervisor.

Past and current EpiCentre gangs, I could not be here without you all. Juan, you are so matured as human and I don’t know how many times you inspired me. Together with Dani, you guys always took care of us. Thank you for all the parties, BBQs, and Piscos. But also, thank you for stimulating discussions on epidemiology things. I always considered you as a comrade who loves classic epidemiology. Nelly, you always have provided us with fun and chaos, which eventually I enjoyed in the hindsight. I truly appreciate discussions we had about network analysis and other epidemiological issues, from which I learnt a lot of things. Thank you for making great cakes always. Your perfection and elaboration always made our parties so good. Together with Sebou, we had so much fun over the last six years! Masako, I am still so grateful that you let me use your bikes, although it was not the best one, when I came to Palmerston North for the first time in 2011. I think me and you provided some fun by us always giving different opinions about some Japanese culture and stuffs, although I still believe I am right. It was relieving and encouraging for me to know that somebody from Hokkaido University was doing PhD. EuTteum, thanks for putting up with me in the same office when you were finishing your PhD. I can never forget that you were eating frozen pies everyday, having made me thought how hard the life would be in the last stage of PhD. I think I have manged to eat something better though. Thank you for taking us to fishing and sharing Tui with nice Sashimi. Jun, it has been so amusing to see how your life has changed over the last four years. Nevertheless, thank you for always being there to listen to my obsession stories and giving us laughs, including your dance which we miss a lot. Also I appreciate your endless questions. Those sometimes really have made me realise I don’t know or understand enough my research, which eventually has improved my work. We have done so many things together—

fishing, trip, drinking, hang-over, surfing. Sabrina, you are like a mum for us. I mean you take care of everyone at EpiCentre! Thank you for always organising something fun and willing to do something that people don't want to do. Your devotion is truly impressive. Jun and Sabrina, thank you for sharing pains sometimes. And I am proud of you two for having continuously pushed your mental and physical limits. Alicia, thanks for sharing some ridiculous fun with us. Your excitement over something has been often hilarious, which provides energies to EpiCentre. I won't forget all these things in Vinegar Hills, which still makes me laugh! Milan, we have known each other for such a long time since 2011! I still remember you and Prakriti invited me for dinner. You are always such a gentleman and inspirational human, Milan. I am proud that our journey has come so far! I hope you all the best! Katja, you are another great event organiser. Thanks for bringing us together. Yuni, your balancing of work and fun (a.k.a badminton) has always reminded me of the importance of having some fun. Am, I always enjoyed our small talks. Your smiles cheered up us all. Catlin, thanks for giving me a ride from Hamilton. I am looking forward to working with you more! Ben, thank you for always having made the office atmosphere much better with your laughs!

Thank you Drs. Naomi Cogger, Simon Fireston, and Pauline Ezanno for generously spending your time to read my thesis and provide me valuable comments and suggestions. I just cannot imagine how on earth you can spend a lot of time to read this thesis in detail. I really appreciate your kindness.

I also thank for our extended colleagues and friends from Hopkirk and other parts. Springer, it was so nice to attend Taming the BEAST workshop with you. Thank you for never failing to invite us for your parties. It's so amazing what you have done for us. I am so glad that I could make it to your engagement party. Reed, you also brought so much fun to us with your love for whisky and beer. Marie, your French humour was often confusing and amusing but your serious thoughts on many things stimulated my thinking. Terrance, you further blended some Kiwi humour to Marie's, which was always nice to taste. Kruno, thanks for your friendship although it had been hard time for you to finish PhD. Ali, Sara, Felipe, Maca, you are all gone to somewhere in this world but I will remember the good time we had together. Guilhem, thank you for being fun and providing some wisdom. All the surf trips we had are gem memories. Jason, although you stayed for a short time, you are such a lovely person and I am so happy to meet you. Diana, you are like a comrade with who I tried to finish PhD. Thank you for being happy despite the hard time in your PhD. You can soon feel so good again, so hang in there! Martin, Ivayla, Kurt, thank you always for trying to make Palmy better place. Nicola, I think your music and generosity will save the world. Lena, it was so good to have you at Palmy! Omar, Paul, Brian, thank you for your friendship! Alexis and Olivia, you guys really make me happy. Sarie, you added some eccentric spice to the life of EpiCentre, which I always enjoyed. Gemma, thank you for stimulating me by your extreme hard works and love for your work! You too Aaron, although sometimes extreme, your passion towards Stats ignited interesting discussions, which I appreciate. I am sure I have missed some names here—thank you so much for being nice and spending good time with me!!!

Chifuyu, thank you so much for introducing GRIT to me. It literally has changed my life. In the hindsight, your positive attitude at the end of your PhD period, when you had loads of difficulties in Stats, is so impressive. Your passion for conserving this world has been always inspiring. Lynsey, thank you for helping my English study and transcribing my interviews with farmers. I am grateful to the conversations we had, small or big, serious or hilarious. Your deep thoughts have sometimes shocked me in a good sense. We had a great time and I hope your life continues to be good and that your goal to make this country better will be achieved. Aniek, thank you for being such a great ambassador for PhD students. You have done a fabulous job to organise workshops, bring PhD students together, and spreading positive attitudes to us. I am truly grateful to you. Kandarp, or shall I call you Pintu, thank you for being such a nice office mate and friend. I will never forget the time at your wedding and the food poisoning I had there. It was so moving for me that you got so worried about Nelly and try to help her when she was in the hardest time of PhD. Anou and Lesley, thank you for taking care of me as a big sister. Your calm passions and wisdom towards life really inspired me. Hannah, thank you for being a lovely flatmate, who always took care of the house properly. Your stick-to-routine attitude is something I don't have and it has been always motivating. Claudia and Eri, thank you for being a nice temporal dweller and giving us a great time. Petra, thank you so much for being an amazing landlord and friend, for all your support, care, love and friendship over the last four years. Panther, I know you are not interested in my thesis, but I love you. Thank you so much for giving me all the cuddles and heals. You are the sweetest black cat I know.

My wholehearted appreciation goes to Massey Recreation Centre crews, especially GRIT team, Janine, Laura, Tessa, Mel, and Dean, and our manager Orla. My PhD journey would be never ever pleasant as it has been in the last three years without a group fitness class at the Rec Centre. You gave me opportunities to realise that overcoming a physical challenge is about overcoming your mental barrier. When I realised this, my life has changed since then. We cannot control everything but we realise there are things that we can certainly change. This realisation leads to make us think 'OK, what can I do in this situation?', which has helped my life in general and PhD days. Being a GRIT and ZUU instructor has coloured my everyday life incredibly beautifully. To Janine, Laura, and Tessa, thank you for being such an amazing mentor! You built an initial step for me to become a better group fitness instructor. I cannot thank you enough for you being generous for your time and effort to teach me what I had to learn. Body balance instructors, especially Toni and Rebecca, thank you for delivering amazing classes. Your class really calmed me down, which I often needed when I get too switched on after getting too excited about work and exercises. I truly love your class.

Special thanks go to my family. My father, who has been always incredibly open-minded, has supported me mentally and physically for the last decade when I struggled to make me understood what I want to do. Thank you for always enlightening me with your wisdom for science and life. I am truly grateful to you for what you have done. My grandfather, now turning into his 98, has long waited my completion of PhD. I am glad I could finally make it. Thank you for always being excited about listening to my research interest and reading my publications. I have been so inspired by your research and work ethics, and wills to do good for this world. I am so proud of you. My sister, thank you for being an amazing sister. Your curiosity towards this world and footwork going everywhere have literally raised me up. My mother, I assume it has been difficult time for you that your son is so stubborn and moving everywhere around this world doing something called epidemiology, which you think is something related to immunology. I hope I have not had given you too much hassle but I would appreciate it if you could see what I have achieved in the last four years. To my late grandmother, Kimiko, thank you for always being happy about what I have been doing in the last 10 years. Please feel relieved, I am done now! To my grandmother in Yamaguchi, I am grateful to you for encouraging me to walk the way I believe.

Lastly, I would like to convey my deepest appreciation to all the funders who supported my PhD. Thank you JASSO and administration staffs there for providing me with scholarship for the last four years. Without this, I could never be able to start my PhD. Thank you for staffs at Hokkaido University for managing JASSO scholarship for me. I am so indebted to all your hard work for checking my application documents for twice to ensure there was no mistake. My appreciation also goes to MAFF and NIAH Japan for providing project funding, which immensely helped my PhD work. Particularly, Drs. Toshiyuki Tsutsui, Takehisa Yamamoto, Yoko Hayama, Sota Kobayashi for managing this project and all the supports you have given me in the last four years. Thank you so much for your patience. Also thanks goes to staffs at NIAH and Massey University, especially Ms. Hiromi Shimura, Mika Mori, Susie Bronlund, for doing all the administration works for this project. I cannot thank you enough to Kathleen Spragg Agricultural Trust for supporting my PhD project. Truly, I could never have had all these workshop and conference opportunities without this funding. Thank you for believing and trusting me. To School of Veterinary Science, formerly IVABS, and Massey University, thank you for financially supporting for trips to conferences. Thanks Debbie Hill for managing all administrations, which have been massively helpful to carry on PhD. Thank you to GRS, especially Dr. Julia Rayner, for organising great postgraduate events. Having done Clifton Strength workshop was really useful for me to perform better in my PhD and know myself more. Thank you A/Professor Tracy Riley for initiating such a wonderful writing Bootcamp and all the supports for PhD students. I know there are far more people at all these institutions mentioned that have worked for this project to happen. I thank you all sincerely, I really have enjoyed my PhD and achieved my goal, with which I hope I can return something good to New Zealand, Japan, and this world and society. To conclude my gratitude, thank you Melanie for your cheer, smiles, support and love.

List of publications

Hidano, A., Carpenter, T.E., Stevenson, M.A., Gates, M.C., 2016. Evaluating the efficacy of regionalisation in limiting high-risk livestock trade movements. Preventive Veterinary Medicine 133, 31–41. https://doi.org/10.1016/j.prevetmed.2016.09.015 (Chapter 3)

Hidano, A., Enticott, G., Christley, R.M., Gates, M.C., 2018. Modelling dynamic human behavioural changes in animal disease models: challenges and opportunities for addressing bias. Front. Vet. Sci. 5. https://doi.org/10.3389/fvets.2018.00137 (Chapter 2)

Hidano, A., Gates, M.C., 2018a. Why sold, not culled? Analysing farm and animal characteristics associated with livestock selling practices. Under review in Preventive Veterinary Medicine. (Chapter 5)

Hidano, A., Gates, M.C., 2018b. Influence of population sampling strategies on making phylodynamic inferences for livestock diseases. Under review in PLoS Computational Biology. (Chapter 6)

List of conference presentations

Hidano, A., 2018. Farmer behaviour and veterinary epidemiology: challenge and opportunity. Presented at the New Zealand Veterinary Association conference, Hamilton, New Zealand.

Hidano, A., Carpenter, T.E., Gates, M.C., Stevenson, M.A., Enticott, G., 2017a. "Big data" and "small data": towards a better understanding of disease data by modelling human behaviour, in: 3rd International Conference on Animal Health Surveillance. Presented at the 3rd International Conference on Animal Health Surveillance, Rotorua, New Zealand.

Hidano, A., Gates, M.C., 2018. Human behaviour and livestock disease spread: challenges and opportunities in modelling dynamic feedback loop in disease simulation model. Presented at the The Australian and New Zealand college of Veterinary Scientists Science Week Conference, Gold Coast, Australia.

Hidano, A., Gates, M.C., Carpenter, T.E., Enticott, G., 2017b. Farmer decision-making around livestock trading practices: the multifactorial roles of stock agents, in: Inaugural Meeting of ISESSA 2017. Presented at the International Society for Economics and Social Sciences of Animal Health, International Society for Economics and Social Sciences of Animal Health, Aviemore, Scotland, p. 91.

Hidano, A., Gates, M.C., Stevenson, M.A., Bosson, M.A., Carpenter, T.E., 2015. Influence of disease control zones on dairy cattle movement patterns. Presented at the The 14th International Symposium of Veterinary Epidemiology and Economics, Merida, Mexico.

Hidano, A., Gates, M.C., Vaughan, T.G., 2017c. Towards an unbiased phylodynamics inference in livestock disease systems: A validation framework using an individual-based model. Presented at the Epidemics 6, Sitges, Spain.

Contents

List of Figures

List of tables

Chapter 1

Introduction

1.1. Background

In this increasingly connected world, infectious diseases can spread long distances over short time periods. During the last decade, we have observed local epidemics can suddenly become global threats as in the case of SARS, Ebola, and foot-and-mouth disease. In order to implement an appropriate disease control strategy, policy-makers need to have good information on population demographics and risk factors that determine how disease spreads in the population. Traditional epidemiological studies have long served to provide this information. However, there is a growing interest in using 'big data' since it often captures a wider population base. Indeed, the emergence of 'big data' has been claimed to transform science in many disciplines (Baraniuk, 2011; Bell et al., 2009; Lazer et al., 2009). Big data relevant to epidemiology include electronic health records, surveillance data, digital traces in social media, cell-phone logs, genetic sequence data and many more (Bansal et al., 2016; Mooney et al., 2015). Data analogous to these exist in the veterinary epidemiology discipline as well including animal health and production records, livestock movement records, and genetic sequences of animal pathogens.

These data have a significant potential to provide epidemiological information that was not traditionally available or difficult to collect. For instance, call data records from mobile phones provide unprecedentedly high resolution data of individual human mobility; whenever an individual makes a phone call or sends a text, his or her approximate geographical location can be identified using the location of mobile phone tower as a proxy (Wesolowski et al., 2016). This data can be in turn used to represent human contact patterns, which are essential components of a mathematical and simulation model for infectious diseases. Traditionally, human contact data were collected through a questionnaire survey, collecting individuals' demographic information as well as information regarding whom they had contacts (Edmunds et al., 1997; Mossong et al., 2008). These traditional methods are costly and have various limitations such as recall bias, where 'stronger' connections are more likely to be recalled and reported (Eames et al., 2015). Similar to human contact data, livestock contact data were traditionally collected through a questionnaire or a diary survey (Bates et al., 2001; Sanson, 2005), which is labour intensive and holds the same limitations as traditional human studies. Over the last few decades, a national-scale animal traceability database has been established in many countries, which allowed much larger scale livestock contact data to be retrieved.

Pathogens' genetic sequence data, in particular whole-genome sequence (WGS) data, are another pivotal example of big data. Although molecular epidemiology is not a new concept (Hall, 1996), the advent of WGS technology enables us to describe the genetic relationships between sequence samples in a high resolution. Combined with epidemiological information, such a high resolution allows us to estimate a transmission chain of infections, describe host contact networks, and identify a potential reservoir of emerging pathogens (Kao et al., 2014). Coined by Grenfell and colleagues, phylodynamics is an emerging field to better understand the relationship between epidemiological processes and pathogen evolution by integrating many different disciplines including epidemiology, evolutional biology, and immunodynamics (Grenfell et al., 2004). The development of user-friendly software that allows us to access sophisticated analytical methods make phylodynamic analysis popular in veterinary epidemiology.

As pointed out by many researchers, however, big data are not silver bullets that can completely replace traditional data collection and analysis (Bansal et al., 2016; Chen et al., 2014; Gates et al., 2015; Lazer et al., 2014). Big data are often biased and noisy. Moreover, big data analysis is often conducted for the purpose that is not originally data were collected for. Here, there are two layers of bias; one is the bias in big data itself, and the other is the bias we researchers bring into science by the way we use data. The former bias may be partially addressed by employing novel statistical methods, which has been discussed elsewhere (Khoury et al., 2013; VanderWaal et al., 2017). With the big data deluge many studies have focused on simply using these data, however, there are much fewer considerations as to the way we apply these data to make a better epidemiological inference. The central theme of this thesis is to fill this gap.

Amongst many different data types, we identified two key data streams that are particularly important in veterinary epidemiology; livestock movement data and genetic sequence data. As described above, the former plays a key role in a livestock disease simulation model, however, the way we currently use livestock movement data is surprisingly limited. On the other hand, there are growing applications of the latter to a phylodynamic analysis. Nevertheless, very few studies discuss the validity of these applications in veterinary epidemiology. In the remaining of this chapter, we provide a brief background of how these two data have been used in this field. Then, the structure of this thesis is described.

1.2. Use of livestock movement data

Given livestock movement is an important transmission mechanism for many economically important livestock diseases, a large number of studies analysed these data over the last decade. Notably, social network analysis has been applied in many countries to describe the statistical properties of movement networks (Dorjee et al., 2013; Dubé et al., 2010; Marquetoux et al., 2016). These studies provided important insights such as a strong temporal and spatial heterogeneity in movement patterns (Bajardi et al., 2011; Ensoy et al., 2014), some consistencies across years (Grisi-Filho et al., 2013; Valdano et al., 2015), and how various farm characteristics are associated with their contact patterns (Valdes-Donoso et al., 2017).

Many social network studies have focused on quantifying various global characteristics of networks such as clustering coefficient, degree distribution, and assortativity (Dubé et al., 2009; Martínez-López et al., 2009). We have also gained good understanding on how these global characteristics influence disease spread patterns (Christley et al., 2005; Moslonka-Lefebvre et al., 2009). However, it has become evident that these global characteristics may not be good measures to describe livestock movement network because, for instance, networks with the same degree distributions and cluster coefficients can show very different higher-order structure, where disease spreads in a completely different manner (Ritchie et al., 2014).

Other studies also described node (farm) level characteristics such as centrality and betweenness because these measures were assumed to be good indicators for a target of disease surveillance (Dubé et al., 2009). However, these network characteristics are usually generated on a snapshot of the network—static network—and studies have shown that these characteristics change over time (Berger-Wolf and Saia, 2006; Uddin and Hossain, 2011). Moreover, another disease simulation study showed that disease spread patterns are highly influenced by its initial condition such as the location of the first infected farm (Bajardi et al., 2012). Given that not all farms have the same probability to be the initial infected farm, such network characteristics do not truly represent the probability of these farms being infected in a disease outbreak, undermining the usefulness of these network measures.

What is largely missing in the current literature is to use livestock movement records to understand farmers' livestock trading behaviours—livestock movement is fundamentally a result of farmers' livestock selling and purchasing practice. As recognised in many other disciplines, big data provide a great opportunity to identify dynamic human behavioural patterns (Bansal et al., 2016; King, 2011). Moreover, understanding farmers' trading behaviours is particularly important because there is evidence that farmers change their behaviours in response to many factors such as disease dynamics and new animal health legislations (Vernon and Keeling, 2012). This, in turn, significantly influences how disease spreads (Vernon and Keeling, 2012).

Nevertheless, many livestock disease simulation models do not account for farmers' dynamic behavioural changes and assume they remain constant over time.

1.3. Use of genetic data

The use of genetic data is not new—molecular epidemiology as a discipline has existed for a few decades. Molecular epidemiology is defined as a discipline that uses molecular typing methods for infectious pathogens in the study of the distribution, dynamics, and determinants of health and disease in population (Field et al., 2014; Hall, 1996). The advent of WGS, however, enables us to identify substitutions in nucleotide in an unprecedentedly precise manner. This high resolution, in turn, allows us to quantify the relationships between genetic samples using an evolutional analysis framework. This is the fundamental idea of Bayesian evolutional analysis that has been emerging in epidemiology disciplines. There are three key concepts that need to be highlighted to understand evolutional analyses; phylogenetics and population genetics, substitution models, and molecular clock hypothesis.

1.3.1. Phylogenetics and population genetics

There are two traditional approaches to describe the evolutionary history of nucleotide sequences; one is phylogenetics and the other is population genetics. These two share some similarities—both use trees—but the fundamental concept is very different. While phylogenetics is interested in determining the topology of tree that describes the evolutionary relationship between sequences, population genetics often aims to identify factors that are associated with the observed genetic diversity. Population genetics still rely on a tree to quantify the relationship between sequences, however, the tree topology is usually treated as a nuisance parameter. This distinction is rather important because many traditional molecular studies have ignored the uncertainty around tree topology given genetic data—many different tree topologies can result in the same genetic diversity observed in the data (Lemey et al., 2009). Population genetics is therefore the preferred approach if evolutionary parameters are of interest because the uncertainty in the tree topology is integrated in the analysis. Bayesian evolutional analysis used in epidemiology therefore relies on population genetics approach.

Regardless of which approaches are used to construct a tree, its topology is characterised by branching order of the tree and branch length of the tree. The branch length from a common ancestor represents the expected amount of genetic mutations that occurred since their divergence. Below, we briefly summarise how tree is constructed in phylogenetics and population genetics.

1.3.1.1. Phylogenetic tree

There are various methodologies to construct a phylogenetic tree. These methods are broadly divided into two categories; discrete-character method and distance-based matrix method. The former includes maximum parsimony approach, maximum likelihood (Felsenstein, 1981) or Bayesian framework (Rannala and Yang, 1996) which generally searches the optimal topology that fits the data based on some criteria (e.g. minimum substitution or likelihood). The latter includes Neighbour-Joining methods and Unweighted Pair-Group method with arithmetic mean (UPGMA), and these methods use a certain algorithm to construct a tree by sequentially connecting nodes. The major difference of these two categories is that while the former handles sequence data as sequence of nucleotide characters, the latter only considers the genetic distance between sequences (i.e. the fraction of sites that differ between two sequences). Maximum parsimony method seeks a tree that requires the fewest number of changes in nucleotide to explain the observed sequence data. Maximum likelihood method constructs a likelihood function based on the probability of observing a specific alignment pattern of nucleotide (i.e. a column of all sample sequences) at each site and branch lengths, then searches a tree that maximises the likelihood function. Similarly, Bayesian approach constructs a likelihood function and seeks a set of best fit trees using Markov Chain Monte Carlo (MCMC) simulation, often with Metropolis-Hasting sampling approach. The latter two is usually called model-based approach and basically seeks a “best” tree topology by adding one branch at a time in the topology. The enormous numbers of potential tree topologies, however, usually prohibit an exhaustive search and some heuristic algorithm is applied. It is important to note that these model-based approaches aim to construct a tree by sequentially adding branches as opposed to explicitly modelling genealogical relationship between samples, which is the common practice in population genetics framework as discussed later.

There has been an ongoing debate regarding the selection of a phylogenetic tree construction method; the detailed descriptions about each method as well as its advantages and disadvantages are beyond the scope of this paper and readers are referred to more comprehensive discussions (Whelan et al., 2001; Holder and Lewis, 2003; Steel, 2005; Simmons, 2014). Various cutting edge approaches that can construct a phylogenetic tree for a large number of samples with less computational time have been also proposed (Harris et al., 2010; Bertels et al., 2014). Regardless of methods used, it is important to bear in mind that the homologous sites of genes or amino acids should be compared to provide a meaningful tree. For this reason, aligning the sequences in an appropriate manner is a crucial step in phylogeny construction (Lemey et al., 2009).

1.3.1.2 Genealogical tree

As highlighted above population genetics approach focuses on modelling the demographic of genetic population and hence stochastic genealogical patterns, whereas phylogenetic approach models the random gene substitutions but not underlying population demographic. More practically speaking, population genetics approach first explicitly creates all possible random genealogies using a certain model we discuss below. For each constructed genealogy, mutations are then assumed to occur at certain probability and when it occurs a site is randomly chosen where a mutation happens according to specified substitution model we discussed in Section 1.3.2. Recall traditional phylogenetic approaches do not explicitly account for genealogical relationships between sequences. There are two popular models that create genealogical relationships between sequences; coalescent model and birth-death model. We briefly summarise these two models and their limitations below.

1.3.1.2.1 Coalescent model

The coalescent model was first formulated by Kingman as a mathematical model to randomly create genealogies backward in time (Kingman, 1982). In the most basic coalescent, n sampled individuals from the current population are traced back along their descendent line until their most recent common ancestor (MRCA) (Aldous, 2001). The coalescent model describes a likelihood function of the observed data by a function of various population demographic parameters such as migration rate and population growth rate, thus allowing these parameters to be estimated. As we have discussed above, genealogy-based models construct stochastic genealogies first, and then model the genetic mutations that took place along the branches of this tree using one of substitution models described later (Rosenberg and Nordborg, 2002). We briefly introduce the concept of the coalescent model under the simplest case of idealised Wright-Fisher population. Here we consider the haploid population of the constant size N under the neutral selection; that is, only a very small fraction of new mutations are selectively advantageous (Hein et al., 2005).

The traditional coalescent model randomly selects a pair of individuals and merges them into a single ancestor from the present (generation *t0*) to the past (generation *t1*, *t2*,...). Then, the probability of two individuals randomly selected from the population sharing a common ancestor at generation *t1* is 1/N and the probability of not finding a common ancestor is 1 – 1/N. This is because the first individual selects its parent freely out of N population that existed in generation *t1* and the second individual has to choose the same parent. This can be easily extended to describe the probability that any two of n (n<<N) randomly selected individuals from the population share their MRCA at past generation *tk*, $P(k) = (n(n-1)/2N)(1-n(n-1)/2N)^{k-1}$. Although this is described in a discrete-time scale, it can be approximated in a continuous-time scale with the molecular clock hypothesis (Section 1.3.3), as a

density function of the exponential distribution that has a mean of $2N/n(n-1)$ and variance of $4N^2/[n(n-1)]^2$ (please see further details in Lemey et al., 2009). The coalescent interval is equivalent to the branch length of the genealogy. Many extensions have been made to relax unrealistic assumptions in the basic coalescent model; however, the concept that a distribution of generation interval (often referred to as waiting period) and hence the time to the most recent common ancestor (TMRCA) can be described as a function of population size is fundamental to any of these models. For instance, the population size was allowed to vary over time (Griffiths and Tavare, 1994), and the change in population size is often described in a various form of parametric or non-parametric model such as exponential distribution and Skyline plot (Drummond et al., 2005).

The mutation process is then modelled along the branch in constructed genealogies as follows. Genetic sequence from a single ancestor is passed to its descendants with probability μ of mutation being occurring. When mutation occurs, a site among the sequence is randomly chosen and a mutation is added according to a pre-specified substitution model. In the discrete-time coalescent model, the probability of observing a mutation between a lineage at generation *t0* and its ascendant at generation *t1* is μ, and more in general, the probability of a lineage having a mutation at the first time in generation *tj* in the past can be described as $P(j)=\mu(1-\mu)^{j-1}$. As above under the condition of N being large, this can be approximated in a continuous-time scale as $1-e^{-\theta t/2}$, where t = j/N and θ = 2Nμ. Here, θ is referred to as the scaled mutation rate and mutation events can be now modelled along with coalescent events, therefore, enables the likelihood function of data to be formulated (Hein et al., 2005). Given these parameter estimations are of primary interest in the coalescent model, estimations are performed by integrating likelihood over nuisance parameters such as genealogies and branch lengths.

Using the coalescent model, we can calculate the time-varying growth rate of infected population size, and hence basic reproduction number (Bouckaert et al., 2014a; Drummond and Rambaut, 2007). A word of caution here is, however, that the coalescent model calculates basic reproduction number based on the growth rate and the average duration of infectiousness, therefore, the validity of this approach is less clear for diseases which have varying infectious period such as tuberculosis and HIV.

Some key assumptions made by coalescent model should be noted. As already mentioned, the number of sampled individual is much smaller than the total number of infected individuals. Sampled sequences need to be tips of genealogical tree and none of them are direct ancestors of other genetic samples. This assumption can be in particular violated in the early phase of an epidemic and violation of this assumption is shown to cause serious bias in inferences (Gavryushkina et al., 2014). The other key assumption of traditional coalescent models is that the demography of pathogens changes deterministically. As has already been well understood in epidemiology, the stochasticity plays a pivotal role in infectious disease dynamics, namely in the early phase of epidemics. Boskova and colleagues suggested that birth-death model is less sensitive to the fluctuations of early outbreaks than coalescent model that assumes a deterministic exponentially growing infected population (Boskova et al., 2014). Rasmussen and colleagues developed a methodology to fit stochastic epidemiological models to genealogy in the coalescent framework (Rasmussen et al., 2014).

1.3.1.2.2 Birth-death model

Having its root in Yule's process that provides a 'null' model for the stochastic process that generates a phylogenetic tree of n extant species (Aldous, 2001), birth-death process has been widely used in many field of biology to describe population dynamics. This model simulates the branching process forward from single initial individual with each individual having certain probabilities of extinction (i.e. death) and branching (i.e. birth). In the application of birth-death model to infectious disease epidemiology, birth events represent the disease transmission and death events represent the cease of being infectious. Death events can occur due to treatment of infected individuals and behavioural changes of infected individuals. The traditional birth-death model assumes that all n extant individuals are sampled, which is rarely the case in infectious disease epidemiology. Stadler (2009) extended the birth-death model to incorporate a random sampling process to account for the incomplete sampling (Stadler, 2009). Also, traditional birth-death models assume a constant probability of sampling throughout the evolutionary history of pathogens, which is apparently invalid under some situations. A birth-death skyline plot model was developed to relax this assumption by allowing the sampling proportion to change in a piece-wise manner (Stadler et al., 2013).

In contrast to coalescent model, basic reproduction number can be estimated solely from genetic sequence data in birth-death model (Stadler et al., 2012).

1.3.2. Substitution models

Genetic distance is a measure to quantify the distance between two or more genetic sequences. The simplest genetic distance between sequences is referred to as observed distance or p-distance. This metric calculates the number of nucleotide sites that differ between a pair of sequences based on Hamming distance. This metric is, however, often inappropriate when the time span between sequence samplings is not short and the substitution rate of nucleotide is high. This is because a single nucleotide site can experience multiple substitution events which might be masked by more recent substitutions and cannot be accounted for by this simple distance measure, resulting the calculated distance being underestimated. Such underestimation can be adjusted by either explicitly formulating a relationship between the observed and expected distance or constructing a likelihood function that describes the probability of observing a set of nucleotides at each site of alignments given the expected distance. Both methods are based on models that specify the probability of substitutions (i.e. transitions and transversions) between each of four nucleotides (i.e. A, C, G and T). Such specification model is called as substitution model and the collection of these substitution probabilities in a matrix form is generally referred to as Q-matrix.

Typically these models assume Markov process for nucleotide substitutions i.e. the instantaneous substitution rate from one nucleotide base to another nucleotide base at a given site is independent from the base that constituted the site previously. Another assumption often made is stationarity where each sequence analysed has similar nucleotide compositions. Different substitution models have different parameters to be estimated. For instance, the simplest Jukes and Cantor (JC69) model assumes that any nucleotide has the same probability of being replaced by any other nucleotide. In contrast, general time reversible (GTR) model assumes that the rates of change between a pair of nucleotide base are reciprocally equal (i.e. substitution rate from nucleotide i to nucleotide j is the same as that from j to i) but each nucleotide pair is allowed to have different substitution rates. The detailed descriptions of each substitution model are beyond the scope of this review and interested readers are referred to elsewhere (Lemey et al., 2009).

These models originally assume a constant transition rate across sites in sequences, however, this assumption may be often violated. For example, it is well known that mutations in protein coding regions will accumulate faster at the third codon compared to that at the first or second. The homogeneous rate assumption can be relaxed to allow heterogeneous transition rates across sites and such heterogeneity is often assumed to follow gamma (Γ) distribution or negative-binomial distribution (Lemey et al., 2009). Heterogeneous substitution rates across sites can be further relaxed by allowing the presence of sites that do not undergo mutations (i.e. invariant sites) and such model is often described as $\Gamma + I$ approach in literature. Another method includes discrete gamma distribution in which sites of sequences are partitioned into a number of classes where different substitution rates are estimated, non-parametric estimation (Huelsenbeck and Suchard, 2007) and hierarchical model (Pond and Frost, 2005); however, these are beyond the scope of this review. Selection of a better model amongst various substitution models and relaxed assumptions such as heterogeneous substitution rates can be performed in the same manner as with the common statistical model selection. That is, likelihood ratio test is applicable when one model is nested to the other and otherwise various information criterion approaches such as AIC and BIC are available.

1.3.3. Molecular clock hypothesis

As discussed above, the observed genetic distance can be calibrated to infer the actual amount of mutations by using different substitution models. Genetic distance is, however, a relative measure to describe the evolutionary relationship between sequences and does not tell us the relationship between the evolution and the time. Phylogenetic trees constructed based solely on genetic distance cannot identify the overall common ancestor for all sampled taxa and referred to as unrooted-tree (Pybus, 2006). The tool which can associate the amount of substitutions with the time is widely known as molecular clock. The molecular clock hypothesis was originally introduced in the 1960's by Zuckerkandl and Pauling who proposed that genetic mutation accumulates at constant rate over time (Zuckerkandl and Pauling, 1962). This assumption is of particular importance because this enables us to estimate the historical dates of events such as time to the most recent common ancestor (TMRCA). This means the evolutionary history of pathogens can be now aligned with epidemiological events that occurred in comparable time.

Trees constructed based on the molecular clock have a common ancestor for all isolates and referred to as rooted-tree, and their branch lengths now represent time rather than genetic distance. The molecular clock can be broadly divided into strict molecular clock and relaxed molecular clock, although the term "molecular clock" has been sometimes used to infer strict molecular clock. The former assumes the evolutionary rate to be constant over every branch of tree (i.e. all taxa have the same evolutionary rate). Note that the evolutionary rate here is fundamentally different from substitution rate we discussed above in that the former describes the rate of evolution since the divergence from its ancestor to the sampled point.

Since the emergence of the molecular clock hypothesis, along with the findings of situations where the strict molecular clock is violated, this assumption has been fairly contentious (Kumar, 2005). A large number of studies have therefore not employed the molecular clock hypothesis and resorted to unrooted tree. As pointed out by Pybus, such approach eventually assumes, however, the variation in evolutionary rates among branches is infinite and includes unnecessarily large number of parameters, which might result in increased statistical uncertainty (Pybus, 2006). Alternatively, a number of relaxed molecular clock models have been developed including the local clock, autocorrelated clock, and uncorrelated clock models. In brief, these models allow evolutionary rate varies across tree branches in different ways. For instance, the uncorrelated relaxed molecular clock model assumes the evolutionary rate follows a specific underlying distribution such as exponential, log-normal and inverse Gaussian distribution (Drummond et al., 2006). As similar to the model selection in substitution model, a standard statistical selection method such as likelihood ratio test and AIC can be used for molecular clock model selection (Arbogast et al., 2002). The details of each model can be found in excellent reviews (Ho and Larson, 2006; Ho and Duchêne, 2014).

1.3.4. Limitations in current phylodynamic studies

Bayesian evolutional analysis requires a prior to be set for various parameters including substitution rate and tree topology. As highlighted already, coalescent and birth-death models are popular approaches to construct a genealogy, and hence used as tree prior. That is, various assumptions made in these models described above are naturally inherited to Bayesian evolutional analysis. The other key assumption made in these models is that samples are collected randomly from the background population. However, many genetic sequences are collected in a biased manner due to the nature of contact tracing investigations, thereby violating this assumption. Several studies have evaluated the impact of this violation on phylodynamic inferences. Silva and colleagues showed that an inclusion of many epidemiologically-linked sequences can result in a significantly biased inference (Silva et al., 2012). A more recent study also suggested that phylodynamic inferences are biased if sampling times depend on the population size—for instance, sampling is conducted more frequently when there is a larger number of infected individuals (Karcher et al., 2016). These studies are, however, conducted assuming a population has a homogeneous mixing structure. The impact of sampling on phylodynamic inferences for livestock diseases, which spread through highly heterogeneous and structured contact networks, is a neglected research field.

1.4. Structure of this thesis

As highlighted above, there are still limitations in the use of livestock movement data and genetic data. Nevertheless, these limitations are largely neglected in the current literature. The objective of this thesis was therefore to highlight these limitations and address why this matters.

As an outset in Chapter 2, the literature on farmers' behavioural changes was reviewed to highlight the knowledge gaps and current assumptions, which allowed us to identify the aspects of livestock movement data that need to be analysed to fill this knowledge gap. Chapter 3 then analysed the national-scale livestock movement records in New Zealand to make an inference whether or not dairy farmers purchase livestock considering a risk of disease introduction. This is the first study in New Zealand that attempted to understand how farmers' perceptions towards disease may influence their livestock trading patterns. In Chapter 4, farmers' livestock purchasing behaviours were then studied in-depth using a qualitative interview method. Qualitative methods are a great approach to understand the underlying mechanism, which was farmers' trading behaviour in this particular context. This study highlighted how erroneous inferences could be made by using solely big data without understanding the mechanisms that generated the data. This study also provided new insights in farmers' livestock selling practices, which were further analysed in Chapter 5. In Chapter 5, livestock movement records and production records were analysed together to confirm the hypotheses that were generated from the qualitative interview study. An individual-based disease model was then developed to simulate disease spread and genetic mutations in Chapter 6. This model was used to explore how different genetic sampling strategies influence phylodynamic inferences for livestock diseases that spread through structured population and complex contact network.

A discussion of the findings of these studies, future challenges and opportunities in veterinary epidemiology in the big data era then follows. The study limitations and the challenges encountered during the research presented in this thesis are also discussed.

Chapter 2

Modelling dynamic human behavioural changes in animal disease models: challenges and opportunities for addressing bias

Arata Hidano[1], Gareth Enticott[2], Robert M Christley[3, 4], M Carolyn Gates[1]

[1] EpiCentre, School of Veterinary Science, Massey University, Palmerston North, New Zealand

[2] Cardiff School of Geography and Planning, Cardiff University, Cardiff, United Kingdom

[3] Department of Epidemiology and Population Health, Institute of Infection and Global Health, University of Liverpool, Neston, United Kingdom

[4] Institute of Veterinary Science, University of Liverpool, Neston, United Kingdom

2.1. Abstract

Over the past several decades, infectious disease modelling has become an essential tool for creating counterfactual scenarios that allow the effectiveness of different disease control policies to be evaluated prior to implementation in the real world. For livestock diseases, these models have become increasingly sophisticated as researchers have gained access to rich national livestock traceability databases, which enables inclusion of explicit spatial and temporal patterns in animal movements through network-based approaches. However, there are still many limitations in how we currently model animal disease dynamics. Critical among these is that many models make the assumption that human behaviours remain constant over time. As many studies have shown, livestock owners change their behaviours around trading, on-farm biosecurity, and disease management in response to complex factors such as increased awareness of disease risks, pressure to conform with social expectations, and the direct imposition of new national animal health regulations; all of which may significantly influence how a disease spreads within and between farms. Failing to account for these dynamics may produce a substantial layer of bias in infectious disease models, yet surprisingly little is currently known about the effects on model inferences. Here, we review the growing evidence on why these assumptions matter. We summarise the current knowledge about farmers' behavioural change in on-farm biosecurity and livestock trading practices and highlight the knowledge gaps that prohibit these behavioural changes from being incorporated into disease modelling frameworks. We suggest this knowledge gap can be filled only by more empirical longitudinal studies on farmers' behavioural change as well as theoretical modelling studies that can help to identify human behavioural changes that are important in disease transmission dynamics. Moreover, we contend it is time to shift our research approach: from modelling a single disease to modelling interactions between multiple diseases and from modelling a single farmer behaviour to modelling interdependencies between multiple behaviours. In order to solve these challenges, there is a strong need for interdisciplinary collaboration

across a wide range of fields including animal health, epidemiology, sociology, and animal welfare.

2.2. Introduction

Since the seminal application of infectious disease models to the 2001 foot-and-mouth disease (FMD) outbreak in the United Kingdom (UK), the number of published modelling studies for livestock diseases has increased dramatically (Keeling, 2005; Pomeroy et al., 2017). These models can be useful tools for evaluating the efficacy of different disease control strategies especially in situations where it may not be ethically justifiable or where it may be too time consuming and expensive to perform research studies in the real world. Livestock infectious disease models are generally built as follows. First, the population demographic structure and various disease transmission pathways are identified based on existing knowledge about the disease system. The within-farm and between-farm disease transmission dynamics are then modelled, if necessary, by defining contact patterns over time and relevant parameters to describe the likelihood of transmission occurring through those contacts. Once the baseline disease dynamic model has been developed and validated, the effectiveness of various control strategies may then be evaluated by imposing modifications on the system.

Model structure has also become increasingly complex, evolving from simple compartmental models where each farm does not have an identity all the way to sophisticated individual-based models where the population is divided into a number of subpopulations that are typically spatially separated and each animal is individually identified and traced throughout the simulation. With the increasing availability of national livestock movement records, researchers can also explicitly replicate livestock movement patterns that occurred in the past to realistically simulate how disease spreads along with movements. With more thorough sensitivity analyses being performed, it is also possible to test the influence of recognised model assumptions and limitations on the final control recommendations (Probert et al., 2016). Using these tools and methodologies, infectious disease dynamics can be studied with an unprecedentedly high resolution. However, most models still assume constant human behaviour meaning that the patterns of within- and between-farm contacts as well as the risk of transmission through the contact remain constant in the models even though farmers in the real world may need to adapt their behaviours to deal with the disease and the control strategy imposed upon them (Barnes et al., 2015; Ferguson, 2007; Sebastian Funk et al., 2010; Funk et al., 2015; Verelst et al., 2016).

Both empirical and theoretical studies for human diseases clearly show that behaviour can have a substantial impact on disease epidemiology. For instance, in the recent Ebola outbreaks in Africa, it was found that traditional funerals in West Africa that involve family members washing the corpse contributed significantly to the number of secondary infections with Ebola virus (Tiffany et al., 2017; Victory et al., 2015). Vaccine refusal or vaccine hesitancy can also occur in a spatially clustered manner for various reasons, including shared views within a community or poor financial status in an area, and this can substantially increase a risk of infection in the geographical area (Goldlust et al., 2017). Not only that, it is known that humans change behaviours in response to various factors including disease occurrence, increased awareness towards a disease risk, social norm, and the perceived efficacy of a disease control strategy (Sebastian Funk et al., 2010; Verelst et al., 2016). In the same example of the Ebola outbreak, Abramowitz et al., (2017) showed how the local community's beliefs about the source and transmission of Ebola changed during the outbreak period, which subsequently changed how people implemented infection prevention and control measures to protect their own health status (Abramowitz et al., 2017).

Similar findings have been observed with animal diseases. For example, a qualitative study on horse owners' perception towards Hendra virus revealed that some people share a belief that vaccinations may lead to adverse reactions in horses such as decreased performance, abortion, and death (Wiethoelter et al., 2017). Some owners also believed that the vaccine was not tested rigorously enough due to it being developed in a rush, which has been identified in other human behavioural studies as yet another reason why people may fail to adopt vaccinations (Karafillakis and Larson, 2017). Regarding other dynamic human behaviours, empirical observations suggest that live bird market closures in response to avian influenza outbreaks may induce an undesirable behavioural change in poultry owners, such as increasing the frequency of movements of high-risk animals to avoid culls or performing illegal trading through an underground markets (Manabe et al., 2012; Paul et al., 2013) both of which contribute to further disease spread (Nguyen et al., 2017). Other studies have similarly reported that movement restrictions can result in infected livestock being sold from an area where disease outbreak occurs (Gunarathne et al., 2015; Yupiana et al., 2010). Emergence of undesirable behaviours has been also observed amongst some UK farmers, who performed illegal badger culling to control bovine tuberculosis (bTB) because they did not trust the government and hence its legislation (Enticott, 2011).

Two sets of behaviours in particular have been identified as being highly influential for human disease spread; one as behaviours related to determining the mixing or contact patterns between hosts and the other as behaviours related to disease prevention and control (Funk et al., 2015). Since livestock populations are managed by humans, it is only natural that similar behavioural factors can influence the epidemiology of livestock diseases. In the context of livestock diseases, these behaviours translate into livestock contact patterns and the biosecurity practices farmers take to prevent disease from spreading through these contacts. The complex interrelationship between disease spread and dynamic human behaviour needs to be accounted for in disease simulation models to minimise potential bias in inferences (Vernon and Keeling, 2012).

This logically leads to the questions of how much detail of dynamic human behavioural change do we actually need to capture to make valid modelling inferences and how should we best model dynamic human behavioural changes? Answering these questions will require an understanding of (1) the disease-related factors that are most likely to cause behavioural change including epidemiological factors (e.g. knowledge of disease prevalence, incidence, and mortality rates) as well as other broader psychological and social factors (e.g. farmers' perception of disease risk and the disease experience of neighbouring farmers) and how they change human behaviours, and (2) what methods we can use to quantitatively model the association between changes in these disease-related factors and changes in a behaviour. Fortunately, the rapidly growing literature on farmers' behaviours provides greater knowledge on 'what' may affect their behaviours. However, we still lack a solid understanding of 'how' these factors operate and interact to deliver a dynamic human behaviour, as represented in Figure 2-1.

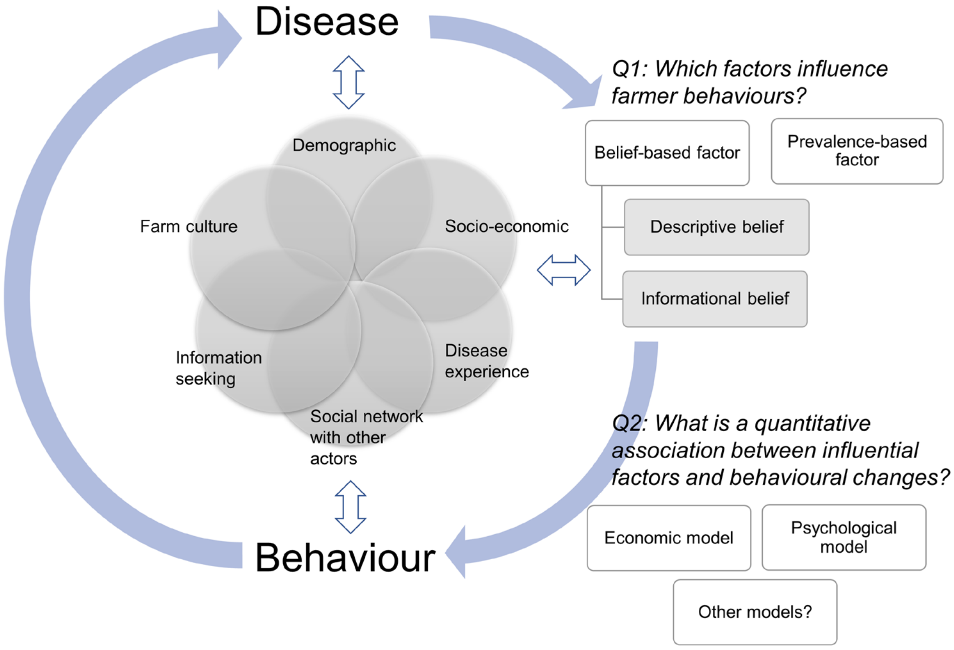


Figure 2-1 Schematic diagram of the feedback loop between disease and behaviour.

Disease influences farmer behaviours through prevalence-based and belief-based factors, which are influenced by various farms' and farmers' characteristics (e.g. demographic factors, socio-economic status, and social network with other actors). Behaviour in turn changes disease dynamics. Farms' and farmers' characteristics influence both disease and behaviour, which also influence farms' and farmers' characteristics. We highlighted two key areas (Q1 and Q2) that need further studies in order to accurately capture the inter-relationships between disease and dynamic human behaviours.

In this review paper, we first summarise the existing literature around the disease-related factors that are responsible for changing farmers' behaviours (Section 2.3). We then discuss different methods that currently exist for building dynamic human behavioural change into disease simulation models (Section 2.4). This paper concludes with a discussion of challenges and opportunities for future research (Section 2.5).

2.3. Disease-related factors relevant to farmers' dynamic behavioural change

Funk et al., (2010) proposed a system for classifying the disease-related factors that can lead to human behavioural change based on both the source of information (global or local) and the type of information (prevalence-based versus belief-based) that individuals routinely use to make personal health decisions (Sebastian Funk et al., 2010). Global information refers to disease information that is widely available in the public domain through national television, newspapers, magazines, and government information services. Local information refers to disease information that is only circulated amongst close social neighbours such as discussions between neighbouring farmers or local farming groups. It is important to distinguish between these two sources because local knowledge may lead to significant local and regional variation in human behaviour (e.g. clustered vaccination), which can have a substantial impact on patterns of disease spread through the global population (Funk et al., 2009; S. Funk et al., 2010).

Prevalence-based information includes direct factual knowledge about how commonly the disease occurs in a population (prevalence or incidence) as well as distribution of outcomes from a disease (e.g. number of cumulative deaths). For example, a previous study modelling human mobility patterns in response to an infectious disease epidemic assumed that people would avoid travelling to areas with a high disease incidence to minimise their risk of becoming infected (Meloni et al., 2011). Belief-based information on the other hand includes information that influences people's beliefs and perceptions about the risks of disease, which may not have any correlation with the true disease situation. For example, individuals may choose to avoid vaccination because they perceive their risk of developing severe adverse vaccine reactions is greater than their risk of getting the disease even though this is statistically untrue (Karafillakis and Larson, 2017). That is, prevalence-based information is based on *incidence*, whereas belief-based information is based on *incidents*. It is important to distinguish this difference when modelling dynamic human behaviour because the former will fluctuate according to disease prevalence, whereas the latter has a more complex mechanism that operates largely independently from the disease dynamics.

Below, we first summarize the current literature about how prevalence-based and belief-based factors may influence farmers' decisions to either change the livestock contact patterns or adopt different control measures in response to a disease outbreak. We then highlight the key knowledge gaps that hinder modelling a dynamic human behaviour.

2.3.1. Prevalence-based factors

2.3.1.1 Local prevalence

There is evidence in the literature that suggests perception of risk—perceptions of threat, vulnerability, and severity—plays an important role in determining human health behaviour (Brewer et al., 2007; Ferrer and Klein, 2015). Both human and livestock disease literature suggests that disease incidence influences perception of risk, which in turn affects uptake of disease preventive measures (Koh et al., 2005; Wiethoelter et al., 2017). Local disease incidence, in particular, is often reported to trigger farmers' behavioural change. For instance, Garforth and colleagues reported that UK sheep farmers often demonstrated that they were willing to vaccinate animals against bluetongue once they heard the disease occurred in their region (Garforth et al., 2013). Another qualitative study on Johne's disease suggested that farmers may not invest resources into controlling disease until they see clear evidence of disease on their farms (McAloon et al., 2017). This 'wait and see' attitude of farmers towards implementing on-farm biosecurity practices has been repeatedly reported in literature. For example, Alarcon et al. (2014) studied the reasons UK pig farmers decide to control disease and found that these reasons include observations of sick animals, reduced production, and increased mortality (Alarcon et al., 2014). Brennan and colleagues reported that some UK dairy farmers perceived that they can change the intensity of on-farm biosecurity practices when necessary, such as in the face of disease outbreaks (Brennan et al., 2016). A rise in the local disease incidence can be, therefore, a legitimate parameter to model change in farmers' biosecurity practices, although there is a considerable knowledge gap on what threshold incidence may trigger behavioural change, or even whether or not such a threshold exists.

2.3.1.2 Global prevalence

The abovementioned UK sheep study also reported that a disease incidence at a wider spatial scale is less likely to motivate farmers to vaccine their animals against bluetongue (Garforth et al., 2013). This relatively weak influence of global prevalence, as opposed to local prevalence, has been also reported for farmers' trading behaviours. Although some farmers, but not all, avoid purchasing livestock from a high disease risk area for both endemic (Enticott et al., 2015a; Young et al., 2010) and exotic diseases (Hidano et al., 2017b, 2017a), it is unclear whether or not farmers are engaged in the risk-averse trading in response to disease prevalence. In fact, a UK study showed that the proportion of farmers from low bTB risk areas who mentioned they do not purchase cattle from high risk areas was larger than that of farmers from high bTB risk areas; however, farmers listed maintaining an existing trade channel as the main reason for this behaviour (Little et al., 2017a). Studies from New Zealand also suggested that the use of a stock agent may result in farmers' apparent risk-averse trading behaviours, although they may not be necessarily concerned about disease status (Hidano et al., 2016, 2017b, 2017a). There was evidence that farmers avoid purchasing from certain geographical areas (Hidano et al., 2017b, 2017a); however, these areas may not necessarily have a higher disease prevalence than other areas or the area may represent a large geographical area (e.g. North Island of New Zealand rather than a specific region). Taken together, these may suggest that farmers' behaviour is more likely to be influenced by their interpretation of disease prevalence, rather than the absolute prevalence, although this needs further empirical studies. This emphasises the importance of belief-based factors, which we discuss below.

2.3.2. Belief-based factors

The term 'belief' is used to represent anything farmers believe; belief therefore includes perceived control of behaviour (self-efficacy), perceived efficacy of behaviour, perceived severity of disease, perceived benefit of controlling disease, social norm, and so on. In psychology literature, belief is assumed to form in one of three ways (Fishbein and Ajzen, 2010): (1) descriptive belief—personal beliefs arising from direct observations, (2) informational belief—beliefs arising from accepting information from outside, and (3) inferential belief—beliefs arising from processing other beliefs. When modelling human behaviour and hence modelling belief formation, it is important to distinguish descriptive and informational belief: the former develops through farmers' personal experience and the latter through acquiring information from other actors such as peer farmers, veterinarians, government, and media. In this section, we focus on three key factors that contribute to a belief formation: disease experience, perception towards disease control measures, and social norm. We discuss how these factors develop both descriptive and informational beliefs.

2.3.2.1 Disease experience

One of the most studied factors that contribute to forming farmers' descriptive belief may be actual disease experience. Enticott et al., (2015) investigated practices and attitudes towards bTB among farmers in an area which had been recently designated as bTB endemic (Enticott et al., 2015a). This study found that the proportions of farmers that avoid purchasing from a high bTB risk area were similar between those previously had a bTB breakdown and those not (Enticott et al., 2015a), which may suggest a direct bTB experience may not necessarily change farmers' trading behaviours. On the other hand, a Dutch study suggested that previous direct experience of having bluetongue-related clinical cases was associated with a higher probability of vaccinating their livestock (Elbers et al., 2010).

This discrepancy in the effect of actual disease experience on behaviour change may be attributable to, at least in part, the difference in how risk perception is updated by the disease experience. Ferrer and Klein (Ferrer and Klein, 2015) summarised three types of risk perceptions recognised in health behaviour discipline: (1) deliberative risk perception—systematic, logical, and rule-based perception to estimate, for instance, the likelihood of negative event occurring, (2) affective risk perception—affect associated with risk such as worry or anxiety about a threat of negative event, and (3) experiential risk perception—rapid judgements made from deliberative and affective perception, which can be described as intuition or 'gut' feeling.

As has already been discussed, disease experience seems to contribute to high deliberative risk perception—after experiencing a disease, individuals may come to perceive that they are at high risk of contracting the disease. Studies also suggest that farmers who are not worried about future disease incidence (i.e. a low affective risk perception) have poor uptake of on-farm biosecurity practices (Schemann et al., 2011). Interestingly, however, having both high deliberative and affective risk perception at the same time does not necessarily lead to implementing a preventive measure. For instance, a study investigating intention to quit smoking demonstrated that those who had high perception of risk—individuals that perceived they have a higher risk of contracting lung cancer—and high worry—individuals who were more anxious about contracting lung cancer—were more likely to have a lower intention to quit smoking (Klein et al., 2009). High levels of deliberative and affective risk perception may result in specific experiential perception which provides 'fatalistic' belief about disease risk (Ferrer and Klein, 2015). Indeed, 'fatalistic' belief has been observed among farmers who are at high risk of infection to disease which is difficult to prevent such as bTB (Enticott, 2008) and Hendra virus (Wiethoelter et al., 2017). Farmers' emotion towards disease may be shaped by many events in their farming life and a single devastating event can also have a prolonged effect on their emotions. For instance, a study on UK farmers reported that a few farmers listed FMD, rather than other diseases, as a particular concern even though more than 4 years had passed since the 2001

FMD outbreak in the UK (Brennan and Christley, 2012). These together emphasise the importance of understanding how previous experiences form farmers' emotion and how such emotion influences their behaviours, which is substantially missing in current literature.

Studies also suggest that peer farmers' disease experience is likely to act as informational belief. In fact, Lupton (2013) argues that risk perception and emotion are fluid, shared, and developed through interaction with others, material objects, and space (Lupton, 2013). For instance, a study on UK pig farmers demonstrated that stories about negative impact of disease circulating among farmers triggered information seeking behaviour of some individuals which did not have the disease (Alarcon et al., 2014). New Zealand farmers from a high bTB risk area also share a belief about bTB that they are always 'one test away from being infected'—the local community developed understanding that bTB breakdown is unpredictable and inevitable after observing many bTB cases that occurred without any clear reasons (Enticott, 2016). It is therefore important to model transmission of informational belief from affected farmers to other farms and how this updates individual's risk perception, and hence behaviour.

2.3.2.2 Perception towards disease control measures

Beliefs including perceived efficacy and safety of control measures, perceived benefit of controlling disease, self-efficacy and perceived behavioural control are similarly likely to transmit directly between farmers. Detailed descriptions for each term can be found in elsewhere (Mankad, 2016; Ritter et al., 2017). Alarcon et al., (2014) reported that farmers may start implementing a specific disease control measure when they obtain 'word of mouth' information on how effective the measure was on the other's farm (Alarcon et al., 2014). Studies also suggest that the lack of self-efficacy—one's belief to his/her ability to perform a behaviour to obtain a desired outcome (Bandura, 1978)—is shared by farmers within a community who observe disease control measures have not worked on other farms (Enticott, 2016, 2008). Wilson et al., (2015) argued that self-efficacy may be developed among a small group of farmers in which they share knowledge, experiences, and skills, which can lead to a behavioural change in the community (Wilson et al., 2015).

These beliefs may also transmit between farmers indirectly through other actors such as veterinarians and farm advisors: farmers consider, at least to some extent, these actors to have good and reliable local knowledge (Alarcon et al., 2014; Ritter et al., 2017; Wilson et al., 2015). On the other hand, however, a longitudinal study of UK farmers suggests that farmers' views on disease control interventions may change little over time. This study tracked farmers' confidence in vaccinating badgers against bTB to help reduce disease in cattle (i.e. self-efficacy), farmers' confidence, and their trust in Government, identifying that these remained low throughout the duration of the study period (Enticott et al., 2014). While disease prevalence appeared to be unrelated to vaccine confidence, the spread of stories of vaccine failures by local veterinarians and farmers were connected to declining confidence (Maye et al., 2015), suggesting veterinarians play a significant role in spreading information.

2.3.2.3 Social norm

Social norm has been frequently modelled in the context of human disease to account for human behavioural change, particularly for vaccine behaviours (Oraby et al., 2014; Pananos et al., 2017). Social norm is often categorised into descriptive norm—perception about what is typically done—and injunctive norm—perception about what is typically approved and disapproved (Cialdini and Goldstein, 2004). Although within veterinary literature the influence of social norm on farmers' behaviour has been repeatedly mentioned, there is only little knowledge on how social norm actually acts on farmers (Sok et al., 2016; Swinkels et al., 2015; Toma et al., 2015). As highlighted by Maye and colleagues (Maye et al., 2017), this lack of knowledge may arise from the lack of studies that separated the influence of social norm from attitude or the lack of studies that identified a full range of influential actors.

A study on farmer antibiotic use for mastitis treatment, however, provides interesting insight on how social norm influences the duration of antibiotic use by farmers (Swinkels et al., 2015). This study identified that the duration of antibiotic use was associated with the duration of clinical cure. However, the increased cost due to the extended antibiotic treatment (e.g. more waste milk) was not a concern for any of the farmers studied. The mastitis treatment practices of the studied farmers seemed to be little influenced by perception of society such as media and government, which tend to be against prolonged antibiotic use due to its potential association with the development of antimicrobial resistance. The authors hypothesise that farmers choose to provide the perceived best possible treatment, which farmers believe is approved to be a good practice by other farmers. Extended antibiotic treatment therefore provides farmers with a feeling of being a 'good farmer' (Swinkels et al., 2015).

The concept of the 'good farmer'—how the identity of being a good farmer influences farmer behaviour—has been recently highlighted in social science studies. Naylor et al., (2018) identified three identities of 'good farmer' in the context of exotic disease control (Naylor et al., 2018); 'good stockman', 'good neighbouring farmer', and 'good public facing farmer'. Health and welfare of animals is valued by farmers with the 'good stockman' identity, which may encourage farmers to identify and report suspicious disease quickly. Farmers with 'good neighbouring farmer' identity have a feeling of responsibility to local farmers, which encourages them to minimise disease spread to other farms. The last identity, 'good public facing farmer', is associated with maintaining a positive image of farmers' industry. The role of perception of responsibility on farmers' behaviour has been similarly reported for (potential) zoonoses. A study on farmers' intention to control *Escherichia coli* O157 suggested that farmers who feel they are responsible for controlling the disease were more likely to be willing to use disease control measures (Toma et al., 2015).

As these studies highlight, pressure from peer farmers, industry, and society have, to some extent, an impact on farmers' behaviour. However, there is currently a significant knowledge gap that prohibits modelling this impact. It has been recently shown that the influence from other actors on farmers' behaviour varies depending on the context and disease. Using bTB as a case example, Maye et al., (2017) showed that while farmers perceived their decision to implement badger culling would be influenced by peer farmers, that for vaccinating cattle against bTB would be influenced by their veterinarians (Maye et al., 2017). This study result not only suggests the difficulty in determining which actors to include in modelling the impact of social norm on farmers' behavioural change but also raises an important question: Would farmers' decision to implement which control options be influenced by actors they perceive most important? In other words, farmers may simply decide to implement a practice recommended by a specific actor (e.g. veterinarian) because they perceive the actor's opinion important. Should this be the case, a question to ask is, who an influential actor is for farmers?—rather than, whether or not a specific actor's influence is important for farmer's intention to perform the practice. This is linked to the problem arising from looking at only single behaviour, which we discuss in the next subsection.

2.3.3. Knowledge gaps and limitations

2.3.3.1 Lack of understanding on qualitative behaviour change of farmers

As highlighted above, there is relatively rich information on disease-related factors relevant to farmers' dynamic behavioural changes. Nevertheless, we have currently very limited knowledge on how these factors actually change their behaviours. This is partially due to the current research approach: A majority of studies focus on how economic and psychological determinants lead to a single behaviour, which is pre-defined by researcher. This is reasonable if disease control is well-established and its option is very limited e.g. vaccination for exotic disease. Nevertheless, in reality, this is not the case for many important livestock diseases; farmers often have multiple options of disease control measures and it is unlikely farmers choose one measure through a full assessment such as a cost-benefit analysis (Anneberg et al., 2016; Garforth, 2015).

Indeed, recent studies suggested that each farmer develops a different control strategy depending on their situations, risk perceptions, and disease understanding (Hidano et al., 2017b, 2017a). If farmers are already implementing their 'biosecurity' practices—which may be supported by previous findings that farmers feel they are doing sufficient practices—in response to disease and their farming experience, this raises an important question which is already covered by Shortall and colleagues (Shortall et al., 2016): What does 'good biosecurity' really mean to farmers and other actors? In the animal welfare context, it has been clearly highlighted that veterinary experts and farmers frame a behaviour differently: while veterinary experts frame record-keeping practice is the key to improving animal welfare, farmers consider this practice as something to satisfy external accountability demands (Escobar and Demeritt, 2017). This may well be the case for biosecurity practices—we assume a certain practice is essential to reduce disease risk, but farmers may have a totally different idea for the same purpose. This is a critical assumption we make, perhaps unconsciously. We need better understanding of why farmers choose a specific behaviour—this is as important as why farmers do not practice a recommended practice, which is the focus of current literature.

As pointed out by Barnes and colleagues, there is also a critical knowledge gap in the interactions between economic and socio-psychological factors on farmer decision making (Barnes et al., 2015). This knowledge gap is critical not only for modelling farmers' behaviour but also for improving overall biosecurity practice in livestock industry. Future studies are warranted to empirically and longitudinally observe how farmers actually change their behaviours (or not) in response to disease experience or disease outbreak and understand why they do so.

The literature also suggests that implementing disease preventive measures may reduce one's risk perception. This reduced risk-perception in turn changes one's other behaviours which are relevant to disease risk. This phenomenon is well-known as 'risk compensation theory' (Hedlund, 2000). For instance, a study reported that horse owners relaxed horse and property management practices after they vaccinated horse against Hendra virus because their confidence in vaccination reduced the risk perception (Wiethoelter et al., 2017). Thus, biosecurity practices are interdependent on each other. Several studies provide useful information on the static interdependency between farm practices (Firestone et al., 2014; McCormick et al., 2017), however, modelling farmers' behaviour change requires knowledge on dynamic interdependency; how implementing one practice leads to a change in risk perception, and hence other behaviour changes. An interesting insight is provided by a randomised control study on UK beef farmers. The authors assessed how tailored biosecurity advice may reduce the prevalence of selected diseases (Cardwell et al., 2016). This study found that farms in the intervention group that received specifically-tailored advice were significantly less likely to be seropositive for BVD and *Leptospira hardjo* in the end of the study period than those in the control group who received only generic advice. Nevertheless, farms in the intervention group were more likely to be positive for bTB in the end of the study period than those in the control group, despite the observation that biosecurity practices on farms in both groups were observed to be improved during this study. The authors speculated that farmers in the intervention group may have put more efforts to purchase from source farms that are accredited as free from diseases such as BVD and *L. hardjo*, which may not necessarily have been free from bTB for a long time (Cardwell et al., 2016). It is therefore important to understand whether or not farmers prioritise a specific disease over others and how this prioritisation may change over time because a biosecurity intervention for or an experience of a specific disease can substantially change farmers' trading practice, which in turn may alter the infection risk to other diseases, as suggested by this study. Again, this knowledge gap can be only filled by investigating longitudinal changes in farmers' behaviours and perceptions in response to various disease-related factors—taking account of relationships between diseases rather than a single disease in isolation—and other wider factors such as animal welfare, environment, and economic components.

Many studies in the literature suggest that farmers will not continue to perform a practice if they do not perceive it to be effective, beneficial in terms of cost, or feasible in terms of both labour and cost (Alarcon et al., 2014; Garforth, 2015; Ritter et al., 2017; Shortall et al., 2017a). Moreover, these farmers' assessments are not fixed in time—milk price, for instance, may drop and practices currently feasible may suddenly become costly. When modelling disease that spreads over a prolonged period of time or that can infect farms over multiple times, it becomes particularly important to account for the maintenance and cessation of changed behaviour. Behaviour change maintenance is, however, a neglected research area (but see (Racicot et al., 2012a) for a veterinary example and (Kwasnicka et al., 2016) for examples in human health).

2.3.3.2 Lack of understanding on the transmission mechanism of beliefs and information

A seminal study by Delabouglise et al., (2015) showed how information on poultry disease outbreak flows between stakeholders, and that this information is likely to trigger various farmers' behaviours such as implementing a preventive measure and selling animals (Delabouglise et al., 2015). However, regarding transmissions of belief-based factors between farms, the literature provides inconsistent evidence. While agricultural studies suggest that information from trusted and credible farmers is the key determinant of one's uptake of knowledge and technology (Garforth et al., 2004; Oreszczyn et al., 2010), disease studies often highlight that farmers do not exchange their disease information (Alarcon et al., 2014; McAloon et al., 2017; Santman-Berends et al., 2014). A lack of communication between farmers on disease problems may be attributable to stigma attached to disease and potential damage to farm's reputation (McAloon et al., 2017; Swinkels et al., 2015), which is particularly important if the farm sells animals to others (Alarcon et al., 2014). Disease information is likely to spread through specific social network of farmers and we need better understanding of the characteristics of such networks. For instance, important questions include: does such information network change between peace time and disease outbreak time? A longitudinal study on farmers' knowledge transfer, such as one by Wood and colleagues (Wood et al., 2014), can be carried out in livestock disease context and would provide useful information to fill this knowledge gap.

A belief-based factor transmission between farms via other actors, such as veterinarian, is another important pathway to be considered. Nevertheless, modelling this pathway is not straightforward for several reasons. First, the role of veterinarians on-farm disease prevention is still unclear. Shortall and colleagues reported that some veterinarians see their current role as 'test and treat' rather than 'predict and prevent' meaning that farmers often seek their advice only when they have problems (Shortall et al., 2016). If informational belief relevant to disease prevention transmits from veterinarians to farmers after these farmers get infection (i.e. test and treat situation), for instance, modelling this belief transmission has a minimal impact on the change of farm susceptibility to disease infection—although such informational belief may have a large impact on within-herd disease transmission patterns. This, however, also means there is an interesting opportunity for modelling studies to demonstrate, for instance, how large the financial benefit that may be gained among farming communities, and countries, by shifting from 'test and treat' to 'predict and prevent' mode—this can be a good incentive for governments to invest onto a communication training for veterinarians so that they can be more involved in disease prevention.

Second, there seems to be large heterogeneity in veterinarians' advice and farmers' uptake of such advice. The former may be influenced by veterinarians' previous experience with specific measures (e.g. having positive or negative experiences with a specific vaccine), confidence in performing the intervention, knowledge of disease, and general attitudes towards disease (Alarcon et al., 2014; Higgins et al., 2017; Richens et al., 2016; Shortall et al., 2016). The latter may be influenced by relationship and trust developed between farmers and veterinarians, and it is known that veterinarians often provide an advice and treatment tailored to each farmer (E. Kristensen and Jakobsen, 2011; Shortall et al., 2016).

Third, little is known about how, why and to what extent veterinarians' practices, such as diagnosis and surveillance activity, evolve over time. This is particularly important for diseases such as bTB that are often non-detectable by farmers—veterinarians define a farmer's disease experience. Enticott (2012) identified that bTB surveillance protocols employed by veterinarians are adapted to the situation at hand: Shortcuts are learned and passed on between veterinarians within veterinary practices in doing so developing their own cultures of testing, both creating and reflecting what are seen to be the central facets of the 'good vet' and veterinary identity (Enticott, 2012a). Studies have, therefore, shown variation in performance between veterinarians in areas of different disease prevalence (Clegg et al., 2015), and where organisational structures and cultural distance between veterinarians and farmers varies (Enticott, 2014). For example, vets working for government organisations find more disease than those in private practice who test their own clients' cattle (Northern Ireland Audit Office, 2009), while other studies find differences in performance between male and female vets (Enticott, 2012b). These variations in behaviour are not strictly confined to veterinarians either; studies of the detection of disease at post-mortem have revealed significant differences between abattoirs (Pascual-Linaza et al., 2017). While these variations may call into question apparent objectivity of disease data, they also suggest the need for greater understanding—both of why variations occur (and what can be done about it), and whether these behaviours change over time in relation to the spread of disease.

These, together, emphasise the need for better understanding of how information spreads between farmers and other actors and how this might change farmers' behaviours. Further studies are warranted in order to incorporate these mechanisms into a disease simulation framework without them being too complex.

2.4. Methods for modelling dynamic human behavioural changes

This section focuses on reviewing the methods used to estimate quantitative associations, which are the information required to incorporate a dynamic behaviour into a disease simulation model. Theoretical studies in animal and agriculture disciplines often model these associations using either of two major approaches—economic or psychological models. Note that other social science methodologies such as grounded theory can be used to generate new theories that can model behaviours; however, to our knowledge, there are very few or no such studies in animal health discipline (but see (Kuehne et al., 2017) for an agricultural behaviour example). We therefore briefly highlight previous applications of economic and psychological approaches. Then limitations of these approaches are discussed and we highlight other potential approaches to associating a human behaviour to a disease-related factor. More details of these methods as well as applications to human diseases can be found in recent excellent review papers (Sebastian Funk et al., 2010; Verelst et al., 2016).

2.4.1. Economic models

For livestock diseases, the seminal papers in this field exclusively model human behaviour in an economic framework: that is, a disease-related factor is an economic cost incurred by a disease and a behavioural change occurs to minimise such a cost. In particular, game theory has often been applied to understand an interdependent nature of decision making on infectious disease control. It assumes that one's decisions about controlling an infectious disease influences local disease epidemiology and hence the disease risk imposed on others, which in turn influences others' decision making. Typically, these studies focus on one particular behaviour, either a single biosecurity practice or a single trading practice. For instance, Hennessy modelled farmers' on-farm biosecurity practices using a simple spatially structured disease model, which accounted only for farm profit (Hennessy, 2007). Kobayashi and Melkonyan performed a theoretical and empirical study using farmers' biosecurity behaviours at a livestock show in California to investigate how the decisions made by individual farmers in a trading pair influenced each other's subsequent decisions (Kobayashi and Melkonyan, 2011). Murray applied game theory to an aquaculture setting to model whether or not fish farmers purchase tested pathogen-free stock or untested stock that may carry a pathogen (Murray, 2014). This study identified that the key motivator for a farmer to uptake a disease preventive measure is often the confidence in other farmers performing the measure.

Given that the objective of these studies is to identify a disease control strategy that maximises the collective benefit under a given human behavior, they typically use very simple disease transmission models. An exception is the work by Tago and colleagues, which modelled dynamic livestock selling behavior of farmers in an economic framework and simulated disease spread using both a network-based model and a spatial transmission model (Tago et al., 2016). This study showed how an inferred effectiveness of a movement restriction policy on a disease spread is overestimated when a dynamic behavioural response is ignored. Hoscheit et al., (2017) modelled French livestock movement patterns accounting for livestock supply and demand, although disease-related factors were not considered in this study (Hoscheit et al., 2017).

These studies however typically consider only a one-off behavioural change. As an exception, Rat-Aspert and Fourichon modelled a dynamic voluntary vaccination behaviour that changes according to a disease prevalence, which in turn influences an economic incentive of vaccination (Rat-Aspert and Fourichon, 2010). However, the behavioural change in this study is assumed to occur only once a year and farmers' decision to vaccinate does not get updated in response to a disease spread situation.

2.4.2. Psychological models

The other class of approaches to modelling human behaviours use psychological models. Unlike economic models, psychological models do not make the assumption that humans behave in a manner to maximise a certain utility. Rather, they assume various psychological factors have an independent association with an intention to perform a certain behaviour, which in turn associates with the actual performance of a behaviour. The psychological factors used to model a behaviour depend on different models. For instance, Theory of Reasoned Action (TORA) assumes one's intention to perform a specific behaviour can be explained by one's attitude and subjective norm towards the behaviour. One's attitude is in turn determined by a belief about, and evaluation of the outcomes of the behaviour (Ajzen and Fishbein, 1980).

Theory of Planned Behaviour (TPB), an extension of TORA and a popular approach in recent veterinary epidemiology literature, assumes that one's perceived behavioural control towards the behaviour also influences one's intention, in addition to the two factors in TORA. This additional component of TPB implicitly accounts for self-efficacy, which is one's belief that one can achieve the behaviour, and other factors facilitate achieving the behaviour such as personal skills, information, opportunities (Ajzen, 2005). The Health Behaviour Model (HBM) has also been frequently applied to human diseases, but less so for livestock diseases (Valeeva et al., 2011). This model assumes engagement in a specific behaviour towards a disease can be explained by factors such as one's belief about the disease problem, perceived benefits of a behaviour, and self-efficacy of a behaviour (Glanz et al., 2015). The greatest strength of these models is that the probability of performing a specific behaviour can be quantitatively described by these factors using a questionnaire survey. These approaches have been often used to investigate why farmers do and do not engage in a specific biosecurity behaviour.

However, one notable exception is a recent work by Fischer and colleagues (Fischer et al., 2017). Using an individual-based model framework, this study accounted for farmers' dynamic treatment behaviours with antibiotics, which in turn influence how disease spreads within a farm. Farmers' dynamic behavioural change was modelled using TPB and their intention to change behaviours is assumed to depend on three factors; the expected economic gain from changing a behaviour, the satisfaction in their own behaviours, and social norms. Although this model still includes several strong assumptions (e.g. farmers have perfect information regarding the cost of measures and the actual behaviour is determined by an intention to perform the behaviour), it is an excellent example of incorporating dynamic human behaviours in a disease simulation model.

2.4.3. Knowledge gaps and limitations

Traditional economic frameworks often assume that humans behave in a manner such that it maximises a certain utility. Game theory assumes that individuals have perfect information as to the cost and effectiveness of a disease control strategy. It is, however, increasingly known that human behaviours do not hold to these assumptions. In reality, farmers do not have sufficient information to evaluate the true cost and effectiveness of a control measure. Moreover, as we discuss in the next section, it is unlikely that farmers go through a full cost-benefit analysis on a control measure accounting for the influence from others' decisions.

Although the ability to quantify an association between each psychological factor and a resulting behaviour is beneficial, particularly for modelling studies, these methods are not without limitations. First of all, TORA and TPB were not originally developed to model behavioural changes (Ajzen, 2015); although they have been applied for this purpose in many studies, the validity of modelling behavioural change using these methods remains unclear. Second, there is evidence of a discrepancy between intention and actual behaviour (Orbell and Sheeran, 1998; Webb and Sheeran, 2006), the so called intention-behaviour gap, which fundamentally violates the assumptions of these models. In fact, it has been long recognised that having an intention to perform a behaviour is often insufficient motivation to actually carry out that behaviour (Heckhausen and Gollwitzer, 1987). Third, these models do not explicitly account for how experience of performing a certain behaviour influences cognitions, which are the impacts of doing the behaviour on a person's attitude, subjective norm, and perceived control (Sniehotta et al., 2015). Literature on farmers' adoption of new technology suggests that establishment of new practices takes time, going through an active assessment period, an implementation period, and a consolidation period where farmers iteratively seek options, invest resources to implement the new practice, and evaluate its effectiveness (Sutherland et al., 2012). Therefore, the lack of a mechanism that captures the process of establishing new behaviour may be a constraint in modelling farmer behaviour, as it is known that farmers are more likely to implement practices that they are experienced in performing (Garforth, 2015).

Not relying on these models, Higgins and colleagues investigated how veterinary clinicians make a treatment decision based on a result from the previous treatment action (Higgins et al., 2017). The authors compared observed clinicians' treatment practices to those theoretically predicted assuming they logically update their beliefs using a Bayesian framework. Although farmers' treatment decision may not exactly match to that of clinicians, the decision of clinicians should be still influential to farmers' decisions given farmer reliance on veterinarians to advise about the best course of action for disease issues (Garforth et al., 2013). This study provides useful information for the disparity between actual human behaviours and expected behaviours that are derived from a certain theory.

While there are no applications in the context of livestock diseases, diverse theories have been developed, tested for their validity, and used for modelling behavioural change in other disciplines. Several key distinctive features of these theories include acknowledging: nonconscious factors (e.g. impulsive and automatic factors) (Hagger and Chatzisarantis, 2014), cognitive habits and socially shared values (Zimmerman, 2013), and emotions (Gutnik et al., 2006; Kirman et al., 2010). Theoretical models are useful in that they can readily inform researchers of factors they may want to consider when investigating a specific behaviour. Nevertheless, with a significant difference between a health behaviour and on-farm behaviour, we contend that we may need to develop a tailored behavioural change model in this field rather than borrowing models that are developed for other purposes. This can only be achieved through acquiring more knowledge of farmers' behaviours and their behavioural changes using empirical qualitative and quantitative studies.

2.5. Discussion

Throughout this manuscript, we have highlighted knowledge gaps and limitations specific to two questions that need to be answered to model dynamic human behavioural changes: (1) the disease-related factors that are most relevant to motivate behavioural change, and (2) the quantitative association between a change in these disease-related factors and a change in a behaviour.

Here, we list six general challenges in veterinary epidemiology that we need to overcome to improve our understanding of human behaviour.

Challenge 1: Little focus on capturing farmers' true behaviours

Many current studies on farmer behaviour rely on questionnaire survey asking self-reported practice; however, this type of study needs a careful consideration because the discrepancy between self-reported and actual behaviour has been repeatedly identified (Conner and Norman, 2017; Racicot et al., 2012b; Rhodes et al., 2017). It may be the time to employ a more rigorous qualitative method such as biographical narrative interpretive method (McAloon et al., 2017; Wengraf, 2001)—a method to acquire interviewee's real-life experience—and quantitative studies using objective measures of farmers' behaviours (Cardwell et al., 2016; Millman et al., 2017; Racicot et al., 2011).

Challenge 2: Lack of empirical longitudinal data

We contend that employing theoretical psychology models to predict behaviour may be a useful quantitative tool but the validity of and assumptions behind models should be rigorously examined rather than merely applying a model to data (Hagger et al., 2017). We need more longitudinal studies that follow how actually farmers' attitude, perception, belief and behaviour change over time in response to various factors; not only disease-related but also wider animal welfare, environment, and economic factors because these can all lead to a change in farm biosecurity practices although improving biosecurity may not be farmers' primary purpose. There is much to learn from human health behaviour discipline, where various interventions to change human behaviours have delivered a mixture of success and failure (Prestwich et al., 2015). Analysis of increasingly available big data is also useful to validate findings from in-depth qualitative studies and provide a hypothesis on human behaviour patterns, which can be further investigated in qualitative studies. Such big-data analysis itself should be carried out by accounting for miscellaneous bias arising from human behaviours—data is essentially a product of, for instance, a decision to participate in a surveillance system and report a disease case (Bansal et al., 2016).

Challenge 3: Tendency to focus on a single disease

Another significant challenge is the development of a disease model that captures dynamics of multiple infectious diseases. Most available models simulate a single disease spread. However, spread patterns of each disease is not independent. As highlighted in this review, disease spread influences farmers' behaviours and trading patterns, which in turn will influence the spread of other infectious diseases. Modelling multiple diseases can be complex and computationally expensive: nevertheless, we do not need to simulate every single disease because humans cannot make a decision considering many complex factors (e.g. diseases) either. We need to understand farmers' decision making from their perspective.

Challenge 4: Barriers to interdisciplinary collaboration

Of course, the call for greater interdisciplinary working has been made by others working in the field of animal disease (Enticott and Wilkinson, 2013; Fish et al., 2011). Nevertheless, institutional boundaries and disciplinary norms can frustrate good intentions (Lowe and Phillipson, 2009), rendering interdisciplinarity an attractive but distant prospect. Potentially, as suggested in this review, a focus on the dynamic nature of human behaviour may provide both disciplinary and interdisciplinary methodological and theoretical challenges, in doing so creating a critical mass that overcomes barriers to interdisciplinary working.

Challenge 5: Gaps in framing behaviours between scientists and lay people

As it has been repeatedly pointed out in human health behaviour research, the use of theoretical psychology models, such as TPB, to identify only 'correlations' between psychological factors and implementation of behaviour would not substantially fill this knowledge gap: a causal model for behavioural change remains unknown (Conner, 2015; Sniehotta et al., 2014). Moreover, as highlighted in this review, the fundamental problem may be that we try to answer why farmers do not practice a certain behaviour, which we pre-defined. Farmers, however, frame behaviours differently from we do. What if farmers are aware of disease problem but implement their own 'biosecurity' practices they believe effective? This review clearly points out, from behaviour modelling perspective, that we lack understanding of 'how' farmers change their behaviours.

Challenge 6: Over-simplification vs. over-modelling

As highlighted throughout this manuscript, the dynamics of human behaviour can be challenging to model, especially when there is significant heterogeneity in behaviours between different groups of farmers. One may therefore argue that these complexities can be ignored as long as the model inferences are robust to sensitivity analysis. However, it should be noted that the most commonly used sensitivity analysis in veterinary epidemiology evaluates only the impact of parameter uncertainty and not the uncertainty in the model structure itself (Garner and Hamilton, 2011). Whether or not a specific dynamic behavioural component needs to be considered can be only evaluated by comparing inferences from models with and without the component, and this evaluation may be necessary for different diseases, populations, time-scales, and objectives of the study (Funk et al., 2015; Mancy et al., 2017; Probert et al., 2016). While we contend unnecessary complexities should be avoided, it is important to carefully evaluate if the simplicity of a given model adequately fits for the study objective (Evans et al., 2013; Sun et al., 2016).

2.6. Conclusion

An existing collaborative environment between scientists from veterinary epidemiology, animal welfare, and social science provides an exciting opportunity to provide a better understanding on behaviours and decision making of not only farmers, but also humans in general. At the same time, within the discipline of epidemiology itself, more theoretical studies that incorporate dynamic human behaviour and detailed infectious disease modelling continue to be necessary to identify behaviours that we should focus on understanding more. Studies should be self-critical about making unconscious and conscious assumptions—be it a behavioural study based on existing theories or a modelling study for an infectious disease spread—and discuss potential biases inherent to making such an assumption.

2.7. Acknowledgements

We would like to thank Tim Carpenter, Mark Stevenson, Lynsey Earl, JunHee Han, Sabrina Greening, Roman Konopka, Anna Alba Casals, Noboru Hidano, and Remo Job for stimulating discussions through which the basis of this review was formed. We thank two reviewers for their constructive comments that we believe improved the quality of this manuscript. Special thanks goes to Julian Drewe from Royal Veterinary College for having asked me if I was interested in submitting this work to the special issue of Frontiers in Veterinary Science and for providing very useful suggestions on this paper.

Chapter 3

Evaluating the efficacy of regionalisation in limiting high-risk livestock trade movements

Arata Hidano[1], Tim E. Carpenter[1], Mark A. Stevenson[2], M. Carolyn Gates[1]

[1] EpiCentre, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North 4442, New Zealand

[2] Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville 3010 Victoria, Australia

3.1. Abstract

Many countries implement regionalisation as a measure to control economically important livestock diseases. Given that regionalisation highlights the difference in disease risk between animal subpopulations, this may discourage herd managers in low-risk areas from purchasing animals from high-risk areas to protect the disease-free status of their herds. Using bovine tuberculosis (bTB) in New Zealand as a case example, we develop a novel network simulation model to predict how much the frequency of cattle movements between different disease control areas (DCAs) could theoretically change if herd managers adopted the safest practices (preferentially purchasing cattle from areas with the lowest risk of bTB), if herd managers adopted the riskiest practices (preferentially purchasing cattle from areas with the greatest risk of bTB), or if herd managers made trade decisions completely at random (purchasing cattle without consideration for bTB disease risk). A modified configuration wiring algorithm was used in the network simulation model to preserve key temporal, spatial, and demographic attributes of cattle movement patterns. The simulated frequencies of cattle movements between DCAs in each of the three behavioural scenarios were compared with the actual frequency of cattle movements that occurred between 1st July 2010 and 30th June 2011. Our results showed that the observed frequency of cattle movements from high-risk areas into low-risk areas was significantly less than if trade decisions were made completely at random, but still significantly greater than if herd managers made the safest possible trade decisions. This suggests that while New Zealand cattle farmers may have adopted risk-averse trading behaviour in response to regionalisation, there are other underlying factors driving livestock trade, such as established supplier-buyer relationships and heterogeneous individual perceptions towards disease risk, which may reduce the potential efficacy of regionalisation as a disease control strategy. Physical constraints and socio-psychological factors that determine herd managers' livestock trading behaviour warrant further studies to better understand how herd managers respond to future livestock disease regulations. The flexibility of a

network re-wiring framework presented in this study allows such a behavioural response to be incorporated into a disease simulation model, which will in turn facilitate a better evaluation of disease control strategies.

3.2. Introduction

3.2.1. Background

Many countries employ regionalisation as a measure for controlling economically important livestock diseases. This approach typically involves drawing geographical boundaries around subpopulations of farms with similar disease status and then imposing targeted control measures such as movement restrictions, testing, and/or vaccination, to minimise the risk of disease spreading from high-risk regions into low-risk regions (World Organisation for Animal Health OIE, 2015). Livestock trade within and between low-risk regions is generally unrestricted to minimise disruption to normal farming practices. The regionalisation approach is currently used as part of national disease control programmes for Johne's disease in Australia (Geraghty et al., 2014), bovine tuberculosis in New Zealand (Livingstone et al., 2006) and the United Kingdom (Bennett, 2009), and brucellosis in the United States (USDA, 2014).

As highlighted by analyses of national level livestock movement records, regionalisation can significantly alter livestock trading patterns (Vernon and Keeling, 2012). The change in trading patterns is, however, multifactorial. On one hand, negative pre-movement test results may provide incentives for some herd managers to purchase livestock from high risk areas and increase the frequency of high risk movement from high to low risk areas (Christley et al., 2011). One the other hand, the frequency of high risk movement may reduce because some herd managers would be inclined to send livestock directly to slaughterhouse to avoid pre-movement testing costs (Bennett, 2009). The need for testing should also remind herd managers of the risk of disease introduction and this may discourage herd managers from purchasing livestock from high risk areas (Christley et al., 2011). Despite this complexity herd managers are, in general, risk averse (Botterill and Mazur, 2004; Valeeva et al., 2011) and there is some evidence that the frequency of high risk movement in the United Kingdom reduced after the introduction of regionalisation (Gates et al., 2013).

Although regionalisation may encourage herd managers' risk-averse (i.e. non-risky) trading behaviour, market opportunities are limited and herd managers have an inherent need to move livestock when sale prices are at a premium. These limited opportunities may in turn constrain how herd managers can alter their trading patterns in response to regionalisation. Should herd managers not have options for a feasible alternative trading pathway, regionalisation might not affect the livestock movement patterns. The impact of regionalisation on reducing the frequency of high risk movement should be therefore evaluated accounting for these limitations. By developing a novel network rewiring model, we quantified how much the livestock movement pattern can actually vary under these constraints, using regionalisation established in New Zealand for bovine tuberculosis control as a case example.

3.2.2. Bovine tuberculosis in New Zealand

Regionalisation was first introduced to New Zealand in the mid-1990s to aid in the control of bovine tuberculosis (bTB) (Livingstone et al., 2015a). The country is divided into Disease Control Areas (DCAs) that are assigned into one of five categories based primarily on the perceived risk of bTB spreading to livestock herds through contact with infected local wildlife populations. The DCA categories include special testing triennial (STT), special testing dairy (STD), special testing biennial (STB), special testing annual (STA), and movement control areas (MCA). For simplicity, we subsequently refer to these as Area 1a, 1b, 2, 3, and 4, respectively, with the higher number indicating a higher perceived risk of bTB transmission from wildlife. Areas 1a and 1b are considered to have an equally negligible risk. The boundaries of DCAs are defined and reviewed annually and details of each DCA can be found elsewhere (Ryan et al., 2006; Buddle et al., 2015). Figure 3-1 shows the New Zealand DCA boundaries as of 2011.

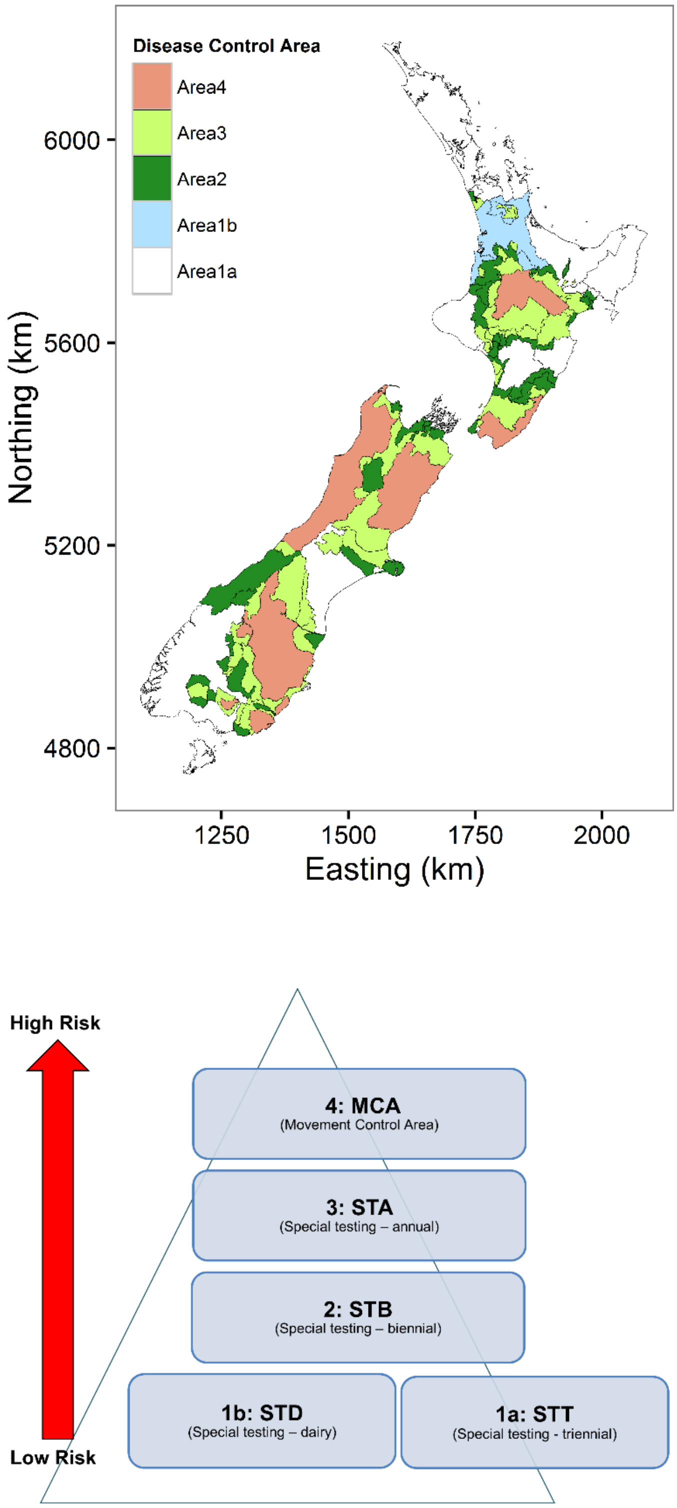


Figure 3-1 Map of New Zealand showing the distribution of 5 disease control areas in the study period (top) and diagram showing the relative levels of bTB transmission risk from wildlife in each disease control area (bottom).

As summarised in Table 3-1, each DCA has a different bTB testing and control regime that scales in intensity according to the disease risk. Cattle and deer moving off farms in Area 4 where the perceived risk of wildlife transmission is the highest must be tested within 60 days prior to the movement (Buddle et al., 2015). Previous surveys on the perception of New Zealand farmers towards regionalisation suggest that herd managers in low-risk areas recognise that purchasing livestock from high- risk areas carries an increased risk of introducing bTB to their herd, whereas herd managers in high-risk areas are either less aware of this risk or have a tendency to preferentially purchase animals from high-risk areas because of their discounted price (Sauter-Louis, 2001; Corner, 2002). If these perceptions translated into practice, we would expect to a see a reduced frequency of movements from high-risk regions into low-risk regions and an increased frequency of movements within high-risk regions compared to the patterns that may be expected if herd managers traded completely at random.

Table 3-1 Details of each Disease Control Area (DCA) at the time of the study period

Name of DCA	Name used in this study	Descriptions	Testing policy
Movement control area	Area 4	Located in VRA (vector risk area), where the presence of bTB infected wildlife has been confirmed or strongly suspected	Annual testing of cattle over 3 months and deer over 8 months + pre-movement testing of all animals over 3 months
Special testing – annual	Area 3	Located in low-risk parts of VRA or neighbouring parts of VFA (vector free area), where wildlife populations are deemed free from bTB infections	Testing cattle over 12 months and deer over 15 months every year
Special testing – biennial	Area 2	Located outside and surrounding of the Special testing - annual	Testing cattle and deer over 24 months every 2 years
Special testing – dairy	Area 1b	Established in Waikato, where bTB transmission risk from wildlife is negligible, as a special voluntary testing area to minimise bTB spread in the region	Testing dairy cattle over 24 months every 2 years and beef cattle and deer over 24 months every 3 years
Special testing – triennial	Area 1a	Located in the parts of VFA, where bTB transmission risk from wildlife is negligibly low	Testing cattle and deer over 24 months every 3 years

3.2.3. Objectives

In this analysis, we developed a novel network re-wiring algorithm that allows us to explore the range of possible movement patterns that could emerge under three different trading behaviour scenarios: (1) the 'safe' scenario where farms in low-risk regions preferentially source cattle from the lowest risk DCAs to prevent bTB introductions, (2) the 'risky' scenario where farms in low-risk regions source cattle from the highest risk DCAs to capture price advantages, and (3) the 'random' scenario where farms make trade decisions without considering the DCA origin of purchased cattle. The results from the re-wired networks were compared with the observed network of movements to determine how effective regionalisation has been in reducing the frequency of high-risk cattle movements in New Zealand.

3.3. Materials and methods

3.3.1. Data collection and processing

3.3.1.1 Cattle movement data

Movement of dairy cattle in New Zealand are captured by the New Zealand Dairy Core Database since 2001, which is legislated under the Dairy Industry Regulations 2001 (Anon, 2001). This database is managed by the Livestock Improvement Corporation (LIC), which is one of the largest dairy farming co-operatives in New Zealand. In this analysis, we used data from 1st July 2010 to 30th June 2011to characterise dairy cattle movement patterns within and between disease control areas. During this time period, approximately 97% of commercial dairy herds were active to record various individual animal and herd-level demographic data (our estimate based on data provided by DairyNZ, 2011).

The relevant individual animal-level variables included the unique animal identification number, date of birth, date of introduction to a registered farm location, date of removal from a registered farm location, and reason for removal (death, culling, or movement). It should be noted that movements in this dataset exclusively mean farm-to-farm transfers and do not include movements to slaughterhouses. For each movement, information was recorded on whether the ownership of the animal has changed, which allowed us to distinguish movements for sales from movements for other purposes such as grazing. No information was available on whether the movements for sales occurred through a livestock market or occurred directly between cattle farms. Animals were further classified into three production groups based on their age at time of movement: calves (< 13 months old), heifers (13 to < 25 months old), or adults ($\geq$25 months old).

The relevant herd-level information included the unique herd identification number, geographical coordinates of the farm, geographical region of the farm (16 regions), and participant code (i.e. identifier for herd owner). In New Zealand, a single participant may own multiple uniquely identified herds. The number of calves born in the previous production season (i.e. between 1st July 2009 and 30th June 2010 and hereafter referred to as the 2009 season) was used as a proxy for herd size. The number of cows culled in the 2009 season was used as a proxy for herd performance, although further studies are required to identify whether or not the number of cows culled in a herd would represent the herd performance. Herds were further classified as being "new" if they first appeared in the database during the study period or "existing" if they were present in the database in or prior to the 2009 season. Using the geographical coordinates, farms were assigned into their respective DCA category (1a, 1b, 2, 3, or 4).

During the study period, there were a total of 801,891 individual animal movements between 11,402 unique herd locations. These herds were owned by 10,891 distinct individuals with ownership information missing for only one herd. Since the primary objective of this study was to evaluate dairy cattle trading patterns, we excluded all movements where there was no change in ownership. Thirteen animal movements that involved one farm with unknown ownership were also removed. The remaining 666,191 movements were aggregated into batch movements such that all animals moved from a given source farm to a given destination farm on the same date were considered a batch. If the batch contained animals from multiple age classes, we selected the predominant age class to represent the batch type. This resulted in 33,634 batches with the following covariates: movement date, number of cattle per batch (i.e. size) and age classification.

On further inspection of the data, there were 6,763 instances where movements occurred back and forth between the same pair of farms during the study period. These movements were likely to be off-farm grazing practice and were therefore excluded. Furthermore, there were 2,643 instances where multiple batches of the same animal type were moved between the same source and destination farms over a short time period (≤ 31 days). We assumed that this behaviour represented a single sales transaction where the animals could not be transported at the same time due to logistical reasons (e.g. transport vehicle availability). We aggregated these movements into a single batch, as described in the Section 1 in Appendix 1. This resulted in a dataset that contained 24,530 batches, which was used for modelling movement distances as described in the Section 3.3.1.2. Among this dataset, 1,087 batches that repeatedly moved over a short distance (<10 km) were removed and the remaining 23,443 batches were used in a network rewiring process (described in the Section 3.3.2) as explained in the Section 3 in Appendix 1. A flow chart that describes the selection process of movements eligible for analyses is shown in **Figure 3-2**.

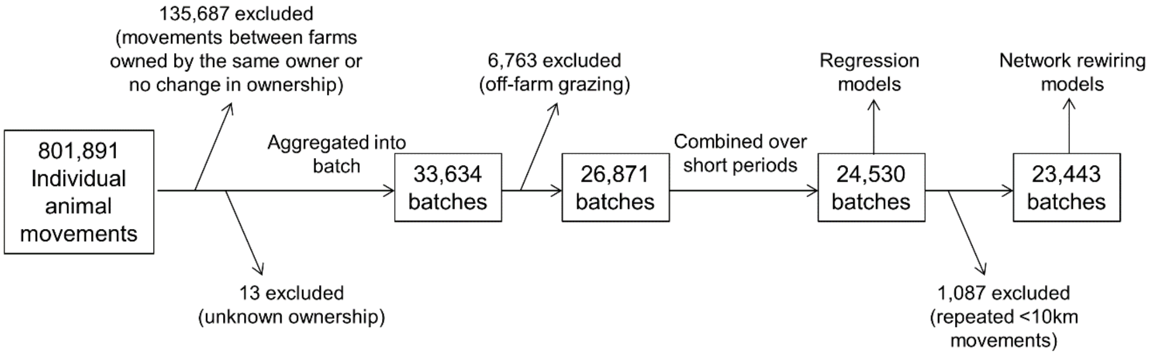


Figure 3-2 Flow chart showing the selection process of movements eligible for analyses.

3.3.1.2 Modelling movement distances

Livestock movements are constrained by geographical distance between farms due to transport cost and animal welfare reasons (Miranda-de la Lama et al., 2014), with local movements occurring more frequently (Mitchell et al., 2005; Sanson, 2005). Movement distances could also be affected by various factors, including animal type, farm region, and season (Lindström et al., 2009; Sanson, 2005). To generate a realistic and stochastic movement distance distribution, we developed a regression model that predicted the movement distance for each batch based on the characteristics of the batch and source farms including the geographical region of farms as described in Appendix 1.

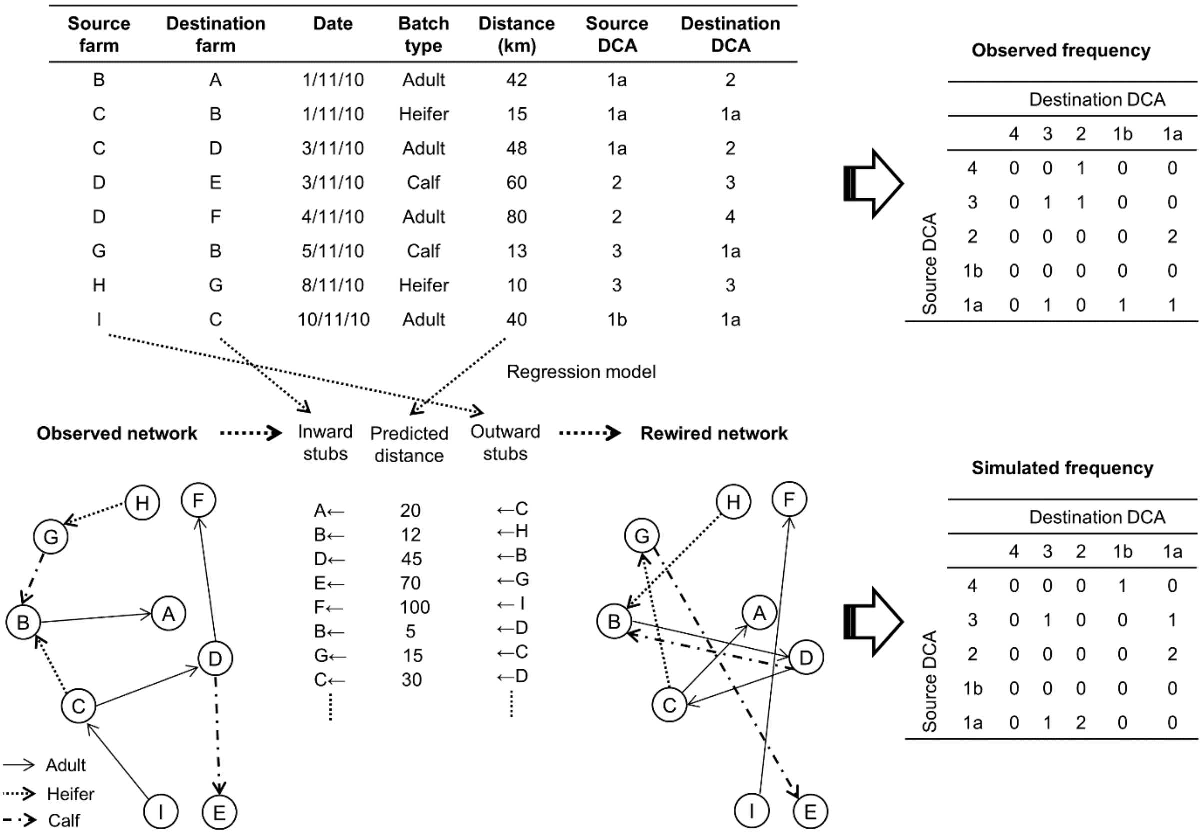
3.3.2. Network rewiring model

To test our hypothesis that herd managers in low-risk areas preferentially purchase cattle from DCAs with equal risk, we developed a network generation model to explore the range of possible movement patterns that could be produced under different behavioural scenarios. Our model used a modified configuration wiring algorithm to preserve important spatial, temporal, and demographic characteristics of the observed cattle movement network during the reconstruction process (Molloy and Reed, 1998; Serrano and Boguna, 2005; Gates and Woolhouse, 2015). This approach further allowed us to adjust the spatial clustering effect of livestock movement, where movements occur, in general, more frequently between farms in close proximities (Mitchell et al., 2005; Sanson, 2005). Unless adjusted, this would create a confounding effect whereby movements occur more likely between farms in the same DCA because farms in close proximities are more likely to be in the same DCA.

Following standard network terminology, we refer to the farms in our network as “nodes” and the movements between them as “edges” (Martínez-López et al., 2009). In the simplest configuration wiring model, each node is assigned a set of virtual objects or “stubs” corresponding to the number of connections it has in the original network. Pairs of stubs are then randomly chosen and connected to form edges until no more stubs remain. Since cattle movements have clear directionality, each farm in our model was assigned a fixed number of inward stubs corresponding to the number of batches received (in-degree) and a fixed number of outward stubs corresponding to the number of batches sent (out-degree). In our model, each stub therefore represented a single batch with the attributes of farm id, batch size, movement date, and batch type (age classification).

At the start of each rewiring process, the lists of inward and outward stubs were sorted at random to ensure stochastic variation in the simulation replicates. Connections were performed by working through the list of inward stubs sequentially from top to bottom until no more stubs remained. For each inward stub, we searched through the list of remaining outward stubs to identify ones that met the following criteria: (1) outward and inward stubs had the same batch type, (2) the movement date for outward stub was the same as or within 7 days of that of a given inward stub, and (3) the source and destination farms were different. The second rule allowed for more stochastic variation in the range of possible trade networks that could be produced through re-wiring. To ensure this 7-day time window does not influence the simulation result, we run the model using 14-day time window and confirmed that both results were almost identical. We then used a hurdle regression model to predict the expected distance for this batch and selected the outward stub that produced the closest possible match to this distance. This allowed us to preserve important spatial features of the trade network.

The matched inward and outward stubs were then removed from the lists and the process was repeated until no more matches could be found. This occasionally resulted in the presence of inward stubs that could not be matched to any outward stubs. Since the total number of unmatched stubs was always less than 20 (<0.1% of total movements) at any given re-wiring replicate, we considered that this would have negligible impact our conclusions and removed unmatched stubs from the analysis. A diagram summarizing the network re-wiring process is shown in **Figure 3-3**.



Source farm	Destination farm	Date	Batch type	Distance (km)	Source DCA	Destination DCA
B	A	1/11/10	Adult	42	1a	2
C	B	1/11/10	Heifer	15	1a	1a
C	D	3/11/10	Adult	48	1a	2
D	E	3/11/10	Calf	60	2	3
D	F	4/11/10	Adult	80	2	4
G	B	5/11/10	Calf	13	3	1a
H	G	8/11/10	Heifer	10	3	3
I	C	10/11/10	Adult	40	1b	1a

Observed frequency

Source DCA	Destination DCA 4	3	2	1b	1a
4	0	0	1	0	0
3	0	1	1	0	0
2	0	0	0	0	2
1b	0	0	0	0	0
1a	0	1	0	1	1

Inward stubs	Predicted distance	Outward stubs
A←	20	←C
B←	12	←H
D←	45	←B
E←	70	←G
F←	100	← I
B←	5	←D
G←	15	←C
C←	30	←D

Simulated frequency

Source DCA	Destination DCA 4	3	2	1b	1a
4	0	0	0	1	0
3	0	1	0	0	1
2	0	0	0	0	2
1b	0	0	0	0	0
1a	0	1	2	0	0

Figure 3-3 Diagram showing the procedure of the network rewiring model and counting the frequency of movement within and between DCAs.

Note that lines in the diagram are not drawn to scale.

3.3.3. Re-wiring scenarios

The network re-wiring model was used to reconstruct movement patterns under three different herd managers' behavioural scenarios (**Figure 3-4**):

1. Random selection scenario
2. Risky scenario
3. Safe scenario.

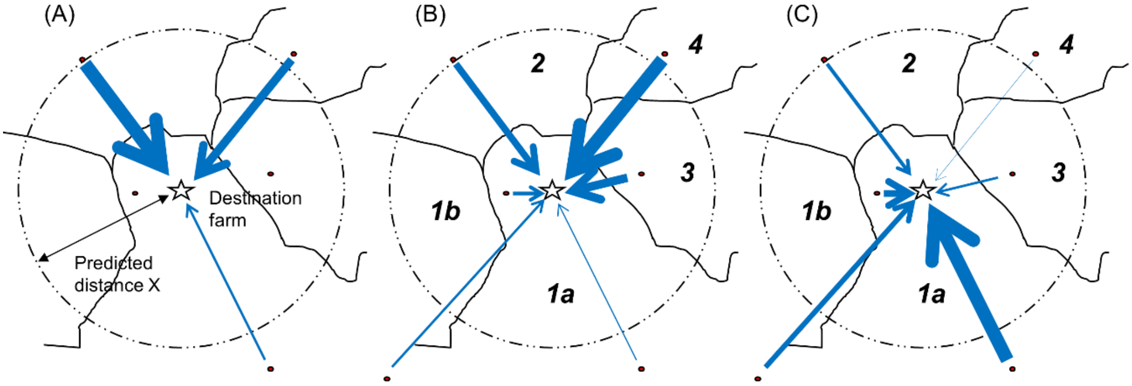


Figure 3-4 Schematic representation of each network rewiring scenario; (A) random selection, (B) risky, and (C) safe scenario.

Location of a destination farm is indicated by Star and the radius of a circle represents a predicted distance from the hurdle models. Red dots are candidate source farms in different DCA areas, which are shown by the numbers in italic. The width of lines represents the order of selection of source farms in the network rewiring model (the wider the higher priority).

In the 'random selection' scenario, herd managers purchased cattle based solely on distance from the source farm with no regard for regional differences in disease risk. In the 'risky' scenario, we explored what would happen if herd managers in low-risk areas (Area 1a) preferentially purchased cattle from the highest-risk bTB area possible. For each inward stub for farms in Area 1a, we selected an outward stub from Area 4 that was the closest possible match to the movement distance predicted by the hurdle regression model. If no eligible match was available from Area 4, we worked sequentially through Area 3, Area 2, and Area 1(a and b) to find the highest risk match possible. In the ‘safe’ scenario, we assumed that farms in Area 1a preferentially purchased cattle from the lowest risk region possible. For each inward stub for farms in Area 1a, we selected an outward stub from Area 1 (a and b) that was the closest possible match to the movement distance predicted by the hurdle regression model. If no eligible match was available from Area 1, we worked sequentially through Area 2, Area 3 and Area 4 to find the lowest risk match possible. To capture stochastic variation in the possible networks, we repeated the re-wiring process 1,000 times for each of the three scenarios. The simulation models were implemented in the C programming language and checked to ensure there were no coding errors. The C code used for the ‘random selection’ scenario can be found at the first author’s repository (https://github.com/arata-hidano/Network_rewire_model_Ccode).

For each of 1,000 iterations, we counted the frequencies of movements that occurred between the different possible combinations of source and destination DCAs. The median and 95% probability intervals were reported. We then compared the frequencies obtained under the three different simulation scenarios against the real data to determine where herd managers in New Zealand fell in the spectrum of risk. The results from the safe and risky scenarios were used to determine the lower and upper limits.

3.4. Results

3.4.1. Descriptive statistics of movements eligible for rewiring

From 1st July, 2010 to 30th June, 2011 there were 10,233 dairy cattle farms that had at least one recorded movement on or off the farm premises. Of these 2,285 farms only received batches, 4,394 farms only sent batches, and 3,554 farms both received and sent at least one batch during the period. Of 23,443 batch movements eligible for rewiring, 13,593 were adult cattle, 5,559 were heifers, and 4,291 were calves. The median and mean movement distances for batch movements were 47 and 130 km, respectively, with minimum and maximum distances being 0 and 1309 km, respectively. As shown in **Figure**

3-5, the frequency of batch movements peaked at the end of the production season (i.e. May and June). **Figure 3-6** shows geographical cattle movement patterns over 16 New Zealand regions in 4 selected months of the early (August), middle (February) and end (May and June) of the production season. The Waikato region served as a sink throughout year, whereas the Canterbury region served as a source at the end of the season (see Figure S1 in Appendix 1 for the location of each region in New Zealand). There was also a high volume of trade between the North and South Islands despite their geographical separation. More detailed descriptive statistics on trade patterns are presented in Table S1 and Figure S7 in Appendix 1.

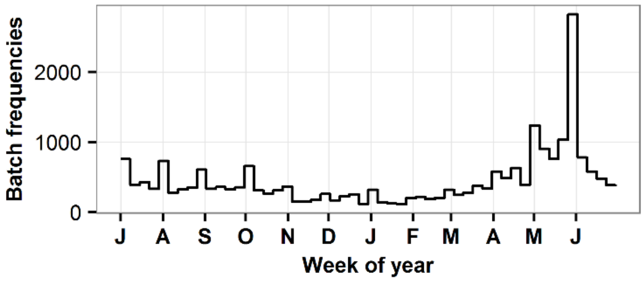


Figure 3-5 Distribution of the batch movement frequencies by week from 1st July, 2010 to 31st June, 2011.

Batch movements were aggregated over 16 regions, whose boundaries are shown in white and locations are indicated by circles. Curve lines represented movements from north towards south in direction (red) and the opposite (blue). Line widths represented the frequency of batch between a given pair of regions. Circle size represented the total number of batches received in each region in a given month. Circle colour represented the proportion of within-region movements to the total movements received in a given month.

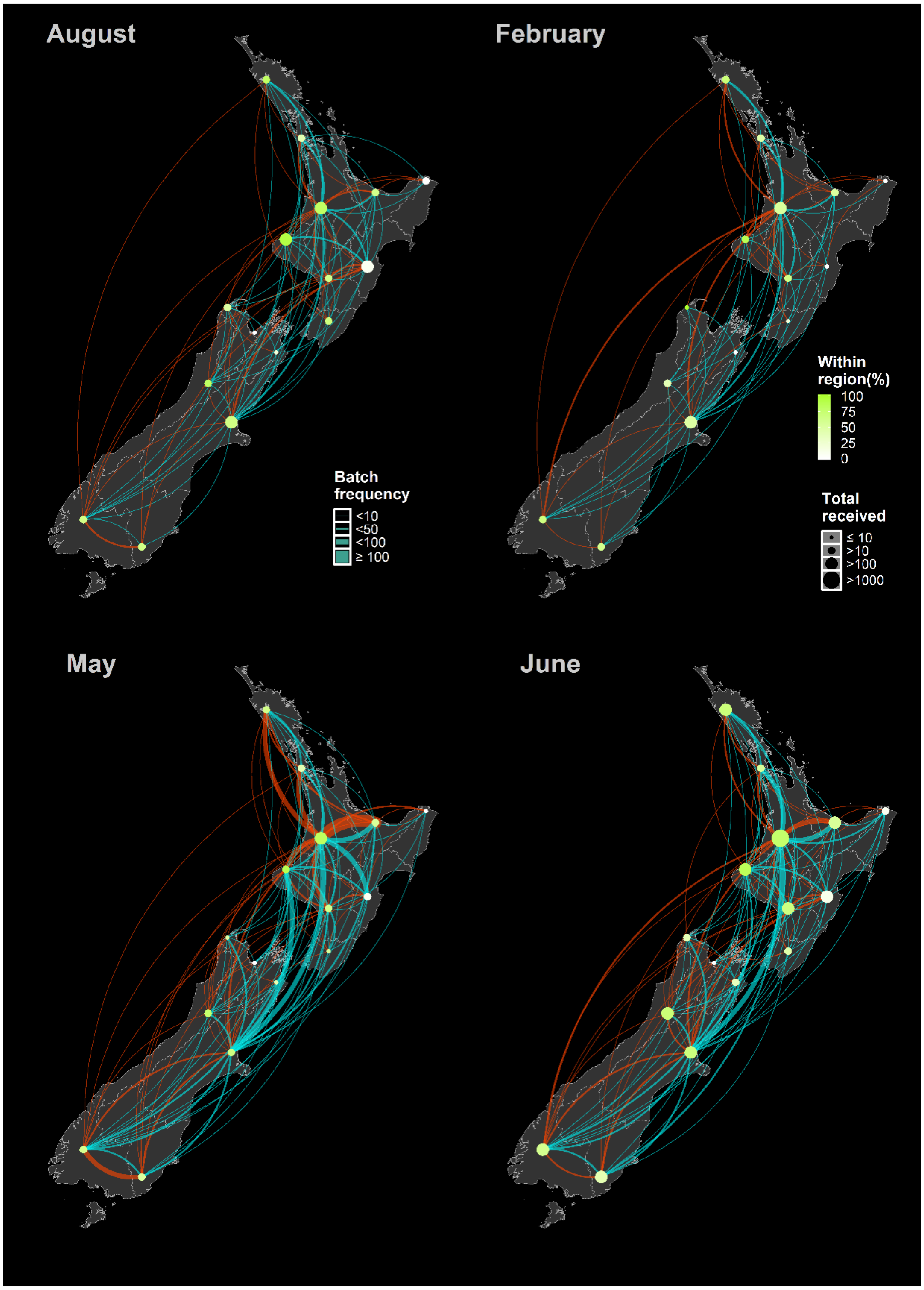


Figure 3-6 Maps of New Zealand showing dairy cattle movement patterns retrieved from Livestock Improve Corporation in August 2010, February, May, and June 2011.

3.4.2. Comparison between the observed and simulated movement distances

Figure 3-7 shows the distance distribution of the observed movement and that of the network re-wiring model under the random scenario. The distance distributions for the safe and risky scenarios are shown in Figure S9 in Appendix 1. These suggest that the network re-wiring model generally captured the observed movement distance distribution, with the exception of slightly underestimating the frequency of short distance movements (≤20km).

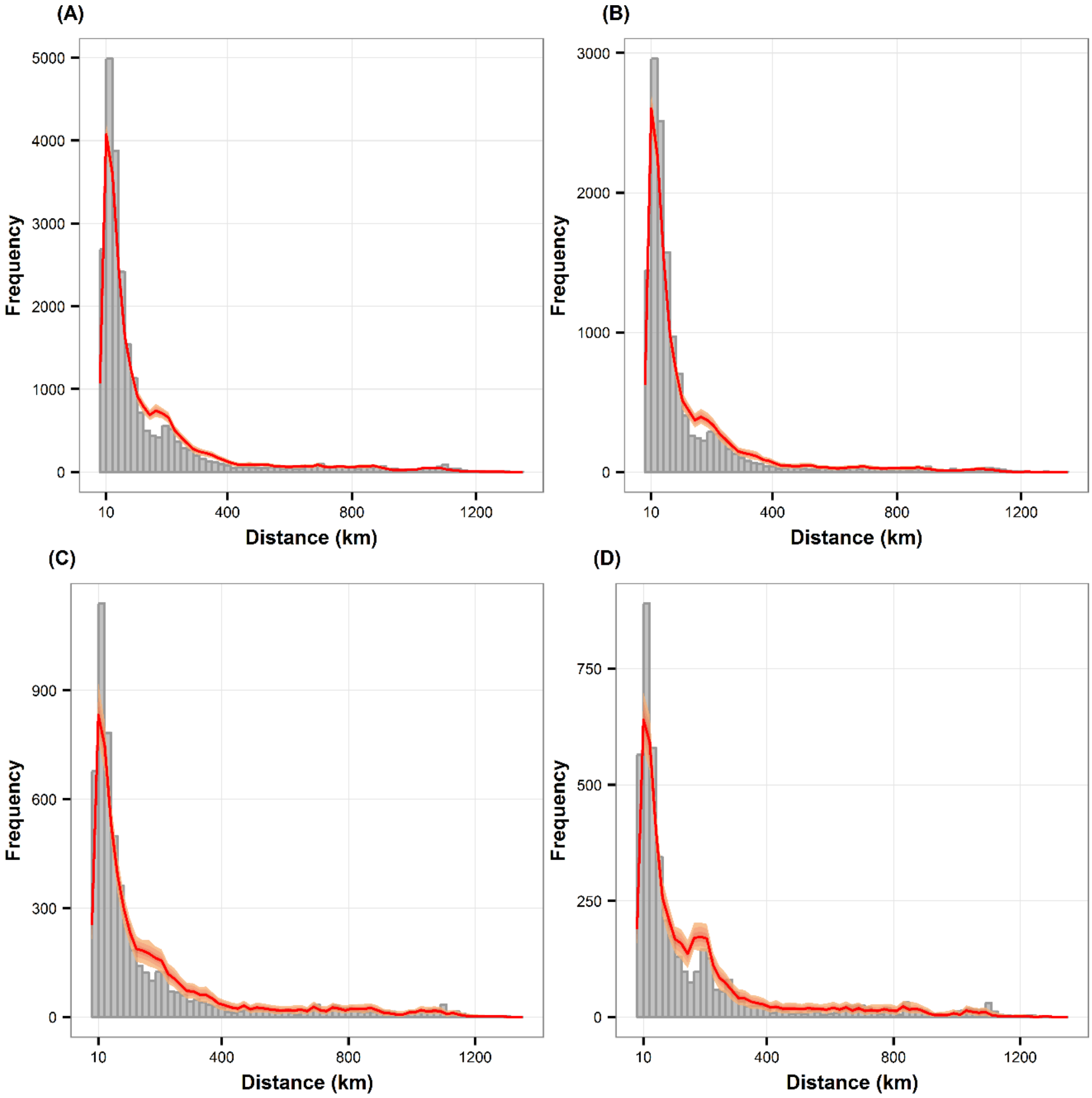


Figure 3-7 Distributions of the observed and simulated batch movement distances.

The histograms show the observed frequency for (A) all batch types, (B) adult, (C) heifer, and (D) calf movements. Red lines show the median frequency of corresponding batch types obtained from 1000 simulations of the network rewiring model under random selection scenario. Shaded strips around the median line were coloured proportional to the probability of frequency at each point occurring over 1000 simulations (Jackson, 2008).

3.4.3. Observed and simulated movement frequencies between DCAs

Table 3-2 summarises the observed frequencies of movements within and between DCAs in New Zealand. The highest frequency ($n = 6{,}066$) was observed for movements within Area 1a, whereas the lowest ($n = 77$) were those from Area 4 to Area 1b. For movements into Area 1a from Areas 2, 3, and 4, the observed frequencies were smaller than the 2.5th percentile of simulated distributions from the network rewiring model under the random selection scenario, whereas the observed frequency exceeded the 97.5th percentile of a distribution for movements within Area 1a (Table 3-3). These results suggested that movements from higher bTB risk areas to Area 1a occurred less frequently than expected if herd managers traded cattle without regard to risk.

Table 3-2 The observed movement frequencies within and between DCAs.
Note that 276 batches were not included in this table because their source, destination or both DCA status were unknown.

	Destination DCA				
Source DCA	Area 4	Area 3	Area 2	Area 1b	Area 1a
Area 4	431	192	100	77	148
Area 3	259	1152	481	831	832
Area 2	168	612	680	517	729
Area 1b	273	1018	774	3634	1154
Area 1a	246	1085	833	875	6066

Figure 3-8 shows where the observed frequencies of movement towards Area 1a lie in the spectrum of the lower and upper limits of movement frequencies that were reproduced in the safe and risky scenarios, respectively. Figure 3-8 (A), (B), (C), and (E) show that the observed frequencies were between those simulated under the random and safe scenarios and farther away from that simulated under the risky scenario. We note that the observed frequency within Area 1a ($n = 6{,}066$) was as high as the 2.5th percentile of a simulated distribution under the safe scenario ($n = 6{,}156$). In contrast, the observed movement frequency from Area 1b to Area 1a was between

simulated frequencies under the risky and random scenarios (Figure 3-8 (D)). Simulated frequencies for all 25 DCA combinations are provided in Table S5, Table S6 and Figure S8 in Appendix 1.

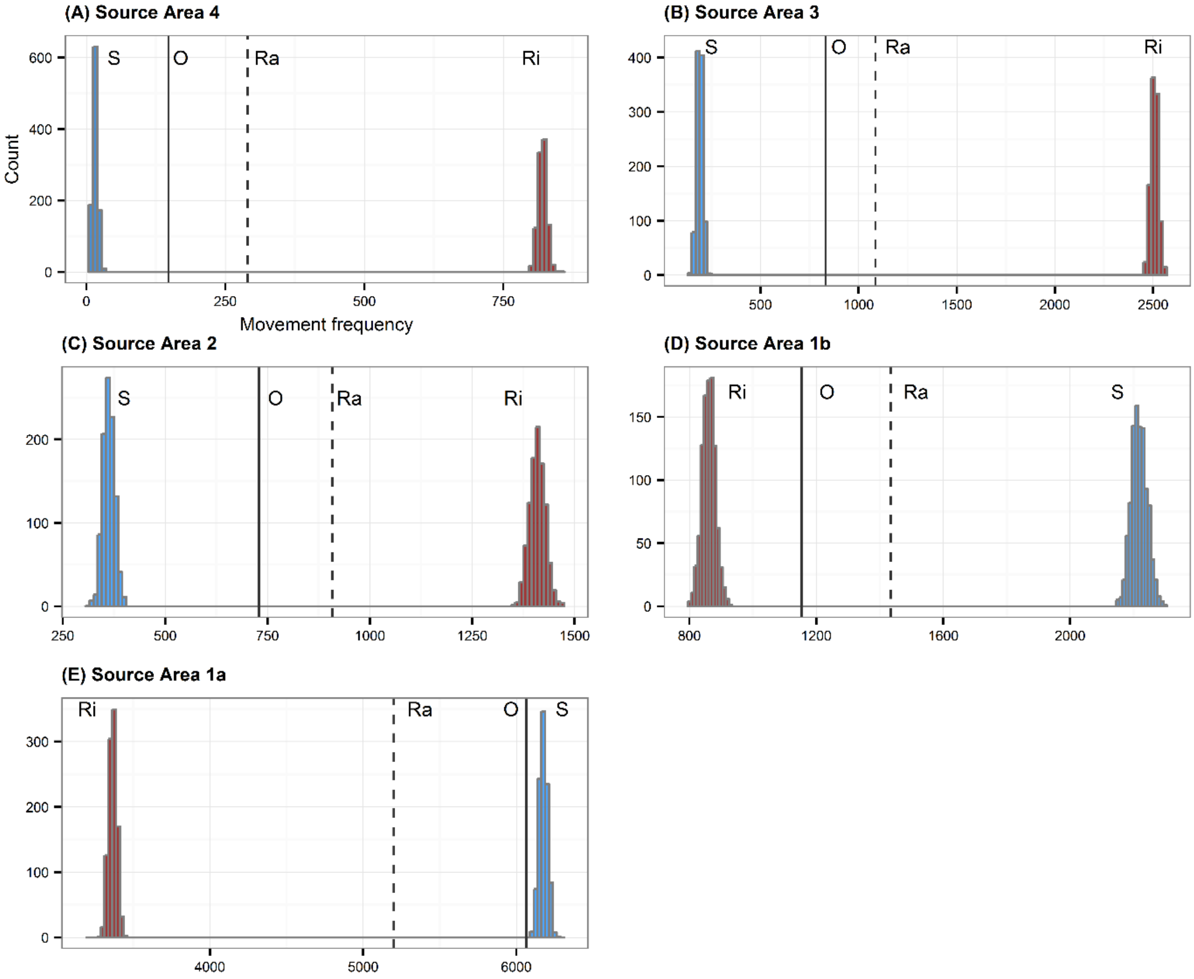


Figure 3-8 Comparison of the observed and simulated frequencies of movement towards Area 1 from (A) Area 4, (B) Area 3, (C) Area 2, (D) Area 1b, and (E) Area 1a.

Solid lines annotated as "O" represent the observed frequency. Dashed lines annotated as "Ra" represent the median frequency obtained from 1000 simulations of the network rewiring model under the random selection scenario. Distributions of simulated frequencies under the safe and risky scenario are shown by blue (annotated as "S") and red (annotated as "Ri") histograms.

Table 3-3 The median movement frequencies within and between DCAs obtained over 1000 simulations of the network rewiring model under the random selection scenario.

Numbers in the bracket represent the 2.5th and 97.5th percentiles of the simulated distributions.

	Destination DCA				
Source DCA	Area 4	Area 3	Area 2	Area 1b	Area 1a
Area 4	267 (248, 285)[a]	184 (164, 203)	143 (125, 163)[b]	80 (66, 94)	290 (268, 313)[b]
Area 3	260 (239, 282)	858 (822, 896)[a]	485 (453, 515)	859 (823, 896)	1086 (1044, 1128)[b]
Area 2	202 (182, 224)[b]	564 (533, 597)[b]	524 (492, 557)[a]	505 (471, 539)	908 (870, 948)[b]
Area 1b	256 (236, 277)	1143 (1107, 1183)[b]	726 (692, 759)	3284 (3233, 3340)[a]	1434 (1390, 1480)[b]
Area 1a	395 (369, 422)[b]	1310 (1269, 1355)[b]	993 (958, 1030)[b]	1195 (1151, 1242)[b]	5200 (5142, 5262)[a]

[a] 97.5th percentile of the distributions calculated over 1000 simulations was lower than the real observed data

[b] 2.5th percentile of the distributions calculated over 1000 simulations was higher than the real observed data

3.5. Discussion

By developing a network rewiring model that preserved key spatial, temporal, and demographic attributes of cattle movement patterns, we were able to explore the range of the New Zealand dairy cattle movement patterns that could theoretically be produced if herd managers changed their trading behaviour in response to regionalisation. The study results confirmed that herd managers in low risk area purchased cattle from farms with higher risk less frequently than expected if between-farm distance were the only factor influencing the trade. We also showed that the geographical proximity does not fully explain why herd managers in low risk areas purchased cattle more frequently from farms with the same risk status rather than farms in higher risk areas. Collectively, these together suggest that herd managers in Area 1a (low risk) performed a risk-averse cattle purchasing behaviour. This result is supported by previous qualitative surveys conducted on cattle farmers in New Zealand (Corner, 2002; Sauter-Louis, 2001). Herd managers in low risk areas may avoid purchasing cattle from higher risk areas because they are well aware of the bTB transmission risk associated with purchasing livestock from high risk areas (Corner, 2002). The tendency to avoid purchasing high bTB risk livestock is also reported for herd managers in the United Kingdom (Enticott et al., 2015b). On the other hand, herd managers in high risk areas might be less hesitant to buy cattle from high risk areas. Sauter-Louis, (2001) reported that some herd managers in high risk areas purchased livestock from bTB-infected farms because these animals were cheaper and they did not consider this behaviour as risky. Such price discount of livestock with high bTB-infection risk might occur due to risk-averse behaviour of other herd managers (Lambie, 1998). Along with the extensive wildlife control and livestock surveillance, we speculate that the suggested risk-averse trading behaviour of herd managers has contributed to the substantial reduction in bTB prevalence in New Zealand over last two decades (Livingstone et al., 2015a, 2015b). Such behaviour means that cattle from high bTB-risk farms are more likely to be purchased by other farms in high risk areas. Therefore, these cattle will be tested more often because of higher testing frequency in high bTB-risk areas. This phenomenon is also discussed in a recent study in the United Kingdom (Adkin et al., 2016).

Despite the equal perceived bTB risk in Area 1a and 1b, herd managers in Area 1b did not follow the same risk-averse trading patterns as those in Area 1a; the observed batch frequencies towards Area 1b from Area 4, 3, and 2 were within the 95% probability interval (PI) of the simulated distributions (Figure S8 in Appendix 1). This is likely to be due to the presence of a livestock dealer with a large number of movements. Out of 1275 farms in Area 1b that purchased at least one batch, the 10 farms that most frequently purchased cattle from each of Areas 4, 3, and 2 contributed to 44%, 22%, and 27% of the total movements towards Area 1b from each corresponding area, respectively. A similar pattern existed in Area 1a among 2547 farms that purchased at least one batch; however, the percentages were lower at 28%, 17%, and 17%, respectively (Figure S11 in Appendix 1). The presence of such dealers in the Waikato region (which occupies the majority of Area 1b) has been known for a long time, however, their role in disseminating bTB are highly controversial in New Zealand (Ryan et al., 1997). Given some of these dealers are registered to exist in low risk areas and the lack of availability of detailed movement history, livestock sold by them may be considered by buyers to be from low risk areas, when they are from high risk areas. This may offset the benefit of risk-averse trading behaviour of farms in terms of containing disease. The improvement in the coverage of National Animal Identification and Tracing scheme and its efficient use are expected to fill this gap in future (Jewell et al., 2016).

In this study, dairy cattle movement patterns were reconstructed using a network re-wiring approach based on the configuration wiring algorithm (Serrano and Boguna, 2005; Gates and Woolhouse, 2015). This flexible framework allowed us to explore a range of plausible networks that could form under different behavioural scenarios while preserving important features of livestock trade; in- and out-degree of each farm, batch type, and spatial and temporal characteristics of each movement. Other approaches to reconstructing contact networks have focused on reproducing global properties of the network (e.g. overall degree distributions, scale-free network properties, and assortativity) at the expense of losing local properties (e.g. node- and edge-level characteristics) that are important in determining how contacts form in the real world (Guimera et al., 2007). In the context of cattle movement networks, these local properties reflect the dynamic decisions of individual herd managers regarding how many animals to trade, the optimal time of year to trade them, and who to select as trading partners. These decisions are in turn based on complex factors such as farm demographics, herd productive and reproductive performance, long-term business goals, social preferences, and strategies for mitigating disease risk (Christley et al., 2011; Gates, 2014). Further studies are warranted to understand how these trade decisions are made and affected by various factors on farm, so that we can develop more sophisticated models of behavioural response to animal health interventions. It is planned to continue ongoing extensive wildlife control with the aim to eliminate bTB from the wildlife population in New Zealand by 2055 (Livingstone et al., 2015b). This will change the DCA status of farms, which in turn will likely alter the trading behaviour of herd managers and hence livestock movement patterns. Ultimately, the efficacy of potential bTB control strategies among livestock populations should be evaluated using a disease simulation model that accounts for such behavioural feedback (Tago et al., 2016).

To replicate a realistic and stochastic movement distance distribution, we used a hurdle regression model, based on various herd-level demographics and batch-level characteristics, to predict the likely movement distance for each batch of cattle sold. This approach is in contrast to that where each movement distance is assumed to be a random draw of an underlying distance distribution aggregated over populations (e.g. all movements). Our approach was employed out of recognition that short distance movements likely occur by different processes than long distance movements (Lindström et al., 2009;

Sanson, 2005). The importance of explicitly accounting for a movement distance for each batch is highlighted by our finding; although, the random selection and the safe scenario produced almost identical overall distance distributions (Figure S9 in Appendix 1), movement frequencies between DCAs were different between these two scenarios (Figure S8 in Appendix 1). These suggest that a movement pattern reconstructed based solely on an overall distance distribution may fail to capture an important epidemiological feature—between and within DCA frequencies in our example—of the actual movement pattern. We also note that the information whether movements occurred through a market was unavailable for this study, although this likely affects movement distances (Sanson, 2005). However, we reason that the impact of this misspecification is minimal; it was reported that movements directly between farms occurred approximately as 30 times frequent as movements through a market in New Zealand (Sanson, 2005). Nevertheless, further understanding on which and why herd managers choose a specific trading channel (e.g. market, direct farm-to-farm trade) should improve the prediction of future livestock movement patterns.

Our network re-wiring model made the simplified assumption that herd managers seek their trading partners solely based on distance as the first criterion. Given transportation accounts for a large part of the logistical costs associated with livestock trading (Miranda-de la Lama et al., 2014), between-farm distance is likely to be a significant factor. It is however increasingly becoming evident that farmers do not make a decision based solely on financial criteria (Garforth, 2015). One alternative approach to avoiding this assumption was to use an exponential random graph model, which is a statistical model developed for social networks to predict the presence of edges or the values associated with edges between nodes, based on various covariates including edge- and node-level factors (Robins et al., 2007; Krivitsky, 2012). However, although ERGM theoretically allows us to examine if DCA affected the movement patterns while adjusting the abovementioned spatial clustering effect (Lusher et al., 2012), this was not feasible for our dataset because this method is not appropriate for large datasets (Goodreau, 2007; Danon et al., 2011).

Other unexamined factors involved in motivating herd managers' trading behaviour could confound our findings, should these factors be also associated with the source farm DCA status. For instance, herd managers in a high risk area may sell a small batch size to minimise the probability of an animal in the batch being detected as bTB-infected in a pre-movement test (Christley et al., 2011), and this might limit a spectrum of farms that wish to buy these animals. There was, however, no evidence that the batch size sold from Area 4 was significantly smaller than those from other DCAs in this study. Regarding data quality, we consider the incomplete coverage of the New Zealand Dairy Core Database on movement information had a negligible impact on our conclusions, both because 97% of farms were captured and because there is no obvious reason to believe that unrecorded movements from recorded farms and all movements made by unrecorded farms would be more likely to be from high risk to low risk areas.

One of the ways to overcome these limitations is to develop a behavioural prediction model of herd managers' trading, with a solid understanding of factors involved in their trading behaviour. Farmers' intention to implement biosecurity and disease control measures is known to be affected by various factors: physical attributes, such as herd demographics and farmers' characteristics, and socio-psychological attributes, such as social norms and habits, perception and actual experience of diseases, and beliefs on the efficacy of a particular measure (Alarcon et al., 2014; Broughan et al., 2016; Garforth, 2015; Erling Kristensen and Jakobsen, 2011; Santman-Berends et al., 2014; Swinkels et al., 2015; Valeeva et al., 2011). It is however largely unknown how such intention is prioritised or not on the verge of an instantaneous decision making for livestock trading. A recent study conducted in the endemic bTB region in New Zealand suggested that herd managers' cattle trading practices are likely to be shaped by far more complex factors than farm's bTB risk status—culture and values that have evolved in the local farming community matter (Enticott, 2016). We emphasise that such a behavioural model can be readily incorporated into the flexible network re-wiring framework, which allows us to explicitly consider human behavioural feedback in a disease simulation model (Ferguson, 2007). This will in turn facilitate a better evaluation of the efficacy of disease control strategies.

3.6. Conclusion

By developing a network re-wiring model, this study provides evidence that livestock movement networks may change based on behavioural response of herd managers towards regionalisation. We confirm that herd managers in perceived low bTB risk areas purchased cattle in a risk-averse manner, which might have contributed to the reduction in bTB prevalence in New Zealand. Yet, the results suggest it is theoretically possible to render the current cattle movement 'safer' in terms of bTB spread containment. Physical constraints and socio-psychological factors that determine herd managers' livestock trading behaviour warrant further studies to understand how herd managers respond to future livestock disease regulations. The flexibility of a network re-wiring framework presented in this study allows such a behavioural response to be incorporated into a disease simulation model, which will in turn facilitate a better evaluation of disease control strategy.

3.7. Acknowledgements

We are grateful to Mark Bosson and Thomas Lopdell for preparing the data used for this study and providing insightful suggestions for this project. We thank Mark Neill, Jane Sinclair, and Kara Dawson for their time spent on constructive discussions. Livestock Improvement Corporation and TBfree New Zealand are deeply acknowledged for providing the data. We also thank anonymous reviewers whose comments have improved the quality of the earlier version of this manuscript.

Chapter 4

Understanding how farmers make livestock trading decisions: A qualitative study

Arata Hidano[1], M. Carolyn Gates[1], Gareth Enticott[2]

[1]EpiCentre, School of Veterinary Science, Massey University, Palmerston North, New Zealand

[2]Cardiff School of Geography and Planning, Cardiff University, United Kingdom

4.1. Abstract

It has been traditionally considered that farmers make an irrational decision when they choose not to implement biosecurity practices that researchers and regulators believe are important to mitigate disease transmissions. However, recent social science studies suggest that farmers behaviours often arise as the natural result of their everyday farming life rather than a lack of intention to improve biosecurity. Nevertheless, farmers' livestock trading behaviours have been almost exclusively studied from a biosecurity perspective, providing little understanding why farmers are engaged in specific trading practices and why they may not verify disease status of source farms and animals. To understand how farmers develop trading practices, we conducted qualitative interviews with 15 New Zealand dairy producers with varying perceptions and experiences of bovine tuberculosis (bTB), where the rudimentary measure of a bTB infection risk for each individual farm is readily available to farmers.

The qualitative analyses suggested that there are three key stages in how farmers develop their own farm culture around biosecurity practices related to cattle trade. In the first stage, farmers naturally come to identify the basic characteristics of cattle they should manage in their given physical and natural farm environment, which determines their baseline purchasing practices. In the second stage, farmers try to maintain their farm culture by developing various trade strategies such as avoiding purchasing from farms with a 'shady' farm culture rather than specifically checking the disease status of animals with diagnostic testing. In the third stage, farmers experience a breakdown of their farm culture due to a severe negative event caused by introducing diseased cattle. This is the period in which farmers appear to scrutinise and revise their purchasing practices.

These results suggest that farmers' livestock purchasing behaviours are complex products of farm culture and past experience. We contend that these behaviours cannot be comprehended if purchasing behaviours are framed as a biosecurity practice. There is a need to shift our framing of farmers'

behaviours—from isolating them as an either disease, production, or animal welfare issue to viewing them as a product of farm system. We also need a better understanding of how farmers gather relevant information and how their experiences influence their decision to change behaviours. These may allow us to develop an effective and trustworthy communication strategy, which can more effectively encourage farmers to employ better biosecurity practices.

4.2. Introduction

Theoretical and empirical research studies have shown that farmer behaviour plays a substantial role in determining how livestock diseases spread within and between farms (Fournié et al., 2013; Manabe et al., 2012; Paul et al., 2013; Vernon and Keeling, 2012). In particular, farmer decisions around buying and selling cattle are responsible for creating the contact network through which important diseases can spread between farms. Although many studies have used basic social network analysis to understand the temporal and spatial variability in movement patterns, there is surprisingly little knowledge available regarding how and why farmers make livestock trading decisions as well as their opinions towards managing the biosecurity risks associated with purchasing in new stock.

The published literature on livestock trading practices almost exclusively frames these behaviours from a biosecurity perspective. Studies often interpret farms being an open-herd as lack of farmers' motivations to maintain high biosecurity standards. Similarly, farmers that do not verify the disease status of purchased animals are generally considered to have poor biosecurity practices (Benjamin et al., 2010; Ritter et al., 2017; Young et al., 2010). Nevertheless, it is likely that other factors, not only farmers' attitudes towards biosecurity, contribute to livestock trading behaviours. For instance, Enticott (2016) reported while New Zealand farmers consider C status—a risk score given to each cattle herd according to the number of years it has been free from bovine tuberculosis (bTB)—when purchasing livestock, farmers also place an importance in whether or not introduced cattle are going to fit their farming conditions. In this study, farmers indicated that they purchase cattle locally even though they are not best rated animals because farmers believe livestock from other regions would not perform well in their wet and harsh environment (Enticott, 2016). A study on animal welfare behaviours also demonstrated that farmers' milking practices can be disturbed due to introducing cattle with undesirable behaviours that had developed because of poor treatment at the previous farms (Burton et al., 2012). This suggests that stockmanship of source farms may be an important consideration for some farmers when purchasing animals.

Other recent biosecurity studies discuss how the farm environment determine the practicality of implementing certain biosecurity practices (Ritter et al., 2017; Shortall et al., 2017a). These studies, however, still view that farmers may be willing to perform such practices, but consciously choose not to because they are hampered by the physical constraints imposed by their farm environments. Nevertheless, these is increasing evidence from social science studies suggesting that farmers naturally come to know 'what to do' through a social milieu without consciously recognising that alternative ways of managing farm biosecurity (Burton et al., 2012; Curry-Roper, 2000). Farmer behaviour may therefore be viewed as a natural product from their farming life rather than a consequence of a clear intention. Segerdahl (2007) used a term 'culture' to describe the ground from which farmers' natural behaviours arise. While there are various definitions for this term, we use the notion 'cowshed culture' proposed by Burton et al., (2012). We describe 'cowshed culture' in details in the methodology section; in short, this notion argues that the farm buildings play a significant role in shaping an everyday interaction between farmers and animals, which leads to a specific farmers' behaviour.

We argue that accounting for the influence of the farm environment on continuous farmer-animal interactions—farm culture—is critical for understanding why farmers are engaged in a certain livestock practice and how they may change these practices. From a disease control perspective, it is then imperative to understand how disease may influence farm cultures and hence farmers' livestock trading behaviours. To the best of our knowledge, there are no studies that frame farmers' livestock trading behaviours in a farm culture framework.

To study this, we used bTB in New Zealand as a case example because, as already mentioned, farmers can easily obtain the bTB risk score for other farms. This unique situation allows us to examine why farmers may or may not consider C status and how such consideration is integrated into farmers' livestock purchasing behaviours. We conducted qualitative interviews with 15 New Zealand dairy producers from both a low and high bTB risk area and those had or did not have a bTB breakdown in the past. We show that the farm physical and natural environment significantly determines the type of cattle to be introduced and that farmers use their own strategies to determine from which farms they purchase animals. Evaluation of other farm cultures is one such strategy, however, we demonstrate that perceptions and experiences of disease also play a various role in evolving farmers' livestock purchasing strategies.

4.3. Methodology

4.3.1. Theoretical background—'cowshed culture'

Burton et al., (2012) introduced a concept of 'cowshed culture', which proposes that farmers' behaviours are developed through an everyday interaction between cattle and farmers, which is in turn shaped by a design of cowshed and characteristics of walking lanes. A smooth flow of milking is critical for dairy producers because they spend a large amount of time in milking practices and it determines farmers' workload. The milking flow can be disrupted by poorly maintained walking lanes (e.g. those with stones that harm cattle hooves) and poorly designed cowshed (e.g. those with pitted floor) because these discourage cattle from walking on lanes or entering into a shed. Under these environments, farmers may push animals harder, causing lameness, and this leads to a further disruption to a milking flow. A disrupted milking flow incurs further workloads and psychological distress on stockpersons, which may result in undesirable farmers' behaviours: this is a negative cycle of farm culture (Burton et al., 2012). The opposite exists: a smooth farm system is likely to improve a welfare of both livestock and farmers, leading to more preferable farmer behaviours. The concept of farm culture therefore acknowledges a continuous interaction between farmers, animals and environments that shape this interaction. We apply this concept and explore how both these physical environments (i.e. cowshed and lanes) and natural environments (e.g. pasture and weather) naturally shape farmer livestock purchasing behaviours among New Zealand dairy producers.

4.3.2. Institutional structure of New Zealand dairy farming

Two distinct features of New Zealand dairy farming system make it suitable to study stockpersons' livestock trading decision making. First, almost all New Zealand dairy farms run an extensive seasonal pastured-based system, where farmers heavily rely on the growth of pasture for animal nutrition. Second, the majority of milk produced in New Zealand is exported to an international market, meaning that the financial status of dairy farms is substantially influenced by international milk prices. These two uncontrollable external factors (weather and international market price) are dynamic and to some extent unpredictable. New Zealand dairy farms therefore need to manage their systems flexibly according to the changing situation. In particular, farmers are required to continuously adjust their herd sizes: the size often needs to go down if there is insufficient pasture to minimise operating costs and go up when the milk price is higher to increase profits. This leads to dynamic and frequent livestock movements throughout the country. The need of a dynamic change in a herd size also means the difficulty for dairy producers in having a constant trading partner because their trading events are irregular in terms of size and timing. New Zealand dairy producers may therefore have to identify a new partner at every trading event with some criteria. Indeed, this need is repeatedly mentioned in our interviews on producers. Taken together, New Zealand dairy farming system offers an ideal opportunity to understand the development process of making livestock trade decisions.

4.3.3. Bovine tuberculosis in New Zealand

Given its importance for public health and international trade, bovine tuberculosis (bTB) in livestock is designated as a notifiable disease in New Zealand. Herds identified with bTB are required to immediately cull bTB positive animals and are placed under restrictions from moving cattle until the disease is cleared, which causes significant economic burdens for affected farms. New Zealand has succeeded in substantially reducing the number of bTB infected livestock herds based on various control strategies (Livingstone et al., 2015a). Namely regionalisation and risk-based trading schemes are assumed to have played a pivotal role in preventing bTB spread between herds (Enticott, 2016; Livingstone et al., 2006). In this context, regionalisation categorises livestock herds into a several groups primarily based on the prevalence of bTB in their geographical areas. Previous research has found evidence that this may result in risk-averse purchasing practice where farmers in low risk regions avoid purchasing cattle from high risk regions (Hidano et al., 2016). In contrast, the risk-based bTB trading scheme in New Zealand reveals whether or not a farm is currently infected with bTB, and it confers a number (maximum 10) to each bTB free farm that indicates how many years the farm has been free from the infection. This system, so-called C status, may provide stockpersons with further information regarding bTB risk; however, the previous study shows that farmers in the West Coast region—an area where bTB has been widely prevalent in the history—do not necessarily interpret a higher C status (i.e. bTB free for a long duration) as a lower bTB risk (Enticott, 2016). These regionalisation and C status systems therefore provide an interesting opportunity to study whether or not disease risk information is considered by farmers and how this affects their livestock purchasing practices.

4.3.4. Interviews

Data were collected from 15 qualitative interviews with New Zealand dairy producers. New Zealand dairy producers can be categorised into three groups: farm operator, share-milker, and worker. A farm operator owns both the cattle and the land and may hire workers. A share-milker owns the cattle, but not the land, and therefore leases infrastructures (e.g. lands and cowsheds). A common type of share-milker is so called fifty-fifty share-milker, who receives 50% of the total profit from the milk production. A worker includes those who work for either farm operators or share-milkers and do not own either the cattle or the land. In this study, we included both farm operators and share-milkers since they are responsible for making decisions around livestock trade —hereafter, we refer them to as farmer.

The interviewed farmers included individuals from both low and high bTB risk areas to investigate differences in how they develop a livestock purchasing strategy. For a low bTB risk area, we purposively chose Waikato, Taranaki (North Island), and Canterbury (South Island) because these are the major dairy producing areas in New Zealand (Anon, 2016). For a high bTB risk area, we chose West Coast (South Island), which has had one of the highest reported prevalences of bTB in New Zealand over several decades (Livingstone et al., 2015a; OSPRI, 2015). Our sample size of 15 was determined to maximise the sample size within the budget and time. The sampling frame was generated by asking researchers, veterinarians, and industry stakeholders to provide a list of candidate stockpersons in each region that may be willing to participate in the study. We also contacted individuals in OSPRI— the organisation responsible for bTB control in New Zealand— to provide a list of farmers who had previously experienced a bTB breakdown and would be willing to participate in this study. The objective was to explore differences in behaviour that may have resulted from the farmers having been directly impacted by bTB.

All potential participants were contacted by phone and the objective of the study (i.e. livestock trading decision making) was explained. After their willingness to participate was confirmed, in-depth face-to-face interviews were carried out between November and December 2016 at the interviewee's preferred location. This was primarily on the farm properties except for one interview that occurred at the office where the farmer was employed for another job. Interviews lasted between 30 and 83 min. Two interviews were conducted with female farmers, 12 interviews were conducted with male farmers, and one conducted with husband and wife couples. The interviews were semi-structured whereby farmers were initially asked several questions about background information of themselves and their farms. Interviewees were then asked if they had purchased or sold any cattle recently and if so they were asked to tell stories about the experience. Subsequently, depending on how interviewees responded, different lines of enquiry were used to ask the following questions; how and when they made a purchasing and/or selling decision; any experience that changed their trading practices. All interviews were conducted by the first author. To compensate interviewees for their time, a NZ$100 gift card was given to each participant after the interview. Transcribed interviews were not returned to the interviewed farmers because of the limited study time.

This project has been evaluated by peer review and judged to be low risk under Ethics Notification Number 4000016617 at Massey University. Consequently it has not been reviewed by one of the University's Human Ethics Committees. The researchers named in this document are responsible for the ethical conduct of this research. If you have any concerns about the conduct of this research that you want to raise with someone other than the researchers, please contact Dr Brian Finch, Director (Research Ethics), email humanethics@massey.ac.nz.

4.3.5. Analysis

Interviews were all audio-recorded and transcribed by the first author. Personal identifiers were removed from the transcribed files to ensure the anonymity of interviewees. Transcripts were imported into NVivo Pro 11 for Windows (QRS International, Australia). Data was initially analysed under a grounded theory framework by the first and second authors to generate a list of codes. This preliminary analysis led to a thematic analysis drawing on the concept of ‘cowshed culture’ proposed by Burton et al., (2012) as described above. The transcripts were coded and then clustered into themes, whose inter-relationship was subsequently analysed.

4.4. Results

Analysis of interviews clearly suggested that farmers' livestock purchasing behaviours can be considered as a reflection of farm culture. Purchasing behaviours therefore change as farm culture changes. We identified the development of farm culture can fall into three key stages: (1) a growth stage, (2) a maintenance stage, and (3) a breakdown and restoring stage. In the first stage, as farm culture grows, farmers naturally come to identify the basic characteristics of cattle they should manage, which determines the characteristics of cattle they purchase. In the second stage, farmers develop various livestock purchasing strategies to maintain their farm culture. In the third stage, farmers experience a breakdown of their farm culture due to a severe negative event, where they may change their trading practices. Below, we describe how each farm culture development stage occurs and how this influences farmers' livestock purchasing practice.

4.4.1. A growth of farm culture: shaping a baseline livestock purchasing practice

Interviewed farmers often highlighted that having an easy farm management is the key issue for their farming and that two farming components are particularly important: a smooth milking flow and smooth grazing on pasture. Farmers therefore try to develop a farm culture that enables these two components. Analysis of interviews revealed that farmers naturally come to identify the characteristics of cattle they need during the process of developing farm culture, which shapes farmers' baseline livestock purchasing practices. In this process, physical and natural farm environment play a significant role.

4.4.1.1 Physical farm environment

Confirming the findings from Burton et al., (2012), interviewed farmers iteratively mentioned how their farming practices have been shaped by their physical farm environments, namely cowsheds and walking lanes. For example, farmer 1 (Canterbury) mentioned:

F: “When we take the heifers into the herd for milking in their first lactation, we will split them between 2 sheds on breed. Because this shed down here is rotary with grain feeding, short tracks… so the tracks aren’t very long and very good tracks. So we put the all Friesian, the big cows, down here. And the other shed, it’s a herringbone shed, old cowshed. Not made for big cows with no grain feeding. Very long walks and the tracks aren’t quite as good. So we put the cross-bred and Jersey, anything with harder feet, we put them in this shed [...].”

I: “So they rarely mix?”

F: “No. [...] It just makes the management easier when you have all your cows are the same. All these cows are roughly the same size, uhmm and all cross-breds, all black and brown, and when they line up in the herringbone it’s easier to have whole lot of cows the same than just to have big cows and small cows and.. or whole big cows and try to fit little one in the middle… they don’t like it. If you keep them all the same, it’s nicer for them, they fit better.”

This extract highlights that farmers observe how cowsheds and lanes influence cattle behaviours, on which they modify their behaviours. Creating a better environment for cows by keeping an appropriate cattle breed and allocating them to suitable sheds and walking lanes is important for this farmer because it makes a farm management easier. In doing so, cattle are less likely to have lameness and feel less stress during milking, contributing a smoother milking flow. In turn, stockpersons save their time and feel less distressed with the easier management.

4.4.1.2 Natural farm environments

Many New Zealand dairy producers run an extensive pasture-based grazing system. Stockpersons observe and learn from weather, paddock, cattle and their interactions. The seasonal weather patterns distinct to each region affect the growth of grass and paddock conditions. Grazing is not only about feeding cattle in New Zealand; it is an important part of farming to control the quality and growth of grass (Holmes et al., 2002). Grazing with too much intensity may damage the soil and grass, and poor paddock conditions may lead to lameness in cattle, which disrupts the milking flow. A successful understanding of this complex relationship enables stockpersons to manage a farm better. Share-milker 3 in Waikato noted that they are trying to down-scale the size of animals in his herd:

“Main reason we wanna bring the size of the animals down is… cos the cows are getting too big and this farm gets quite wet in winter and big cows are gonna sink, so they get a lot of lame feet, and…. Little cows just seem to be more profitable… it is lighter on feet and easy to maintain.”

Here, as a share-milker—who owns cattle but not the land and therefore need to move from one farm to the other regularly—this farmer recognises the need of changing his current practices after having observed the behaviours of cattle in the environment. What was suitable in the previous farm environment—having big cows—is not preferable anymore in the current farm environment.

These two extracts demonstrate how farmers naturally come to realise what they need for a better farm management through everyday interactions with their animals. Given that these interactions occur both in the natural and physical environment, these largely contribute to shaping farmers’ practices. In this process, farmers identify the characteristics of cattle they keep in their farming environment, naturally determining what kind of cattle they need to purchase.

4.4.2. Maintenance of farm culture: developing a livestock purchasing strategy

As described in Methodology section, New Zealand dairy farmers have various reasons why they need to introduce external cattle. Interviewed farmers often suggested, however, that introducing cattle is not an easy task because it can undermine a developed farm culture: an introduction of external cattle means an introduction of external farm culture. For example, the introduction of infectious diseases such as bTB or introducing cattle with undesirable behaviours may disrupt an established farm management system. Although our primary interest was to understand how farmers use bTB risk information (i.e. C status), this information is not always considered important by farmers while they acknowledged the importance of bTB. Analysis of interviews, however, revealed that farmers develop more generic purchasing strategies. Here, we first analyse how farmers interpret C status, then demonstrate why they rely on more generic strategies.

4.4.2.1 Interpretation and use of C status

A farm's C status has a number ranging from 1 to 10 (e.g. C10 means a farm has been free from bTB for at least 10 years). It might be intuitive to assume that farmers prefer purchasing from farms that have been free from bTB for longer: nevertheless, this view was not commonly demonstrated during the interviews. Farmers in a low bTB risk region and who never had a bTB breakdown often demonstrated that the C status does not really matter as long as a source farm is free from bTB.

"As long as they're passing TB test… yeah as long as they pass TB test I don't think I'm too worried. I've never really thought about it. As long as they're clear and not on movement control… that's not a factor when I buy animals… definitely I don't wanna get TB". (share-milker 2, Waikato)

This extract highlights that his indifference to the C status—being 'never really thought about it'—is not because the farmer does not care about bTB: rather, the farmer has his own understanding on a bTB risk. Having being farming a low risk region, bTB has played a relatively minor role in developing a farm culture. It is only a dichotomous bTB status—clear or infected—that matters. Another reason for this indifference may be that many farms in a low risk region have C10 status and farmers do not have opportunities to think about the status, as demonstrated by share-milker 3 (Waikato): "as long as they're clear yeah, it's all good. I haven't looked at it too closely. Because most of us are 10 here".

Farmers from a high bTB risk region, on the other hand, have a very different view on the C status; it cannot be interpreted by its own and needs to be understood in a context. For instance, farmer 14 (West Coast) mentioned:

"Depends where they are and why they are [with a specific C status]. You know, you look into those sorts of things. And where they are coming from… like here in the coast, it's a TB area so you know that it would be the likelihood but… yeah we just go through… check it out".

Here, it is evident that bTB status is understood in a context of source farm's culture—the number in the C status itself does not have much meaning. Moreover, the meaning of the C status further becomes less important due to the understanding on bTB that is shared by farmers in the high bTB risk region:

"You know you can buy C10 and still get TB if you're really unlucky. It's just the way the penny drops." (farmer 4, West Coast)

Farmers in this region consider bTB as a disease due to a bad luck, confirming the finding from a previous study conducted in the same region (Enticott, 2016), just as described by farmer 15 (West Coast): So many various people had been affected, so they understand it's nothing they did or whatever, just one of those things happened."

These together indicate that the C status is unlikely to be a primary factor to consider for farmers when purchasing livestock—for different reasons in a low and high bTB risk area. As described below, further analysis on interviews suggested that farmers rather develop more generic purchasing strategies to prevent introducing cattle that would have a negative impact, not only for bTB but other diseases as well.

4.4.2.2 A generic strategy: avoiding a 'shady' farm culture

Interviewed farmers demonstrated that unless they are exiting the dairy industry, they normally send cattle that are unproductive or have serious health conditions (i.e. repeated mastitis and lameness, and behavioural issues) to slaughter and sell cattle that can still produce milk but only at a suboptimal level on their farms. Nevertheless, they also often demonstrated their concerns about the presence of other farmers that sell cattle which should have been sent to slaughter. This is problematic for farmers; it is difficult to notice these serious mal-conditions when purchasing because it takes a while to recognise these problems or requires observation during specific circumstances such as during milking, as illustrated by following extracts.

"Three quarters [...] people don't want those. Off to the work. Mastitis definitely. We would not knowing sell cows that has got mastitis or repeated lameness, we wouldn't do that. That's not honest. That's a very shady farmer that would buy those and if he is shady he's got selling to somebody else. And our industry needs that... we need to be self-monitoring. We need to be able to trust each other. We don't need shady people. Cos it's a very hard industry to be in." (Farmer 11, Waikato)

"I don't actually like sale yards [...] you don't really know why those animals are on sale yards sometimes. Fine you might look at these animals and the animals are perfectly healthy. These animals might have been sent to the sale yard to go to the works because they've got problems." (Farmer 5, Taranaki)

Here, farmer 5 demonstrated his preference not to purchase from sale yards because it is impossible to know whether selling farmers have a 'genuine' or 'shady' culture. Therefore, it is important for farmers to identify farmers with 'shady' farm cultures. Farmers demonstrated various strategies to achieve this. First, farmers may purchase from farms that they personally know that they have a 'genuine' culture and trustable. Farmer 4 (West Coast) noted:

"I mean we've got neighbours around the road but he's got Friesian. If we wanted to buy Friesian, I'm happy to buy them off from him. Because he thinks the same as we do. [...] Honesty, integrity, you know, if there was a problem he would tell you what it was. There is a really good heifer, but they are pretty toey when they calve you know you have to spend a lot of time to get them quiet down once they've got going they'll be good, righto, that's alright as long as we know that."

Farmers, however, mentioned this type of personal trading is often infeasible for multiple reasons: they do not know many sellers who are selling at the right timing (farmer 4 West Coast, farmer 9 Canterbury); it is difficult to agree on a price (farmer 8 Canterbury); there is no time to set up a personal deal (share-milker 3 Waikato, farmer 11 Waikato).

The second, and more common way to avoid trading with 'shady' farms is to use a stock agent—a middle man. Other trading channels mentioned by interviewed farmers are listed in **Table 4-1**. Most stock agents in New Zealand work locally but generally belong to a large company which has a nation-wide network and therefore have an access to rich information about livestock markets. Farmers value the local knowledge of stock agents; stock agents know which stockpersons had some trading problems in the past:

"he [stock agent] knows… 'this guy is selling cows, selling surplus cows for 5 years or 10 years and we never had problems or he sold some cows and we had a bit of problem 3 years ago so maybe you don't wanna go there'… so he knows all that. Whereas if we're going to trying to deal with other farmer, they don't tell you, you won't know." (Farmer 4, West Coast)

Table 4-1 List of livestock trading channels mentioned by interviewed farmers

Channel	Characteristics
Directly from paddocks	• Farmer purchases cattle directly from seller's farm without going through any physical intermediaries • Stock agents choose the sellers that are likely to be favoured by the buyer • Trade usually occurs locally • Stock agents sometimes need to find cattle from a distant place for their buyers

	• Convenient for farmers who are looking for a large number of cattle from a single or few herds
Sale yards	• Purchasing ‘genuine’ cattle can be tricky • Farmers may ask stock agents to choose animals (Farmer 3, 7, 10) • Not many sale yards exist in South Island (Farmer 1, 7) • It can be entertaining—bidding and meeting other farmers
From a seller that a farmer personally knows	• Often infeasible, although this is in general a reliable way to purchase from ‘genuine’ farms
Through on-line	• Widely recognised by interviewed farmers but they were reluctant to use it because of the difficulty in ensuring that cattle are ‘genuine’ • Price can be cheaper than other channels because of no need to pay commissions to stock agents

Despite relying on stock agents, some farmers also make sure they visit and check the selling farm—this is a further strategy to assess whether the seller is honest and has a good farm culture. This assessment seems to involve either a communication with the seller or visually checking the farm and cattle, or both of them. Share-milker 3 (Waikato) noted:

“He [stock agent] sort of got…3 or 4 herds for me to look at and we went for a drive one day. I think we went to…the first 3 and I was like ‘Hmm, I hope the last one is good’. […] The way the farmer had them… it wasn’t… they were a little bit light and looked ugly. And rough… the coats were rough. They weren’t shiny, healthy looking. So it just sort of gives you an idea that maybe he doesn’t do job properly. When we went to the last one the owner came with us we went around and he told me this cow doesn’t give much production, this is my peak cow here. You know he just knew his herd. He looked like he had more involvement with it and he actually cared. As soon as I walked in there I was like this is what I want. It's a nice looking herd".

This quote highlights two important points. First, the farmer assessed the sellers' farm management as poor based on the 'ugly' appearance of their cattle. The literature suggests that farmers 'read' how other farmers have been managing livestock over years from the appearances of their livestock—an indicator of a 'good' farmer (Burton, 2004; McAloon et al., 2017). The 'ugly' appearance of livestock therefore indicates farmers' poor management and hence links to 'shady' farm culture—cattle on these farms may have some hidden problems. The link between the poor animal care, poor management, and 'shady' farm culture is also mentioned by Farmer 14 in West Coast: "if he is not looking after his animals and records probably are not 100% either". Second, 'knowing their own herd' provided the farmer with a credential that the seller is 'genuine'. Farmers who know their own herds well are likely to be able to identify problems in cattle quickly and minimise stress on cattle, which is an important component of a good farm culture (Burton et al., 2012).

In summary, purchasing cattle from a 'genuine' farm culture is important: animals from such a farm are less likely to have serious problems. Farmers consider good looking animals, other farmers' knowledge on their own herds, and farmers that care for their animals to be indicative of a 'genuine' farm culture.

4.4.2.3 Assess a farm cultural distance

Avoiding purchasing cattle from a 'shady' farm reduces the chance of introducing disease problems to a farm. This, however, still does not guarantee that introduced animals fit well to a new farm environment. Introducing cattle that have incompatible behaviours can disrupt a smooth farm management that has been developed over years. Analysis of the interviews suggest that farmers prevent this by purchasing from a similar farm environment to their own with the belief that a similar farm environment cultivates a similar cattle behaviour. For instance, the cowshed design of a source farm was frequently mentioned as a factor to consider:

"[…] you've got things like cowshed they are coming from as well… like herringbone or rotary… there are always things you got to think about. Some sheds go clockwise and somewhere anti-clockwise. […] You are still gonna disrupt the cow flow when you are training them, yeah it makes a difference. Just a little thing that people are not always interested in. Practical things, you can't explain all these things." (Share-milker 12, Canterbury)

"If they offer rotary that would be better to buy them. If you go to a shed cows up in a rotary… makes it easier to break... doesn't matter for heifers because 'I've never been…I don't know what shed is'. But for milking cows, yeah it is easier, it's not a big, it doesn't take long to break them in, you can take some to go from herringbone to rotary they'll do get to know… you've got to push them on for a bit, train them up. Probably the worse is actually cows from herringbone to rotary is worse than cows from rotary to herringbone." (Farmer 8, Canterbury)

These extracts once again highlight that the physical farm environment plays a pivotal role in shaping farmer-cattle interactions and hence cattle behaviours. While some farmers consider reshaping cattle behaviour to be extremely difficult (share-milker 2, 12), others consider it as a manageable task (farmer 4, 7, 8, 10, 11, 14). The degree of this perceived difficulty not only depends on a cowshed type that farmers have but also; how many cattle they purchase and hence have to teach (farmer 11) and the experience and skill of farmers (farmer 7). Likewise, a source farm's natural environment may be a factor to consider when purchasing. Farmer 11 (Waikato) noted that he purchases from farms that have similarly hilly paddocks, Palmerston North—300km apart from his farm—rather than Morrinsville, which is one of major dairying areas in his proximity:

"Morrinsvile… those cows are not going to move up to hills so we look for better cows coming from Palmerston North."

Given characteristics of a natural environment are often geographically determined, a spatial location of a source farm is an important factor to consider for farmers. Interestingly, farmers seem to evaluate other farms' potential disease status based on the farms' spatial location:

"Trying not to [buy from West Coast]. Because they are at high TB risk, aren't they? Yeah trying to stay away from that." (share-milker 12, Canterbury)

"now, we wouldn't purchase anything from North Island because of Theileria" (farmer 8, Canterbury)

"I mean Taranaki is a good area, there are lots of good cows. It's been there for a long time. But [...], even though they are really good herds, they're still full of Johne's." (farmer 4, West Coast)

We note that farmers also demonstrated their status of living with a disease: "I don't have to do [control] Johne's. It's always there" (share-milker 12), and "I'm sure we probably have them [Bovine Viral Diarrhoea]... but they are keeping their ugly head down" (farmer 1). Although this is beyond of scope for this study, these extracts suggest that a presence of a disease can become a part of farm culture too—farm practices evolve adjusting for a disease presence, which in turn might explain a lack of motivation of controlling a certain disease.

4.4.3. A breakdown of farm culture: revising livestock purchasing practices

In the previous section, we introduced farmers' various strategies to avoid introducing cattle that may bring negative consequences. Here, we describe how farmers deal with this negative consequence once it is introduced. Analysis of the interviews suggests that this is a situation where farmers are no longer able to run their farms by doing what they 'know what to do'—a farm culture breakdown. The experience of dealing with a negative consequence seems to provide farmers with two lessons. First, farmers become better able to anticipate the distress they may experience if a similar breakdown occurs again. Second, their understandings on their own and other's farm culture are updated. As we discuss below, these two lessons are likely to act together to change farmers' purchasing behaviours.

4.4.3.1 Dealing with a negative consequence

A bTB breakdown is a typical example of such a negative consequence: the breakdown leads to culling of individual livestock (and potentially the whole herd as well as restrictions on selling and moving animals. The latter can be particularly critical for New Zealand dairy producers because selling and moving animals to other properties is an important herd management practice when fodder is limited. For instance, farmer 5 (Taranaki) demonstrated how the bTB breakdown imposed not only economic distress but also psychological distress by limiting his farming options:

"When you've got no option, you got into a corner… it's kind of sucks. When you've got option, you're always on the front foot, thinking about what you can do next, and that's kind of where we've got to in the last 12 months. And the part of that is changing the whole farm system. So you know… last 2, 3 years I felt like a death by thousand cuts type things… slow way of dying… you're always fighting fires… you're always wondering where how your next dollars are coming from… whereas if you've got options in your back pocket, then all of sudden your attitude can change. From fighting fires to actually thinking 'Ok where the hell am I going now? What am I gonna do?' And it's easy to say just a mindset but it's actually more than that. To get that mindset you need the options to start with. You can say 'Well…get the mindset and options will come' but it doesn't always work out. You know sometimes mindset is because of lack of options."

This extract exemplifies the process farmers undergo during a farm culture breakdown. Being no longer able to do what they used to do, the farm system and practices that have been developed over years have collapsed. Then, a negative cycle of a farm culture commenced: the oppressed situation put farmers in 'thinking in a silo mentality', which forced them to carry out practices that did not work well. This in turn further imposed a psychological distress on farmers. This negative cycle finally turned over when the whole farming system was entirely changed.

As described before, a smooth milking flow is paramount to a dairy farm. An introduction of cattle that do not know how to behave in a milking shed can therefore break a farm culture—behaviours and practices of both cattle and farmers that have been developed for a better milking flow are no longer effective. Share-milker 2 described the physical and psychological distress due to a disruption in the milking flow caused by introduced cattle:

"Because our shed's quite unusual, you don't get too many internal rotaries.. […] there's not many sheds like this so there's not many cows that know how to come…that's another thing that stops me from trading is that it's bloody hard to teach cows to come in the shed. So you can train them how to do that… so it took us 3 months to teach them how to come in. And even then after years some cows don't wanna come in." (share-milker 2, Waikato)

This stressful experience enabled the farmer to envisage a likely consequence and accompanying distress that would occur if he would again introduce cattle that do not know how to behave in his shed. This extract clearly highlights that the anticipated substantial distress discourages the farmer to experience it, which influences his trading behaviours.

4.4.3.2 Updating an understanding on a farm culture that guides a behavioural change

A farm culture breakdown due to cattle introductions also provides farmers with an opportunity to scrutinise their behaviours that farmers may perform subconsciously. Farmers may analyse the introduced farm culture that is responsible for the breakdown, and then develop an understanding on other's farm culture, which can be subjective. For instance, farmer 9 demonstrated:

"Um we had bulls last year that had a bloody pink eye. Bad… bad strain of pink eye. So we had some teaser bulls last year. So decided not to use teaser bull ever again for that reason […] it was you know the guys had to be very vigilant looking at eyes and making sure that they treated them. […] it was more just … hassle and cost… and stress because you know that they could go through the whole herd and imagine you'd have to put stuff on eyes on every cow… nah." (farmer 9, Canterbury)

Again, this extract highlights that this negative experience enables the farmer to envisage the worst scenario that could occur again. This experience also provided the farmer with an updated understanding on 'teaser bull farm culture'—teaser bulls have a bad strain of 'pink eye'. In the other word, a specific farm culture now links to a disease risk. This understanding can be entirely subjective: nevertheless, this extract suggests that this updated understanding guides farmers' behavioural change. Linking a specific farm culture to a disease risk is also observed for bTB. Farmer 6 (Canterbury) believes his bTB breakdown was caused by introducing cattle from a farm that had an official bTB free status at that time. Prior to the breakdown, the farmer "just presumed if they were clear, they were clear you know", indicating that a dichotomous bTB status on an official document only mattered before as we discussed before. After the breakdown, however, the farmer considers:

"Where they come from, how long they have been on that farm and where they are buying from… share-milkers move around obviously quite a lot so you have to be careful about that." (farmer 6, Canterbury)

Here, it is evident that the bTB breakdown has changed how the farmer interprets a bTB status. A specific farm culture such as share-milkers' now represents a risk of bTB—the understanding on these farm cultures is updated. Similarly, farmer 4, who moved from Canterbury (low bTB risk area) to West Coast (high bTB risk area), described his regret of having purchased from a high bTB risk farm after having being exposed to criticisms and perceptions from other farmers in the local community:

"we bought C1 [a herd that just became clear for bTB a year ago] at the first year we were here. And sort of I wished ever since we hadn't but anyway we didn't get TB, touch wood, as far as we know. We haven't had any since we've been here. Yeah I wouldn't do that again. I wouldn't buy C1 again, ever. It's just too risky."

While this farmer was farming in a low bTB risk region, bTB was not a real and the C1 status did not convey any bTB risk to the farmer. After having moved to the high bTB risk region, however, the farmer has experienced bTB breakdowns indirectly in the local community—bTB became a real and a specific farm culture linked to a bTB risk. These two extracts again demonstrate how updated understandings on a farm culture guided farmers' behavioural changes.

4.5. Discussion

Drawing on the notion of 'cowshed culture'(Burton et al., 2012), this preliminary study argues that farmers naturally develop their baseline livestock purchasing behaviours as their farm culture evolves through an everyday interaction between farmers, livestock, and environments. Farmers come to recognise what kind of cattle they should keep in their given farm environment to make farm management easier. Farmers then further develop this baseline practice in various ways to maintain their farm culture. Despite these strategies, farmers may still experience negative consequences due to introducing poorly suited or diseases cattle, which might lead to a farm culture breakdown. Our analysis indicates this breakdown enables farmers to anticipate distress that could occur in a similar situation, which encourages them to avoid such an experience again. By dealing with the breakdown, farmers also develop a subjective understanding on other farm cultures, which link to a risk of negative consequence. These are likely to act together to change farmers' behaviours in a specific direction. Below, we discuss important implications from these findings.

4.5.1. Need of shifting our framing of farmers' behaviours

The literature on farmers' behaviours often assumes that a behaviour is determined by one's intention, which is in turn influenced by attitude, perceived behavioural control, and social norm towards the behaviour. As we demonstrated, there seems to be a period that farmers have an intention to change their behaviours such as having experienced negative incidents. However, as we argue, their baseline behaviours are more or less naturally cultivated through an everyday interaction between farmers, livestock, and environments. Farmers may reach a status where they naturally 'know what to do', meaning that their behaviours are not necessarily chosen out of feasible alternatives. Each of farm practices therefore build on other farm practices and all practices work together to build a complex farm system (Burton et al., 2012; Whay, 2007). Attempts to change one's behaviour may therefore ignore this complexity and do not consider the role of the behaviour in the whole farm system. This may explain the consistent findings that farmers feel the scientific advice as irrelevant and find difficulty in understanding how such advice fits into their situations—leading to these advices being perceived as impractical.

We also showed that farmers may not necessarily rely on disease risk information that researchers and regulators perceive important for various reasons. It was evident that while biosecurity studies focus solely on minimising the risk of introducing risk inherent to introduction of animals, farmers in reality need to consider many other factors when purchasing animals such as improving farm production, avoiding animals with hidden problems, and maintaining their farm culture. Therefore, the farmers' strategy to purchase from 'genuine' farms makes perfect sense rather than checking items in the long list one by one—this is effectively impossible. This clearly highlights the need of a change in how we frame farmers' livestock purchasing practices—or more in general their farming practices. We contend that we need a wider approach, rather than independently tackling to farm issues such as animal welfare, biosecurity, and production—to understanding the complexity of farmers' behaviours. This may allow us to deliver a consistent and trustworthy message to improve farm biosecurity while meeting farmers' other requirements when purchasing animals.

These together also raise an interesting question around the common argument in literature that veterinarians are in the ideal position to communicate with farmers about biosecurity and diseases because of their knowledge. While we acknowledge the importance of veterinarians in promoting biosecurity practices, there are other actors that build a close relationship with farmers by understanding their farming system such as farm advisors and stock agents. We need better understanding of how these actors influence farmers' behaviours and how they can work together to communicate effectively with farmers.

4.5.2. Perceptions of a disease risk and behaviours

Using bTB as an example, we showed that farmers perceive the risk of bTB infection on a source farm differently depending on whether their own farm is based in a low or high bTB risk area. This is consistent with previous findings that a risk perception is influenced by a current local disease status (Garforth et al., 2013; Toma et al., 2015) and that risk perceptions and disease understandings are fluid and shared with other people and in the community (Enticott, 2016; Lupton, 2013). Nevertheless, the difference in bTB perceptions does not necessarily translate into an obvious behavioural difference. In neither of risk areas, farmers use the C status as a primarily information to prevent a bTB introduction. While farmers in a low risk area are concerned about binary disease status—infected or clear—those in a high risk area or previously had a bTB breakdown evaluate the risk in relation to other farm culture. This has two important implications. First, as often suggested in literature, better knowledge about a disease may not contribute to a higher uptake of disease preventive measures because knowledge may not be directly translated into an understanding on farm culture. Second, given understanding of farm culture is subjective and can vary between farmers, farmers may perform heterogeneous preventive measures. This is critical because we often try to associate farmers' willingness to prevent a disease with their implementation of a specific preventive measure we define. We argue that it is prerequisite to further investigate a process in which an understanding on farm cultures are developed and how we can inject scientific knowledge into such a process.

Interestingly, when talking about a certain disease, farmers often named other diseases, making analogies or comparisons. For instance, farmer 5 (Taranaki) noted that he does not really pay attention to BVD, saying "I know it works like Theileria… but I've probably heard of farmers being hit with Theileria than being hit with BVD". While farmer 4 (West Coast) was demonstrating his view on how his neighbours got bTB, this farmer noted that "Maybe it's been always there. Just needed stress that did bring it out", followed by "It's logical because Johne's work like that". These likely suggest that farmers try to understand a disease that is unfamiliar or uncertain by comparing to other diseases that they already know. Wiethoelter et al., (2017) reported a similar finding that Australian horse keepers assess the Hendra virus risk by using other diseases or events, which they are familiar with, as benchmarks. Analogy or comparison—more generally, relational reasoning or thinking—is known to be an effective way to make sense of something that people have not experienced before (Bostrom, 2008; Visschers et al., 2007). Should this be often used by farmers to understand unfamiliar diseases, there are two important implications. First, if false information is used by farmers to establish their understanding about diseases, it might be difficult to overwrite the initial understanding, leading to misunderstanding or misconception of the disease. Given that information on diseases is communicated in a different way depending on the disease characteristics (McAloon et al., 2017), we need a better understanding on how and when farmers establish such an analogy. Second, we might be able to deliver a better understanding about disease to farmers using a relational thinking by using analogy or comparison that is succinct and accurate. This way of communication based on relational reasoning has been already well studied in the discipline of risk communication and education (Dumas et al., 2013; Guy et al., 2013). By doing so, we might be able to prevent new important diseases from becoming one of diseases on the large lists.

4.5.3. Implications for livestock movement patterns

We found a relatively low level of concern for bTB among farmers in a low risk area. This seems to be somewhat contradictory to our previous finding that New Zealand dairy producers in a low bTB risk areas were significantly more likely to purchase cattle from low bTB risk areas (Hidano et al., 2016). Nevertheless, a similar finding is reported by a study conducted in the UK—cattle producers in the low risk areas prefer to purchase from low risk areas (Little et al., 2017b). This discrepancy may be explained by our finding that New Zealand farmers' purchasing decision is more driven by the options provided by stock agents for two reasons. First, given stock agents work locally and in general match buyers and sellers among their regular clients, trading often occurs within the same region. Second, farms in a high bTB risk areas have specific natural environments (e.g. severe weather, limited land available), which leads to a smaller herd size and few surplus animals to sell. Many interviewed farmers mentioned that stock agents that are looking for a large number of animals do not bother finding cattle from these small herds. These together facilitate cattle trading within a high risk area, limiting opportunities for farmers in a low risk area to even have options of purchasing from a high risk area. This means some of apparent risk-averse purchasing practices can be in fact just a product of this trading system based on stock agents. Such system might be specific to New Zealand, however, farmers in any countries need to rely on a specific system to find sellers, be it a word of mouth or an advertisement. We emphasise the importance of understanding such a trading system to fully understand livestock movement patterns. Note that our current and previous studies looked at a slightly different factor—C status and bTB risk regions—which may contribute to the discrepancy in our findings. However, we reason this is unlikely because the interviewed farmers were anyway not concerned about bTB risk in general.

Although farmers' livestock selling practices were not the scope of this study, the interviews provided an interesting implication that culling and livestock purchasing decision is inter-linked. Farmers building their herds demonstrated their opportunistic attitudes towards purchasing better cattle and culling their inferior animals, even though this is not their urgent task. A similar finding that cattle are culled to secure a place for new coming cattle is reported in the qualitative study conducted in Sweden (Bergeå et al., 2016). A few farmers also noted that they leave the decision of whether to send cattle to other farms or to a slaughterhouse to the discretion of the stock agents. Further studies are required to understand the complexity in farmers' livestock purchasing, selling, and culling practices.

4.5.4. Implications for biosecurity practices

While studies suggest that farmers recognise the risk of introducing diseases when purchasing animals (McAloon et al., 2017), they do not always verify the disease status of animals they purchase (Benjamin et al., 2010; Young et al., 2010). This leads to a discussion as to why farmers may or may not choose to verify the disease status of animals they are purchasing. Shortall et al., (2017) reported how effectiveness and practicality of different verifications practices— verifying disease status of animals by consulting seller's vets, by checking the accredited test results, or by asking the information to sellers—is perceived by UK farmers. Among these, asking the disease status to sellers was considered ineffective and the authors explained this might be attributed to the information asymmetry (Shortall et al., 2017a). Information asymmetry represents a condition where buyers have substantially limited access to the information seller has (Akerlof, 1970). Findings from this study suggest while information on specific disease status is important, farmers seem to judge the credibility of animals based on whether or not the seller is a 'genuine' farmer. Purchasing from 'genuine' farmers means that the sellers give honest information, they keep reliable records, and animals are less likely to have problems—this is more reliable and practical way for farmers to minimise the risk of introducing major negative impacts rather than going through a potentially long checklist of problems and diseases. It is important to recognise that what matters for farmers is whether or not the condition of cattle would undermine their farm cultures—be it a disease, behaviour problem, or productive and reproductive issue—and farmers seem to consider these equally important. On the other hand, scientists may treat these issues independently in each discipline, which may prevent us from understanding the complexity and reasons underlying farmers' decision making. For instance, share-milker 3 (Waikato) noted that he started testing cattle for BVD not because of its potential economic impact but because he was convinced he could take a better care of animals by controlling BVD. This example clearly emphasises that farmers' disease control practices cannot be understood without accounting for their empathetic feelings towards livestock (Wilkie, 2005) and the background culture in which such empathies have been cultivated (Burton et al., 2012). This requires not only taking an inter-disciplinary approach but also viewing farmers' behaviours from their view points (E. Kristensen and Jakobsen, 2011).

Noe that the sample size of the study may need to be increased. A previous study suggested that 'saturation'—the point at which no new information or themes are observed in the data—occurs within the first 12 interviews (Guest et al., 2006), although the saturation point can vary depending on the study subject and population characteristics. A more recent study, however, argued that the meaning of saturation remains unclear in practice and showed that coding saturation—the point at which all themes occur—and meaning saturation—the point at which meanings of occurred themes can be appropriately understood—require different numbers of interviews (Hennink et al., 2017). Future studies therefore may benefit from increasing the sample size, although there is still a critical knowledge gap in the estimating required sample size.

4.6. Conclusion

This preliminary study demonstrated how farmers' livestock purchasing practices may evolve over time as their farm culture changes. Physical and natural farm environment seems to play a significant role in shaping farmers' baseline purchasing practices. Farmers then update their practices in various ways as their understanding on disease and other farm culture is updated. We contend that this updating phase is a critical period in which we may effectively encourage farmers to employ a better purchasing practice in terms of disease control—farmers seem to scrutinise and revise their own practices which they may have done subconsciously in this period. It is imperative to further study how farmers update their understanding on other farm culture, how they gather relevant information, and how they translate such information into new practices in this critical period. This will allow us to develop an effective and trustworthy communication strategy with farmers to improve their biosecurity practices.

4.7. Acknowledgements

We deeply thank all farmers who kindly participated in this study. We are also grateful to Mark Neill, Jane Sinclair, Sonya Shaw, Aaron Yang, Emma Cuttance, Richard Laven, Aaron Chambers, Neil Chesterton, Sarah O'Connell, Chris Houston for identifying and providing contact information of participants. Special thanks goes to Lynsey Earl for helping the transcription of interviews. This study was partially supported by Massey University Research Fund, MAFF Japan, Beef + Lamb New Zealand, and Kathleen Spragg Agricultural Trust.

Chapter 5

Why sold, not culled? Analysing farm and animal characteristics associated with livestock selling practices

Arata Hidano[1], M. Carolyn Gates[1]

[1] EpiCentre, School of Veterinary Science, Massey University, Private Bag 11222, Palmerston North, 4442, New Zealand

5.1. Abstract

Livestock disease simulation models that incorporate animal movements often assume (1) that farmers' livestock trading practices remain consistent over time, (2) that animals sold to other farms are chosen randomly from a herd, and (3) that the animals' fate on the destination farm is not influenced by their past production and movement histories. The objective of this study was to assess the extent to which these assumptions are violated in the real world using records from a national database in New Zealand that captures both milk production and movement data for individual dairy cattle. All individual animal milk test records from 2006 through 2010 were extracted from the database and processed to generate different animal and herd level variables including cow demographics, previous movement history, milk volume, and milk composition (somatic cell counts (SCC), protein percentage, and fat percentage). Various statistical models were used to explore factors associated with farms' selling practice and characteristics of animals being sold. The results showed farms' livestock selling practices were highly influenced by both external factors such as market milk price and internal factors such as previous year's cow mortality and how long farms had been in business. Higher milk price increased both the number of cows being sold and the number of farms selling cows. Compared with cows that remained in the herd at the end of lactation, cows sold to other farms had lower fat and protein percentages, but similar milk volumes and SCCs. Cows that were previously purchased into the herd were more likely to be sold rather than being culled after controlling for the effects of age. Overall, these findings highlight the potential need for disease simulation models to account for dynamics in selling practices and animal characteristics when determining which animals will be sold to other herds.

5.2. Introduction

Simulation models have become increasingly important tools for helping decision-makers evaluate different options for controlling economically important livestock diseases at the animal, farm, and industry levels (Halasa and Dürr, 2017). This is particularly the case for diseases and control measures that would otherwise be too impractical or too expensive to formally study in the real world. For example, simulation models are often used to explore potential foot-and-mouth disease control strategies in countries that are currently free from the disease (Van Andel et al., 2018). However, biological systems are highly complex and even the most sophisticated models must make simplifying assumptions, which could influence the validity of any predictions generated by the model. As highlighted in a recent review paper (Mancy et al., 2017), these assumptions can be grouped into three key areas: (1) demographic assumptions—deciding the level of detail with which to model the host population structure ranging from individual-based models to entirely compartmental models, (2) contact assumptions—deciding how to model the connectivity of hosts ranging from simple mass-action transmission models to fully heterogeneous network models, and (3) disease assumptions—deciding which disease states to model.

Modelling the connectivity of livestock populations (contact assumptions) has received much more attention in recent years with the introduction of national animal identification and tracing system databases, which provide detailed records on the historical movements of livestock between registered herds. A large number of studies analysing livestock movement records have revealed that movement patterns are highly dynamic with a strong temporal and geographical heterogeneity (Bajardi et al., 2011; Ensoy et al., 2014; Vernon, 2011). This heterogeneity has been in turn shown to have a substantial impact on how disease spreads (Bajardi et al., 2012; Vernon and Keeling, 2012) and many disease simulation models therefore now account for these spatial and temporal heterogeneities in between-farm livestock movements rather than assuming a homogeneous mixing or a simple distant-based kernel distribution.

There have been three main approaches to modelling movements using records stored in the national databases. The first approach is to preserve and replicate every detail about the movements that was observed in the real world including the individual animal identity, source and destination farm identities, and movement date. The second is to still preserve the real-world data on farm identities and movement dates, but to assign a probability that disease transmission will occur through the movement based on the prevalence of disease in the source herd and the total number of animals involved in the movements. The third is to calculate parameters for between-farm movements such as a probability of a movement between given sets of farms during a given time period, which may be a function of various factors such as between-farm distance and production type (Lindström et al., 2011; Relun et al., 2017). The latter allows for a potentially more stochastic approach to generating between-farm contacts.

Regardless of which of the above approaches is used, most livestock disease models currently make key simplifying assumptions about the between-herd contact patterns. The first is that future livestock movement patterns will remain the same as the observed historical movement over the course of a disease outbreak or response. However, the retrospective analysis of livestock movement database records before and after the implementation of national animal health policies suggests that farmers may change trade behaviours to minimise the risk to their farm or the impact on their business (Gates et al., 2013; Hidano et al., 2016; Vernon and Keeling, 2012). For example, farmers from regions with a low risk for bovine tuberculosis tend to preferentially avoid buying cattle from high risk regions, which is likely to protect the disease status of farms (Hidano et al., 2016). Thus, any modelling studies that seek to predict the epidemiological and economic impacts of disease control measures need to account for dynamic behavioural changes (Hidano et al., 2018a). It is also important to note that other external factors such as changes in cattle market prices or other disruptions to the supply chain caused by natural or man-made disasters can also influence trade decisions regardless of the presence of disease or the introduction of new animal health legislation. Another example from a qualitative study on New Zealand dairy farmers suggests that farmers are willing to sell greater numbers of cattle when there is opportunity for more profit and that factors such a low market milk prices can encourage farmers to reduce their herd size through animal sales and culling, and others to increase herd size through animal purchases (Hidano et al., 2017a, 2018b).

The second assumption that models often make is that animals are chosen randomly from the herd to be sold or culled on a given date (Brooks-Pollock et al., 2014) or, at best, selected based on age-specific probabilities of a movement occurring (Widgren et al., 2016). This assumption is to some extent unavoidable given the often limited data available on animal characteristics in the national database as well as the strikingly limited knowledge on how farmers decide which animal to sell. Based on our qualitative interviews of New Zealand dairy farmers, animals sold from a farm are very unlikely to be 'top' animals in the herd; farmers often sell animals that are surplus but still useful for other farmers (Hidano et al., 2017a, 2018b). This indicates that farmers or traders likely have specific demographic and production criteria for making decisions about animal trade. Failing to account for this in simulation models may be trivial for fast-moving, highly contagious diseases such as Foot-and-Mouth (FMD) disease where all animals in the herd are expected to be rapidly infected and where all detected-positive farms are expected to be placed under movement restrictions. For other slow-moving endemic diseases, however, there are often specific risk factors that influence the likelihood of individual animals becoming infected. Furthermore, the infections are often subclinical and cause varying levels of production loss, which can increase the risk of infected animals unknowingly being sold to other herds. For example, Nekouei and colleagues showed that cows infected with bovine leukaemia virus (BLV) have a significant lower milk production in their second and third lactations compared to BLV negative animals in the same lactations (Nekouei et al., 2016). A reduction in the fat and protein yield has been also reported for cows subclinically infected with paratuberculosis (Pritchard et al., 2017), while a marginal reduction in the milk yield has also been shown among cows infected with bovine tuberculosis (bTB) (Mellado et al., 2015). Farmers may sell these animals not because they know animals' disease status, but because of their suboptimal productions. This may significantly increase the likelihood that a specific between-herd movement will result in disease transmissions compared with models that assume all individuals are equally likely to be selected regardless of disease status.

The third assumption is that an individual animals' fate is assumed to have Markov properties; that is, the fate of whether an individual is going to stay in a herd, be slaughtered, or sold depends only on the current status of the individual. This means, for instance, the probability of an individual to be sold can be determined based solely on attributes such as age, current production performance, and disease status while other important attributes such as how many times the animal has moved between farms and how long it has stayed at the current herd are often ignored. As highlighted in the above New Zealand study (Hidano et al., 2018b), farmers may sell animals for many other reasons such as the animals having poor temperament or poor behaviour. Farmers are also known to have a varying degree of personal attachment to individual animals that can influence that likelihood of that animal being sold. For example, Wilkie reported that commercial cattle workers maintain a greater emotional distance from their animals, particularly if these animals are bought in rather than being born and raised on farm (Wilkie, 2005). It is therefore not surprising that farmers may choose to sell animals they have bought in rather than those they have raised on their farms. Should this be the case, for example, animals that have moved more frequently would be more likely to be sold again, which would have a substantial impact on disease spread as these animals can theoretically have a higher risk of carrying a disease because of their movement histories.

In New Zealand, there is a unique opportunity to explore issues around reasons for selling cattle due to the existence of a national dairy health database that contains records on animal performance as well as between-herd transfers to other herds registered in the database. The three main study objectives were therefore to (1) identify factors that influence herds' animal selling performance and predict how it would change over time, (2) provide preliminary understandings on the potential difference in production characteristics between three populations: animals that are not removed, those going to be sold, and culled, and (3) investigate how animals' previous movement histories and other production characteristics would affect their fate of being sold or slaughtered. The implications of these findings for building more dynamic models of farmer trade patterns are also discussed.

5.3. Materials and methods

5.3.1. Data description

This study utilized the Dairy NZ Core Database in New Zealand, which encompasses over 70% of commercial dairy herds and has been recording four streams of data for individual animals since 1st January 1990 (Anon, 2001). These data streams include Key data (farm location and herd identification), Static data (individual animal identification, animal's sire, sex, date of birth, and breed), Event data (date animal entered and exits a herd, animal fate including death, slaughtered, or sold to other farms, cause of fate, and calving date), and Production data (production test date, average number of milkings per day, AM and PM milk volume (in litres), fat percentage, protein percentage, and somatic cell count (SCC)—SCC is described as the number of cells ($\times 10^3$) per millilitre in this manuscript). For the purpose of this analysis, we chose to focus on the last 5 years of available data in the extract (1st July 2006 and 30th June 2011) and we defined a milking herd as all lactating animals that existed on a given farm location during a typical milking season in New Zealand from 1st July to 30th June the following year. Since our analyses primarily focused on the influence of animal production levels on movement risks, the study population was also inherently restricted to milking herds that conducted at least one production test during the season and animals that had entered lactation.

5.3.2. Individual-level factors

The data extract was used to generate demographic data and production data for every lactating animal existed on the farm locations during each of the five seasons between 1st July 2006 and 30th June 2011. The demographic factors included the age in years at which they were sold or slaughtered, the frequency that the animal has been sold in the past, and the number of days that an animal had spent on the current farm before being sold or slaughtered. The other key variable was the fate of individual animals during the lactation season (retained, died, sent to slaughter, or sold to other farms). Although the Event data recorded this information for each animal, it was apparent that there were misclassifications in the data. For example, some animals that were recorded as sold were actually slaughtered as indicated by the appearance of a death record in the database. We therefore considered animals were truly sold to other farms only if these animals survived more than 30 days after they were sold. This was verified by referring to the subsequent records of these animals that appeared in the database after the movement. Otherwise they were considered as slaughtered and treated so in the analysis. During the study period, there were 2,667,779 cattle that were recorded to be sold from the milking herds of interest and 153,337 cattle (5.7%) did not survive 30 days after their movements. No animals that were recorded to be culled had any records after the date of culling.

For production variables, we first calculated an average daily milk volume as a sum of AM and PM milk volume if both data points were available. If the average number of milkings was one for a given animal and there was either AM or PM milk volume data, the available data point was used as the value for sum of milk volume. If the average number of milkings was two and there was either AM or PM data, the available value was multiplied by two to represent sum of milk volume. Four production factors used for the analysis—a ranking in percentile of an animal in its herd for each of four production variables (i.e. SCC, milk volume, fat percentage, and protein percentage) on a given year—were then prepared as follows. As explained later, we first randomly chose one observation per animal per herd per year. We then ordered animals in an ascending order for each of four variables, giving a ranking of the animal in the herd. This ranking value was then divided by the herd size, providing a percentile value.

5.3.3. Herd-level factors

Similarly, we obtained herd-level demographic factors and production factors for each of the five seasons between 1st July 2006 and 30th June 2011 by aggregating information about individual animals present in the milking herds. There is no specific data field in the Dairy NZ Core Database that records herd size. Although Production data provides the number of cows tested for production in each herd (N_t), this number can be lower than the actual herd size because not every cow may be tested in a herd. We therefore validated whether N_t accurately represents the true herd size by estimating the herd size in an alternative way; herd size can be approximated as a sum of the number of cows that calved in a herd (N_p) and the number of cows that were purchased from other farms after they calved (N_c). Both numbers were calculated for each herd using Event data. Then, we evaluated how much disparity N_t and $N_p + N_c$ had by calculating percentage error = $100 \times (N_p + N_c - N_t)/N_t$. We chose 10% error as a threshold; that is herds were assumed to have good quality data if $N_p + N_c$ fell off between $0.9\times N_t$ and $1.1\times N_t$. This threshold was arbitrary chosen based on observing that the proportion of eligible herds substantially dropped between at 20% and at 10% (Figure S1 in Appendix 2).

Other demographic factors included the number of animals that exited the herds for each reason (on-farm deaths, sent to slaughter, and sales involving a change in animal ownership), the number of cows that were moved into the herd, the number of distinct farms that a given farm received animals (indegree) from and send animals to (outdegree), the number of years that the farm was in operation, and the geographical region (16 levels, see Figure S2 in Appendix 2). The number of years that the farm was in operation was calculated as follows. We obtained the year each herd started its operation (year 1) by tracing back Production data, which started recording data on 1990. The number of years in operation was then obtained by calculating how many years each herd had been in operation since its year 1. The possibility of left censoring was minimal— the obtained maximum number of years in operation was 12 years, suggesting that none of eligible herds started their operations before 1990.

The production factors included variables representing a herd-average of SCC, milk volume, fat percentage, and protein percentage, respectively. As previously mentioned, one observation was randomly chosen for each cow in a given herd, and these observations were used to calculate herd-average statistics for production variables. The yearly dairy company payout over this time period was obtained from New Zealand Dairy Statistics for the purpose of determining the economic value of animal production (Anon, 2017a).

5.3.4. Selection of eligible animals and herds for the analysis

The overall selection process of eligible animals and herds was shown in Figure 5-1. There was a total of 45,119,122 observations in Production data between 1st July 2006 and 30th June 2011, which represented 5,915,313 unique animals from 12,612 unique herds. We excluded observations that were missing a record for at least one of each four production variables. As described before, we then chose 10% percentage error for the milking herd size as a cut-off value to select eligible herds, resulting in a total of 11,609,095 observations remained. We further excluded observations from which a given herd was in their first or final year of operation because herds should have a very different pattern of selling animals. We then randomly selected one observation per animal per herd per year because, as previously described, one animal may have more than one observations per herd per year. This resulted in 3,304,974 observations from 1,884,524 unique animals from 4410 unique herds. This dataset was used for ***Analysis 2*** (hereafter we refer to this dataset as Analysis 2 dataset), where we conducted a descriptive analysis on the characteristics of four production variables between animals that were slaughtered, being sold to other farms, or not being removed from a farm on a given year.

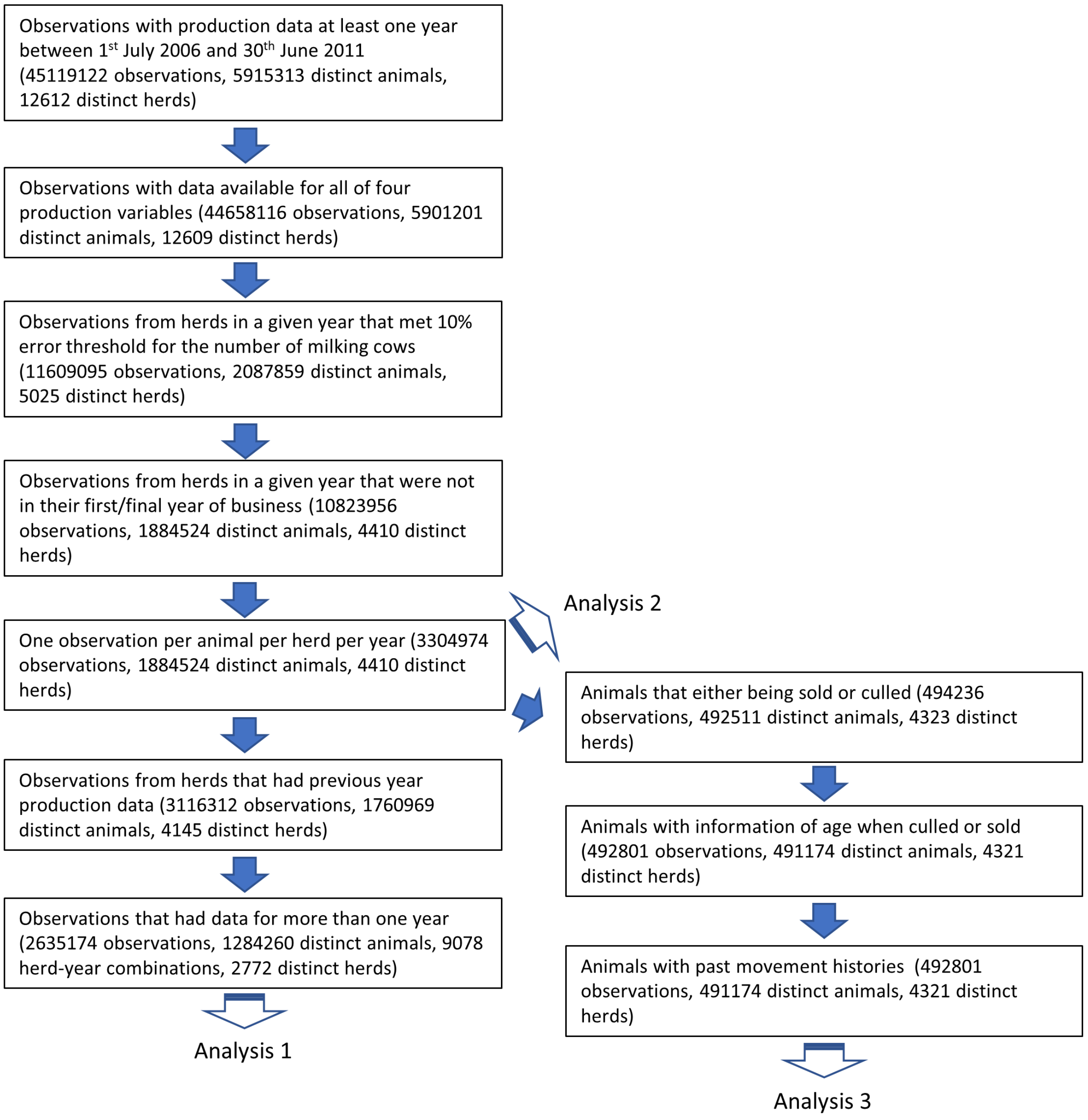


Figure 5-1 A flow chart describing the selection process of eligible observations.

A total of 2,635,174 observations were aggregated on the herd level and used for Analysis 1. Analysis 2 used 3,304,974 observations from 4,410 unique herds. Only observations for animals that were culled or sold in a given year were extracted from these 3,304,974 observations and used for Analysis 3, after removing observations that had missing data for age and past movement histories.

From Analysis 2 dataset, we further removed observations from herds that did not have any production records from the previous year. Subsequently, observations from herds that had observations over more than one year were only included because one of the objective was to investigate the impact of a change in milk price over years on farms' livestock selling performance. The resultant 2,635,174 observations were eligible for ***Analysis 1***, where we investigated how various herd-level factors and milk price influenced the number of animals being sold. Given this analysis was based on a herd level, we aggregated observations over a herd on a given year to obtain herd-level production variables described above.

As follows, we selected data for ***Analysis 3***, where we investigated how various individual-level factors influenced the fate of animals. From Analysis 2 dataset, we selected observations where the animal was slaughtered or sold to other farms from its existing farm in a given year. We then excluded 1,435 observations that did not have age information. All these observations had information for past movements; a total of 492,801 observations from 4,321 unique herds were therefore used for ***Analysis 3***.

5.3.5. Statistical analysis

All the statistical analyses were performed using R version 3.4.2.

5.3.5.1 Descriptive statistics

All relevant national statistics for demographic and production variables were retrieved from New Zealand Dairy Statistics (Anon, 2011). The average milk volume per cow per day was obtained by dividing the average milk yield per cow over a season by the average number of days in milk, which was estimated by Dairy NZ (Anon, 2011). As a reference, we also calculated the time an animal spent in the existing herd for survived cows, by obtaining the number of days between the day the animal had moved into or been born at the herd and the herd test date at which the animal was tested.

5.3.5.2 Analysis1

The associations between herd-level factors and the number of animals sold to other farms on a given year (Yt) were explored by using a zero-inflated count regression model. Each herd therefore could have maximum five observations—one observation per year for five seasons. A zero-inflated model was chosen because the large number of farms did not sell any cattle to other farms. The outcome was the number of sold cows in year Yt. Explanatory variables included: the number of animals that died and were slaughtered, respectively, in the previous year (Yt -1), the number of adults a herd purchased in Yt -1, the number of milking animals in Yt, indegree and outdegree in Yt -1, the number of years that a farm had been in an operation, the geographical region, herd average for milk volume, SCC, fat percentage, and protein percentage in Yt -1, and milk price in Yt. All explanatory variables except geographical region were treated as continuous variables and their best-fit forms were explored as described below.

A regression model was built following a method previously described (Hidano et al., 2016). In short, a manual forward variable selection was performed. Sets of explanatory variables for the binary part of the model were first identified using a logistic regression with outcome being the presence or absence of selling animals. Then, variables for the count part of the model were identified using a Poisson regression model. A fractional polynomial was fitted to each of continuous variables to identify the form that fits data best, selecting a variable that has p-value < 0.25 in a univariable analysis using the R package mfp (Ambler and Benner, 2015). Selected variables were then added to a multivariable model one by one with the order of the strength of p-value from the univariable analysis using the R package pscl (Jackman et al., 2017). Variables were retained in the final multivariable model if the p-value from the Wald test was <0.05. Then, we re-evaluated the form of continuous variables that remained in the final model: we applied the mfp function to one variable at a time under the presence of other variables. Finally, we compared the goodness-of-fit between the zero-inflated Poisson model and the zero-inflated negative binomial (ZINB) model by computing the shape (theta) parameter of the negative binomial. The ZINB model was chosen as the theta parameter was significantly different from 0 (Dohoo et al., 2009).

Parameter coefficients were then estimated using the herd identifier as a random effect both in the binary and count part. To the best of our knowledge, there is no R package that can handle both random effects and covariates in the zero-inflation part under frequentist framework; hence the parameter estimation was implemented under Bayesian framework using the R package brms (Bürkner, 2018), which uses Stan for Bayesian inference (Stan Development Team, 2018). Noe that a compound symmetry was used as a correlation structure for a random effect because other structures were not available in the package—a sensitivity analysis will be required to examine the effect of the correlation structure in a future study. All priors were set to default values (Bürkner, 2017) and four independent chains were run for 2,000 iterations each. The first 1000 iterations were discarded as a burn-in and the second 1,000 iterations were used for an inference, resulting in 4,000 iterations in total from four chains. Note this iteration number may appear to be small; however, Stan uses Hamilton Monte Carlo Sampler, which provides higher quality samples and convergence occurs with a fewer iterations than other methods such as Metropolis-Hastings sampling (Bürkner, 2017). Four chains were run parallelly, taking approximately one hour using Core i7 2.60Ghz with 32.0GB of RAM. Convergence of parameters of interest was visually confirmed as well as ensuring Rhat value being 1.0. The probability of a farm not selling cows and the number of cows being sold given a farm was selling at least one cow were then predicted using parameters estimated as follows. By sampling one posterior value for each parameter, the outcome was predicted for each of 9,078 observations by replacing its value for the variable of interest by a value for which a prediction was to be made. The probability of a farm not selling was calculated as a proportion of 0 count among 9,078 outcomes. The number of cows being sold given a farm was selling at least one cow was obtained by averaging the outcome that was not 0. This process was repeated for 1,000 times, sampling different posterior value at each time, to provide the median and 95% credible intervals of the outcome.

5.3.5.3 Analysis2

A descriptive analysis was performed to compare the relative production performance of three different animal populations: those were not removed from a given herd on a given year (hereafter refer to as ‘survived’), those sold to other farms (‘sold’), and those being slaughtered (‘culled’). Within a given herd, we ranked each individual animal in the ascending order for each of four production variables (SCC, milking volume, fat percentage, and protein percentage). Each animal was then assigned to a percentile value for each of variables, which was calculated by dividing its rank by the number of milking cows in a given herd. For example, an animal with 95th percentile for milking volume and 90th percentile for SCC is one of the top producing but has one of the highest SCC in the herd. The distribution of these percentile values over all herds and years was then obtained for each of survived, sold, and culled groups. We then used a Gaussian kernel estimate with its bandwidth being chosen based on the SJ method (Sheather and Jones, 1991).

5.3.5.4 Analysis 3

Finally, the associations between the fate of animal (i.e. being slaughtered or sold given it was removed) and individual-level factors were investigated using a mixed effect logistic regression model. The outcome was animal’s fate, binary sold (coded 1) and slaughtered (coded 0). Explanatory variables included: the number of previous sales the animal had experienced, number of days the animal stayed on the current farm before being removed, age in years, and animal’s percentile ranking for each of four production variables (i.e. milk volume, SCC, protein percentage and fat percentage) described above. Because the herd test dates were different for each animal and this could potentially confound the effect of production variables due to their seasonality, we included a variable representing the interval between the calving date and the herd test date. A regression model was built in the same way treating all variables as a continuous variable and check their best forms using a fractional polynomial as described above. The final multivariable regression was developed by adding a significant variable one by one. Variables were retained in the final multivariable model if the p-value from the Wald test was <0.05. A combination of herd identifier and year was used as a random effect variable to account for the potential similarities between fates of animals from the same herd. Given that it is not straightforward to interpret coefficients of fractional polynomial models, results were shown by displaying the predicted using different values of a given continuous variable. A predicted value and its 95% confidence interval (CI), where appropriate, were obtained for each continuous variable using bootstrapping as follows. A total of 1,000 observations were randomly selected from the original dataset with replacement and their values of the continuous variable of interest were set to a value that was to be predicted. Using the obtained regression coefficients, the outcome was predicted for each of 1,000 observations and we calculated a mean of the predicted outcomes. This bootstrapping process was repeated for 100 times, providing a distribution of the mean value. The mean, 2.5th percentile, and 97.5th percentile values of this distribution were shown in the figures.

5.4. Results

5.4.1. Descriptive statistics of eligible herds and cows

As shown in **Figure 5-1**, a total of 2,772 distinct herds were eligible for the Analysis 1. These herds provided 9,078 unique herd-year data points between 2006 and 2010 seasons. Figure 5-2 summarises herd-level statistics for key demographic and production variables. The mean herd size of eligible herds was slightly lower than the national average herd size for each year. The previous year's national average statistics ranged between 17.4 – 18.0 litres for milk volume per cow per day, 232 – 253 (×10^3) per millilitre for SCC, 3.68 – 3.76 % for protein percentage, and 4.68 – 4.85 % for fat percentage between 2006 and 2010 seasons. The minimum herd-average milk volume per cow per day among eligible farms was 5.5 litres and the maximum was 37.3 litres for the same period. The minimum herd-average SCC among eligible herds was 30.9 and the maximum was 1044.1 (×10^3) per millilitre. The minimum and maximum herd-average fat percentage among eligible farms was 2.4% and

7.3%, respectively. The minimum and maximum herd-average protein percentage was 3.0 and 5.0%, respectively.

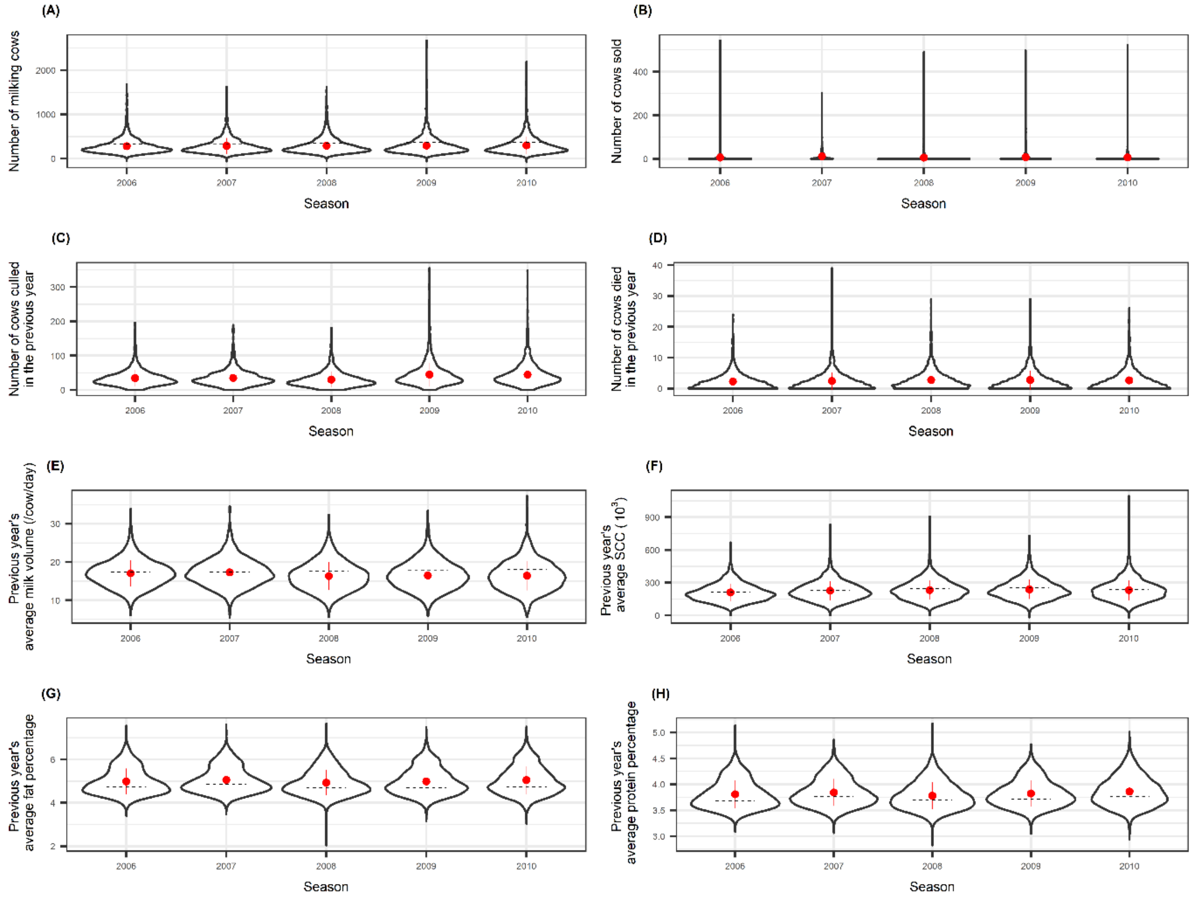


Figure 5-2 Violin plots that describe statistics of key herd-level demographic and production variables

(A) number of milking cows, (B) number of cows sold from each herd, (C) number of cows culled in the previous year, (D) number of cows died on-farm in the previous year, (E) average milk volume per cow per day in the previous year, (F) average SCC (×10^3) per millilitre in the previous year (G) average fat percentage in the previous year, and (H) average protein in the previous year. The red circle and horizontal line represent the mean value and one standard deviation for a given variable over all eligible herds in a given year. The black dotted line, where available, represents the mean value for each variable extracted from the National statistics.

As Figure 5-2 (B) shows, while a large proportion of farms did not sell any animals, some farms sold a substantial number of animals. A total of 975 farms (59.7%) in 2006, 726 farms (40.0%) in 2007, 1198 farms (63.4%) in 2008, 1124 farms (57.2%) in 2009, and 1134 farms (63.8%) in 2010 did not sell any cows during the year. Figure 5-3 shows the overall distribution of the proportion of animals being sold to the total number of milking cows in a given herd between 2006 and 2010 season. Among 2,772 unique farms that were eligible for Analysis 1, 659 farms (23.8%), 716 farms (25.8%), 421 farms (15.2%), 247 farms (8.9%), and 163 farms (5.9%) did not sell animals for one, two, three, four, and five seasons, respectively. Note that not all farms had data for whole five seasons; 804 farms (29.0%) did not sell a single cow in seasons in which data for these farms were available. Figure 5-4 shows the spatial distribution of eligible farms. Among eligible herds, a total of 961 unique farms (34.7%) located in Waikato region, where 30.4% of dairy herds in New Zealand existed in the 2010 season (Anon, 2011).

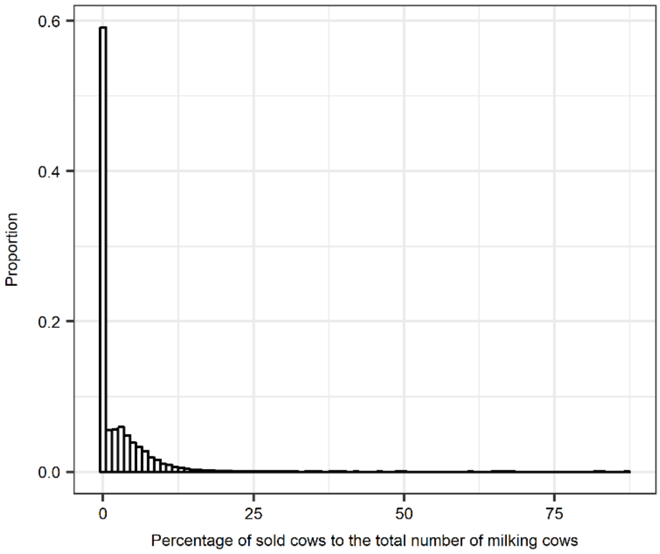


Figure 5-3 Distribution of the percentage of cows being sold from each herd between 2006 – 2010 among New Zealand dairy herds.

X-axis shows the proportion of cows being sold to the total number of milking cows in each herd on a given year and Y-axis shows the proportion of farms that had a given percentage of selling.

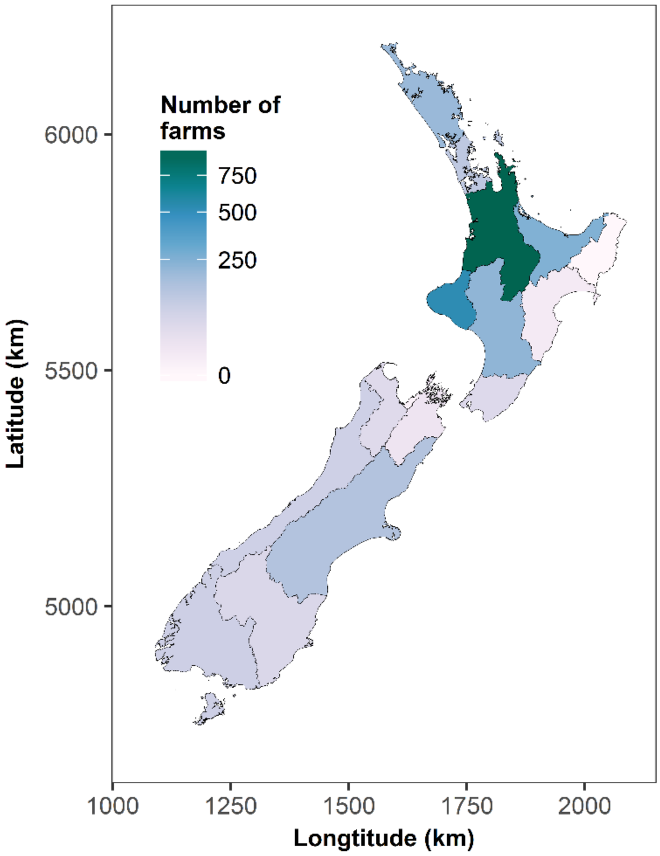


Figure 5-4 Distribution of farms that were eligible for the Analysis 1 between 2006 – 2010 seasons among New Zealand dairy herds.

The number of eligible farms was counted for each of 16 geographical regions in New Zealand and each region was coloured based on the count; the darker colour indicates that the larger number of farms in a given region were eligible for the analysis.

Figure 5-5 shows the descriptive statistics of individual-level demographic variables for each year. Distributions of each variable were shown separately for survived, sold, and culled cow populations. As can be seen, there was no substantial variation in each statistic between seasons. The exception is the frequency of sales and days the animal spent in the current herd for the survived population in 2009. Median ages over seasons were 5, 3, and 7 for survived, sold, and culled populations, respectively. Median frequencies of sales were 1, 0, and 0 for survived, sold, and culled populations, respectively. Median days the animal spent in the current herd were 176, 1366, and 2443 for survived, sold, and culled populations, respectively.

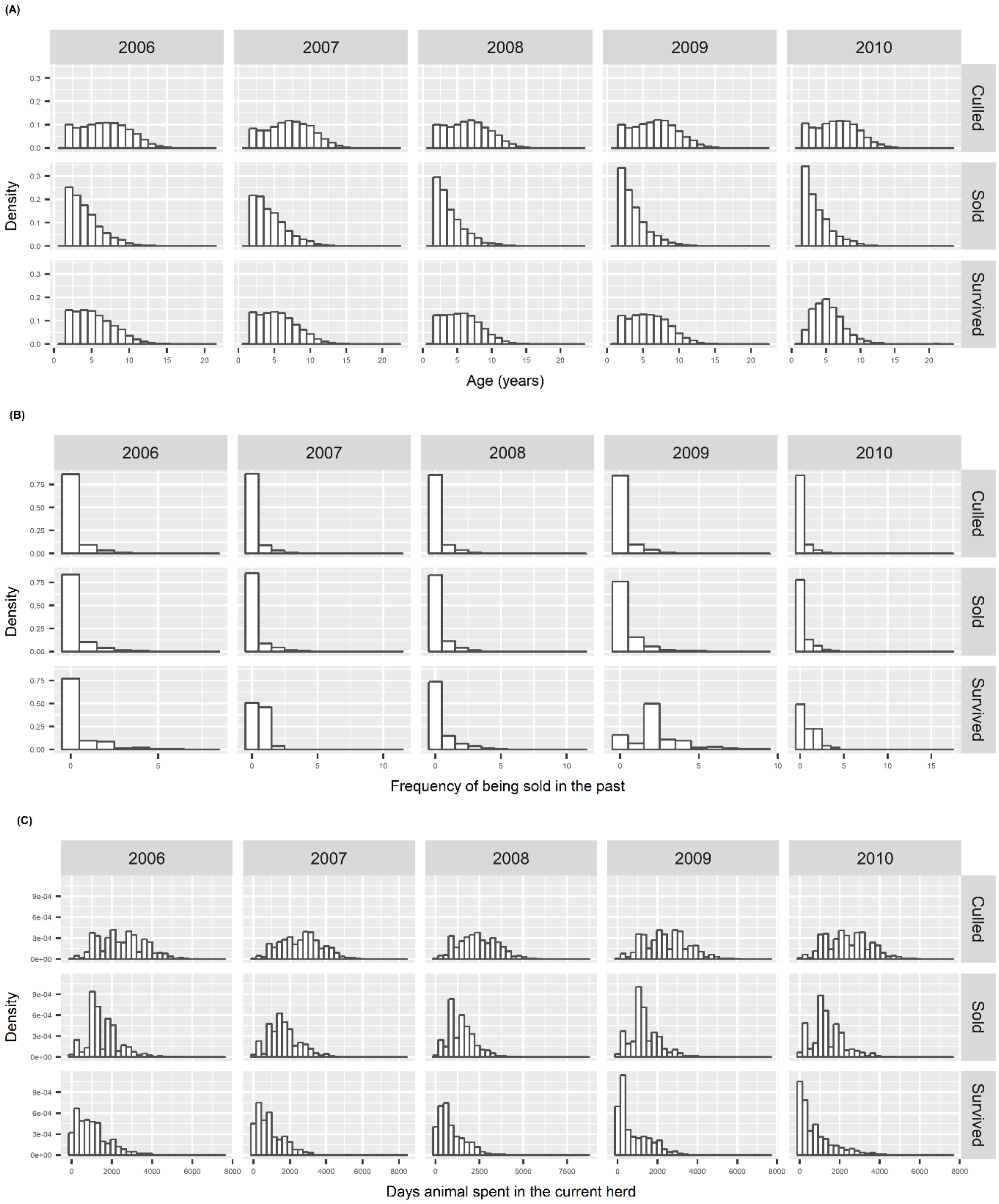


Figure 5-5 Histograms showing the distributions of demographic variables for survived, sold, and culled cow populations between 2006 and 2010.

Demographic variables included (A) age in years, (B) frequency of animals being sold in the past, and (C) days that the animal spent in the current herd.

5.4.2. Analysis 1: Factors associated with farms not selling any cows

Figure 5-6 shows the predicted probabilities of farms not selling any cows for milk price, three herd-level production and seven herd-level demographic factors that were identified to be significant in the final multivariable ZINB model after controlling for the geographical region and a herd random effect. A higher milk price reduced the probability of farms not selling any cows; that is, the higher milk price is, the more likely it was that farms sold at least one cows. All four production variables were significantly associated with the probability of a farm not selling any cows in the univariable analysis. However, the significance was lost for the variable representing for protein percentage in the multivariable analysis, which could be due to the high correlation between the fat and protein percentage (rho = 0.868, $p < 0.001$ in Spearman correlation test). The higher the average SCC, farms had a higher probability of not selling cows. Similarly, farms with a higher milk volume and fat percentage had a lower probability of not selling cows.

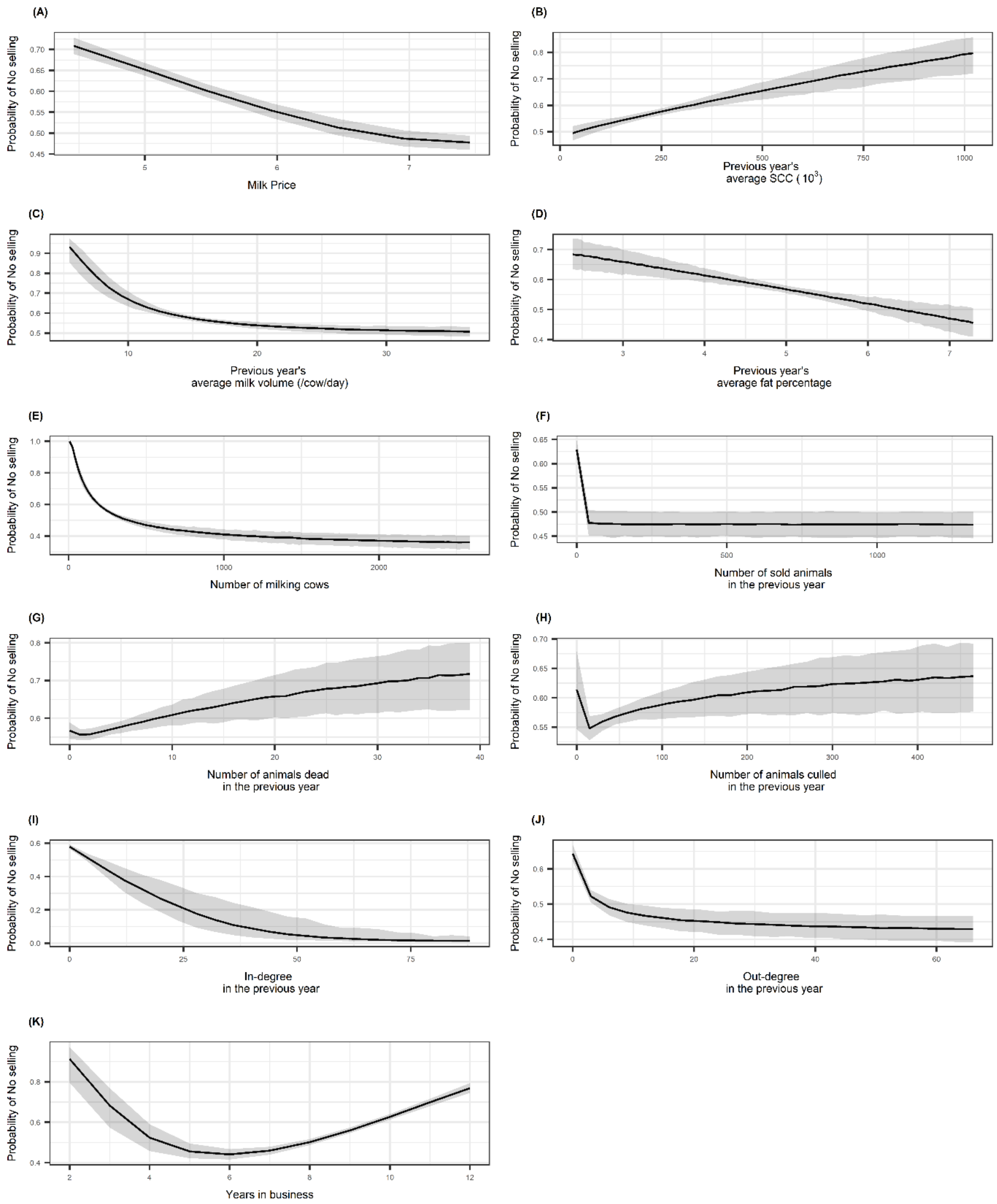


Figure 5-6 Predicted probability of a farm not selling any cows from the final zero-inflated negative binomial model.

Y-axis shows a probability that a farm was not selling any cows in a given year as a function of each of different continuous variable. Black line and grey areas represent the median and 95% credible intervals of the outcome obtained through 1,000 iterations.

Regarding demographic factors, a larger herd size was associated with a lower probability of not selling cows. The probability of not selling cows substantially dropped as the number of sold cows in the previous year increased up to 10; however, the predicted probability remained at the same level for farms that sold more than 10 cows. The lowest probability of not selling was predicted for farms in which two cows died in the previous year. Similarly, farms that had culled 15 cows in the previous year were predicted to have the lowest probability of not selling cows. Both larger in- and out-degree in the previous year were associated with a lower probability of not selling any animals. The probability of not selling was lowest for farms that had been in an operation for six years; those with shorter or longer business periods had a higher probability of not selling. Estimated coefficients can be found in Table S1 in Appendix 2.

5.4.3. Analysis 1: Factors associated with the number of cows being sold

Figure 5-7 shows the predicted numbers of cows being sold as a function of each of continuous variable that remained in the count part of the final ZINB model. There was a general trend that shows the higher milk price is associated with a higher number of cows sold. Among four production variables, only variables representing SCC and fat percentage were associated with the number of cows being sold. The higher the SCC a farm had, the fewer number of cows that were sold. Having a higher fat percentage was associated with a larger number of cows being sold.

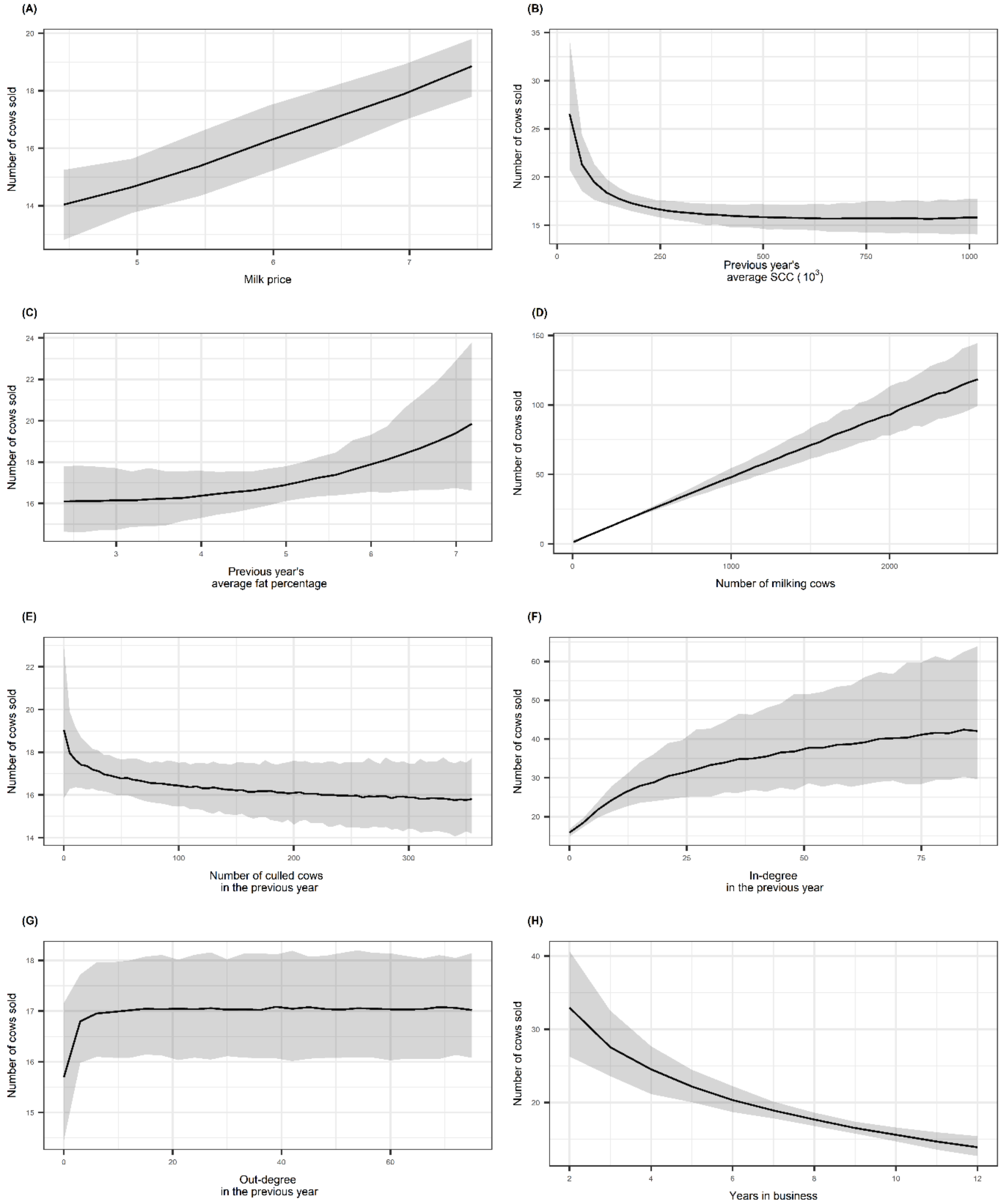


Figure 5-7 Predicted number of cows being sold from the final zero-inflated negative binomial model.

Y-axis shows the predicted number of cows sold given that a farm was selling at least one cow in a given year. X-axis shows a continuous variable of interest. Black line and grey areas represent the median and 95% credible intervals of the outcome obtained by 1,000 iterations.

Regarding demographic factors, larger herds sold a larger number of cows. After controlling the effect of the number of milking cows in a herd, there was a still negative association between the number of cows sold and the number of cows culled in the previous year. Larger in- and out-degree were associated with a larger number of cows being sold. Farms with a longer business period had a lower number of cows being sold. Estimated coefficients can be found in Table S2 in Appendix 2.

5.4.4. Analysis 2: Overall trend of production characteristics between survived, sold, and culled cows

Figure 5-8 shows the distribution of ranking for each of four production variables (SCC, milk volume, fat percentage, protein percentage) for survived, sold, and culled cows. Median ranking for SCC was 43.4 %, 42.4 %, and 56.6 % for survived, sold, and culled cows, respectively. For milk volume, median ranking was 44.1 %, 44.0 %, and 49.0 % for survived, sold, and culled cows, respectively. For both SCC and milk volume, survived and sold cows had very similar distributions of ranking values. Culled cows had a high ranking for SCC; culled cows had a higher SCC compared to other herd mates. Culled cows either had a high or very low ranking for milk volume. The distinction between survived and sold cows is clearer for fat percentage and protein percentage; sold cows had a lower ranking in these two variables (median rankings for fat and protein were 42.2 % and 42.0 %, respectively) compared to survived cows (median rankings for fat and protein were 38.4 % and 37.7 %, respectively).

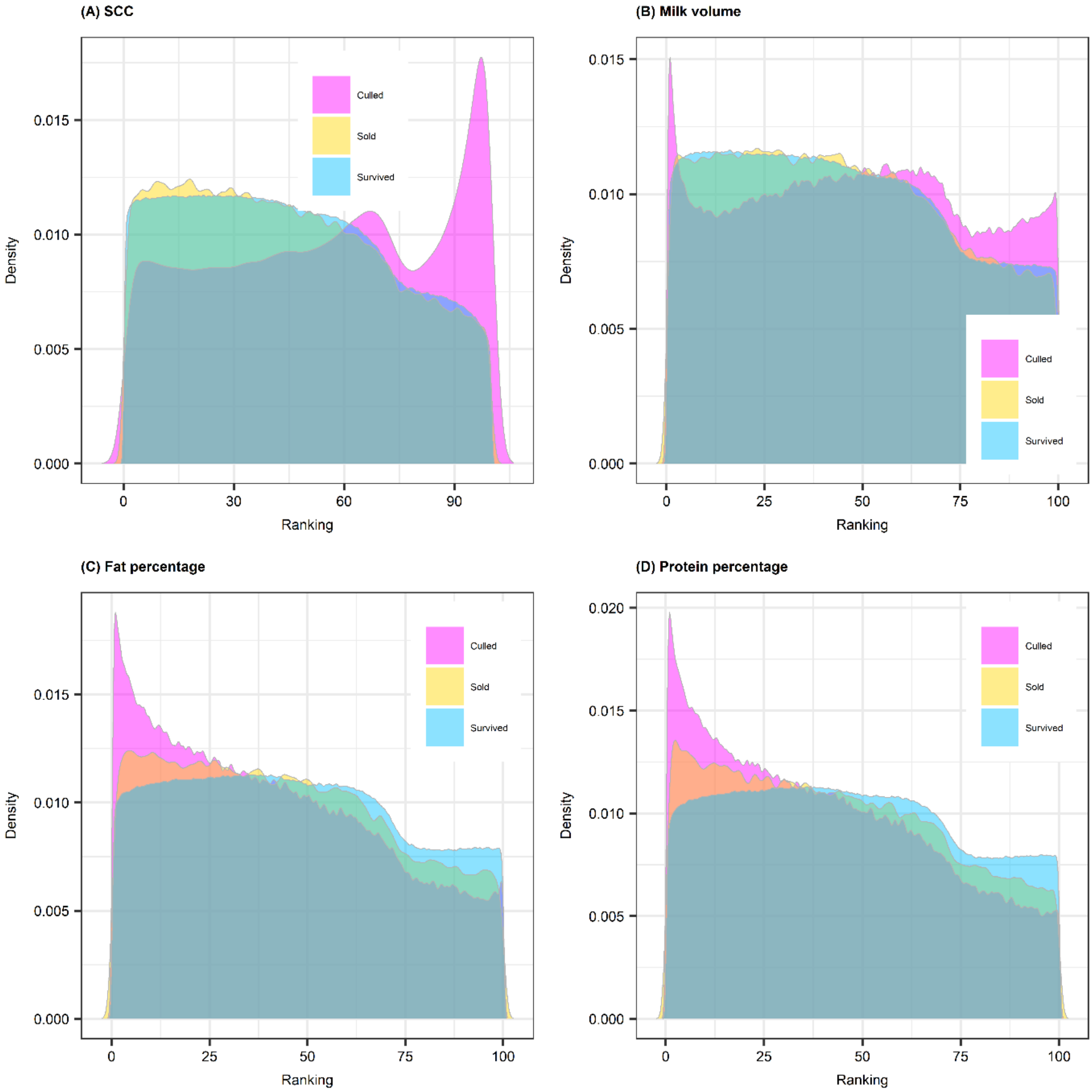


Figure 5-8 Distributions of individual animal's ranking in its herd for (A) somatic cell count (SCC), (B) milk volume, (C) fat percentage, and (D) protein percentage for survived, culled, and sold animal groups.

The ranking value ranges between 0 and 100, representing a percentile which indicates where the animal placed when all animals from the herd were ordered in the ascending order of the production factor of interest. For instance, an animal with a higher ranking in SCC and milk volume means it had a high SCC and high milk volume relative to other herd mates in its herd.

5.4.5. Analysis 3: Individual-level factors associated with an animal to be sold rather than being culled

Figure 5-9 shows the predicted probabilities of a cow to be sold to other farms rather than being culled given that the animal was going to be removed from a herd at the end of the season. After controlling for cow's age and its production ranking, a cow that was sold more frequently in the past had a higher probability of being sold again rather than being culled. For instance, cows that had experienced 17 past sales were approximately three times more likely to be sold again compared with cows that had experienced only five past sales. The probability of being sold rather than culled was higher for cows that stayed longer in the existing farm. An increase in the age was associated with a lower probability to be sold—the older a cow was the more likely it was culled. Three production factors—rankings for milk volume, milk fat percentage, and milk protein percentage—had similar trends; the higher ranking the higher probability of a cow to be sold. For SCC, the higher ranking the lower probability to be sold, suggesting cows with high SCC were more likely to be culled rather than being sold. Estimated coefficients can be found in Table S3 in Appendix.

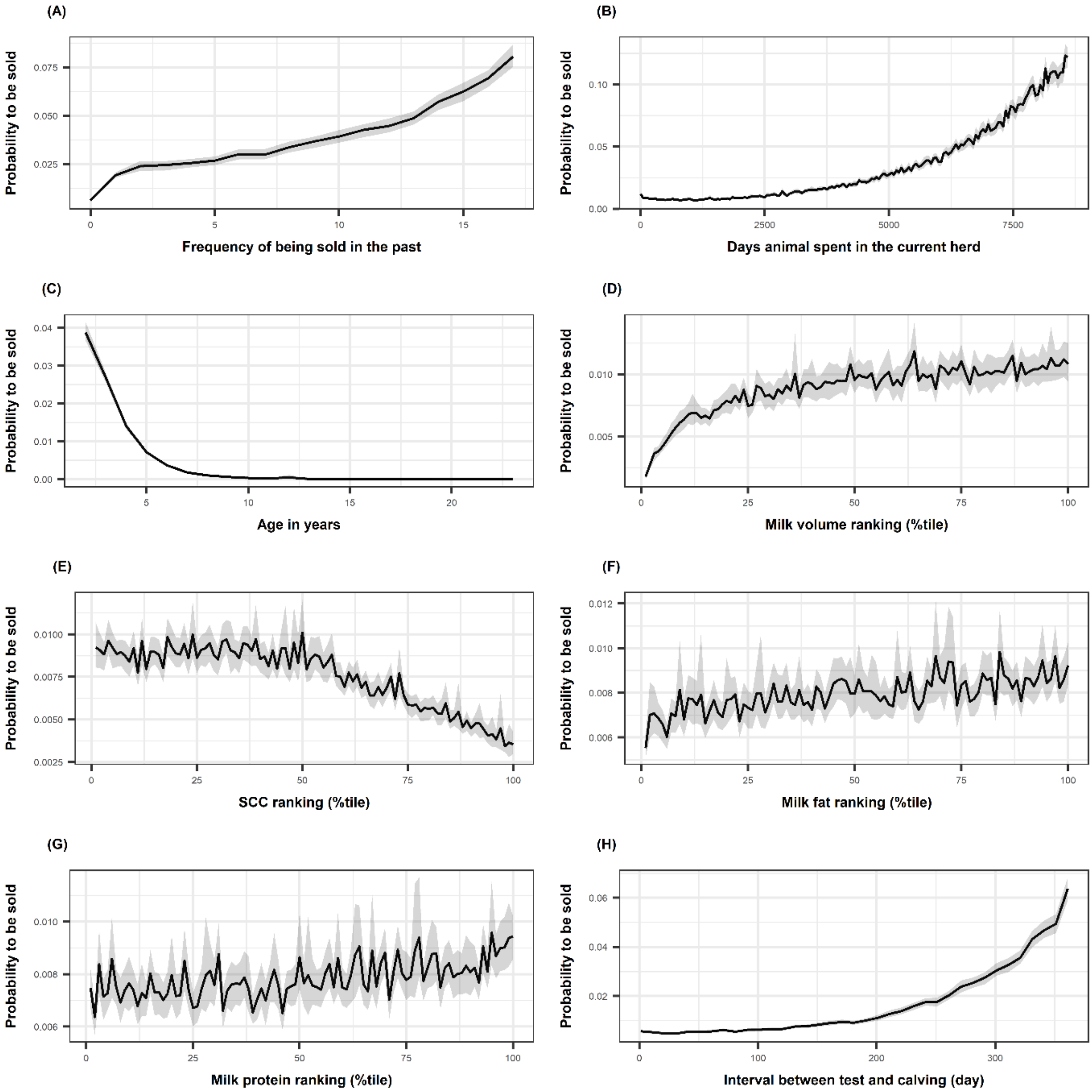


Figure 5-9 Results of the mixed effect logistic regression model.

Y-axis represents the predicted probability of being sold rather than culled given that the animal was removed from a herd. X-axis represents each of continuous explanatory variables. The black line and the area in grey represent the mean and 95% confidence interval of the predicted value, respectively.

5.5. Discussion

To the best of our knowledge, this analysis is the first to explore animal and farm level risk factors for the between-herd sales of dairy cattle. Our findings showed that the three key assumptions made in livestock disease simulation models are likely to be violated. First, the likelihood of farms selling cows in any given year is dynamic and influenced both by external (e.g. market milk price) and internal (e.g. demographic and production characteristics) factors. Second, animals that are sold to other farms are unlikely to be chosen randomly given the differences in production characteristics between animals that survived versus those that were sold or culled. Third, cows' fate if they are culled or sold is not a Markov process with cows' past sales histories likely to influence their fate. Although the impact of the violations of these assumptions is not evaluated in this study, we contend that this impact needs to be carefully examined particularly for slow spreading diseases. Below, we discuss our findings related to each of three assumptions and their implications.

5.5.1. Farmers' cow selling practices are dynamic

There are two points that need to be considered for disease simulation models that account for dynamic change in farmers' cow selling practices. One is the quantitative change—the change in the number of cows being sold from a farm. The second is the qualitative change—whether or not a farm sells cows in a given year. Our results from the ZINB model provided useful information to understand these two changes. Higher milk price was predicted to increase the number of cows being sold from each farm as well as the number of farms that sell cows. The association with milk prices is consistent with results from our previous qualitative interview study that suggested farmers try to increase their herd size when the milk price is high so that they can produce a larger quantity of milk (Hidano et al., 2018b). Other published studies have also suggested that market price plays a substantial role in livestock trade dynamics (Bensemann and Shadbolt, 2015; Delabouglise et al., 2016). For example, an Australian study exploring the factors associated with selling and buying practices among beef producers found that market price was one of the most important factors behind these decisions (Dyall et al., 2010). To the best of our knowledge, however, only limited number of studies quantitatively analysed the association between livestock price and livestock movement patterns. An exception is the previous study that attempted to predict the probability of a livestock movement to occur as a function of livestock prices between two geographical regions in Cambodia (Madin, 2011). Future studies are warranted to take a closer look at how market price fluctuations would influence farmers' selling and buying practices, and subsequently, the overall livestock industry trading network.

All production variables except for protein percentage were significantly associated with the probability of a farm not selling any cows. These results were intuitively reasonable; farms that had the lower probability of not selling (i.e. high probability of selling) were those with lower SCC, higher milk volume, and higher milk fat percentage. The number of cows being sold was also associated with SCC and fat percentage. Previous research has shown that lower SCC levels generally indicate better milking shed hygiene and better farm management practice such as a use of milk recording scheme (Kelly et al., 2009). Milk fat and protein percentage and milk volume all contribute to Breeding Worth (BW), which is a metric used in New Zealand to represent the genetic value of cattle (Anon, 2018a). Therefore, it is reasonable cows that produce a higher milk fat percentage are in a higher demand. We note that although the protein percentage did not remain in the final model, this may be due to the high correlation between the fat percentage and protein percentage. We did not, however, observe a significant association between milk volume and the number of cows being sold. One potential explanation for this is that farmers may simply pay more attentions to the fat and protein percentage rather than the milk volume given that farmers are paid based on their production of milk solid rather than milk volume in New Zealand (Anon, 2017a). Anecdotally, high milk yields can also potentially be unfavourable in years where the pasture quality is insufficient to support the animals' energy demands. Further studies are required to investigate how each of production characteristics including SCC, milk volume, fat and protein percentage, and BW influences cows' price on the market and how farmers take this into account when purchasing cows.

Interestingly, while farms that sold less than 10 cows in the previous year had a high probability of not selling cows, the probability remained at a similar level for all farms that sold greater than 10 cows irrespective of their herd size and production characteristics. This suggests that there seems to be two distinct populations; farms that only sporadically sell cows and farms that remain consistently active in selling cows. Our previous qualitative study suggested that there are at least two types of dairy farms in New Zealand; one is more engaged in livestock selling, which may form a substantial part of the total farm income, and the other is farmers that do not rely on making profit by selling cows. The latter identify themselves as a milk producer and not a trader, which they consider is a common practice of beef and sheep farmers.

Farms with low numbers of culled or dead cows in the previous year in general had a lower probability of not selling cows. The number of cows being sold was also larger for farms that culled less animals in the previous year. These may simply represent that these farms had more surplus cows to sell. Alternatively, these two variables may reflect the quality of farm management; farms with fewer deaths and culls may indicate a good farm management. Many studies in the literature have confirmed that farmers that take better care of livestock are considered to be good farmers (Burton, 2004; McAloon et al., 2017; Shortall et al., 2018); given that high mortality is known to reflect poor health and welfare status of animals (Alvåsen et al., 2012; McConnel et al., 2008; Raboisson et al., 2011; Reimus et al., 2018), it may be reasonable to assume that farms with fewer deaths are considered to have a good reputation. Our previous qualitative study suggested that some farmers are more willing to purchase from farmers with a good reputation (Hidano et al., 2018b), therefore this may explain why these farms sold a larger number of cows. Nevertheless, recent studies highlighted that farmers' culling decision making is complex (Bergeå et al., 2016; Haine et al., 2017)—further studies are required to understand interrelationships between livestock culling and selling practices.

Farms that had been in business for at least six years were predicted to have the lowest probability of not selling any cows. Unfortunately, we cannot conclude whether this effect truly represents how farms' selling practice would change as business goes longer, or this reflects the specific characteristics of dairy farmers who had started business at least six years prior to this study period. It is known that many sheep and beef farms had converted into dairy farms in the early 2000 (Anon, 2017b); these farms in business for six years may be those had converted from sheep and beef farms and more willing to trade cows, although we have no data to formally examine this. The longer business period was also associated with the lower number of cows being sold. One potential explanation is that farms in business for a longer period may have established a business style that does not rely on selling cows for profit, which can be very vulnerable to the fluctuation of market prices. These farms may keep a minimal number of replacement animals (and hence fewer animals to sell) because it is costly to have extra number of animals. Alternatively, the duration that a farm has been in operation may be highly correlated with other variables that are known to influence farmers' behaviours such as age, experience, quality of farm management, and reputations in the area (Burton, 2014, 2006). It is important to better understand how farmers change their farm management, trading practices, and business style over years as well as how farmers' farming background influences these behaviours.

Variables representing farms' in-degree and out-degree in the previous year were associated with a higher probability a farm not selling any cows. These metrics are likely to indicate the degree of direct and indirect connections that a farm has, which represent farms' previous livestock trading practices. Again, interviews from our previous qualitative study demonstrated that stock agents—who often mediate livestock trading in New

Zealand—establish a connection with farmers so that they can maintain the supply-demand relationship that is critical for stock agents' business (Hidano et al., 2018). Therefore, farmers that are more connected to other farms through trading (i.e. farms with a higher indegree and outdegree) are more likely to remain engaged in trading cows every year, explaining why these farms have higher probabilities of selling cows. Farms with a larger in-degree were predicted to sell a larger number of cows. These farms are more likely to run a 'trading' business style; for instance, these farms may farmers purchase non-pregnant cows and sell them in the following year after they become pregnant. Given these farms—with high in-degree and large number of animals sold—have a potential to play a role in spreading diseases between farms, we need a better understanding on how these farms operate their business such as how and where they purchase cows and to which extent they maintain their biosecurity levels.

5.5.2. Individual cattle characteristics that determine their fate

Overall, *Analysis 2* and *Analysis 3* clearly suggested that sold cows are not chosen randomly from the herd. While the distributions were similar between survived and sold cows for both SCC and milk volume, those of culled cows were very different. The high SCC density at the high ranking for culled cows is reasonable given that high SCC is well-known to increase the likelihood of a cow to be culled worldwide (Hadley et al., 2006; McDougall et al., 2009). Lower milk production is also known to be associated with a higher likelihood for a cow to be culled (Hadley et al., 2006), which is likely to explain the high density at the low milk volume ranking for culled cows. Our explanation for the high density observed at the high milk volume ranking is that these high producing cows may have serious udder issues. In an intensified system producing high milk volume, cows tend to have udder problems (such as too high udder depth and loose fore udder attachment) after a few years of milking and they need to cull these animals. Various studies indeed reported the negative associations between milk yield and udder traits including fore udder attachment, udder depth, and front teat placement (DeGroot et al., 2002; Misztal et al., 1992). However, we need to use caution in interpreting these findings. As we discuss below, farmers may consider multiple production characteristics together, rather than consider each of them independently when making sales decisions. The genetic merit and economic value of cows is mainly measured based on two metrics in New Zealand; BW, which we discussed above, and Production Worth (PW). Both BW and PW are calculated according various individual characteristics including milk volume, fat, and protein percentage (Anon, 2017a). These metrics were not available for this study; it would be interesting to study how much farmers value and rely on these metrics when making decisions.

As we hypothesised, the past movement history of animal was significantly associated with a probability of the animal to be sold given it was removed from the herd. Animals that were sold more frequently in the past were more likely to be sold rather than being culled after controlling the effect of other factors including age. There are at least two potential explanations for this finding. First, these animals may have some characteristics that were not observed in this study but that motivate farmers to sell these animals such as poor temperament and behaviour problems. A previous study investigated how farmers' decision of culling is influenced by animals' characteristics related to management (i.e. traits other than production; TOP) including adaptability to milking, milking speed, and animal's temperament (Berry et al., 2005). This study identified that the risk of cows being culled with undesirable scores in these three traits was significantly but weakly higher than other herd mates. Farmers interviewed in our previous study suggested that purchased cows can disturb the smooth milking flow if they do not adapt well to the farm environment (Hidano et al., 2018b). It is therefore plausible that farmers may sell cows that have undesirable TOP, but still produce a sufficient amount of milk, which can result in these cows being sold from farm to farm more frequently. The second potential explanation is that, as discussed earlier, some farms are likely to run a specific business style where they purchase cows and sell them after some time. The longer duration of stay on a given farm was associated with a higher probability of being sold rather than culled. These populations may represent good quality of cows that had survived longer in herds. Once these cows are not required by one farm, there may be a still demand for these cows on other herds. The quality and usefulness of national database for trading such as the one used in this study can be significantly improved by recording detailed and actual reasons for trade; the majority of reasons for trading in this study was ‘other cause’ or not specified, which prohibited us to identify reasons that are not related to production.

5.5.3. Study limitations

There were several limitations in this study related to the challenges of using nationally recorded data to make inferences about farm behaviour. The largest limitation is that the study findings may not be generalisable to the whole New Zealand dairy farm population. Between 2006 and 2010 seasons, approximately 70% of dairy herds in New Zealand used the herd testing service (Anon, 2011). The herd test use seems to have a substantial geographical variation; for instance, while 81.4 % of herds in Taranaki used the herd test service, only 62.2 % used it in South Canterbury in the 2010 season. This regional variation can be attributable to various factors. For instance, Taranaki is the dairy farming region with a long history, and South Canterbury is a relatively new dairy area with many sheep and beef farms having converted into dairy farms over the last decade, indicating these populations can have different farming goals, attitudes, and business styles. The latter has also a substantially large herd size, which may have installed their own milk quality control systems and do not need to perform an external herd test. Furthermore, the fact that the longest duration that a farm had been in business being 12 years among eligible herds poses a question to the representativeness of the studied population because some herds in New Zealand should be running business for much longer. As discussed already, this short duration is not due to censoring because this database runs since 1990.

In order to avoid making erroneous inferences arising from poor data quality, we also employed rigorous inclusion criteria to remove herds that had perceived poor data quality in a given year. Although this was inevitable given the lack of data for herd size in this database system, this may further limit the generalisability of our study findings. For instance, because the herd test cost increases as the number of animal tested increases, farmers may only test a selection of cows in the herd. This selection, however, likely depends on the purpose of the herd testing; while some farmers may test their cows to make a culling decision, others may do so to choose which cows to use for breeding. Farm characteristics and hence trading patterns may be very different between these two types of farms, which were not included in this study. It is therefore difficult to infer how the exclusion of these farms would have affected our results. Further research is required to understand why farmers may or may not choose to perform herd testing so that the results can be adjusted to eliminate the effects of bias.

One might argue the random selection of one herd test result per cow per herd per year may have introduced bias in our results because milk production characteristics change during the milking period. This random selection was, however, unavoidable because the herd test dates for each cow were heterogeneous. The first reason for this is that, again, some farmers tested only a subset of cows at a time. The second reason is that cows entered into and exited from herds at any time during the milking season. These made it difficult to choose a similar herd test date for all cows in the herd. We contend that the bias due to this selection process on herd-level production characteristics is minimal given that all herds were equally subjected to this selection. For individual-level production characteristics, we adjusted this potential bias by including the variable representing the interval between calving and herd test.

The estimated effects of milk price on farmers' selling practices also require a cautious interpretation because the results could be confounded by other factors that had changed between the 2006 and 2010 seasons. For instance, given the pastured-based production system in New Zealand, farm management is significantly influenced by weather conditions such as drought, which influences the availability of fodder. This in turn affects farmers' trading practice; for instance, farmers may attempt to reduce the herd size when fodder is limited. It is also likely that milk price itself influences the use of the herd testing—for instance, the low use (67.3%) of the herd testing in the 2009 season may be attributable to the low milk price at the start of the 2009 season (Anon, 2011). Because the low milk price is likely to shrink the demand for live milking cows—and increases the number of cows to be culled—farms that did not perform the herd testing in the 2009 season because of the low milk price unlikely sold many cows. That is, the number of cows sold captured by the database may be underestimated. Should this be true, we may have underestimated both the positive association between milk price on the number of cows being sold and the negative association between milk price and the probability of a farm not selling any cows.

Although we focused only on selling practice in this study, buying practice also needs to be investigated in future. A similar analysis to this study using the national-scale production database may allow us to study how farmers use production characteristics of cows to make a purchasing decision. Understanding purchasing practices is, however, more challenging because farmers seem to consider wider attributes of source farms such as their reputations and farm culture (Hidano et al., 2018b). This previous study also suggested that, at least among the current New Zealand dairy farm population, other actors such as stock agents and farm advisers have a significant influence on farmers' purchasing practices. The further important aspect is how these livestock selling and purchasing practices are, if any, associated with their farm biosecurity practices; this is critical to understand because it is the combination of these factors that play a role in disease spreads between farms. Farmers see those farms with poor data management as a 'shady farmer', who may purchase and sell cows that have problems such as disease and production disorders (Hidano et al., 2018b). Given that the national database does not capture these factors, it is necessary to conduct both quantitative studies and in-depth qualitative interview studies to fully understand farmers' livestock purchasing practices.

5.6. Conclusion

This study showed that farmers' selling practice are sensitive to both external (e.g. market fluctuation) and internal factors (e.g. farm management change), at least in the New Zealand dairy farm population. Given with a continuous increase in the herd size and decrease in the number of herds over years, as well as the instable market situation, livestock selling patterns will continue to change over time in New Zealand. Our preliminary results provide information that can be used to predict farmers' future livestock selling practice and/or livestock movements that are not recorded in the current database. Although the National Animal Identification and Tracing (NAIT) system has been established in New Zealand, the lack of accuracy in the data has been recently pointed out by the Agriculture and Biosecurity Minister (Anon, 2018b). At the time of writing this manuscript *Mycoplasma bovis* spread has been on-going in the country, and there has been of great interest to evaluate the effectiveness of different control options using a simulation model. Given that missing movement records can introduce a substantial bias on inferences from a disease simulation model (Dawson et al., 2015), any modelling exercises need to consider the dynamic change in the livestock movement patterns over time. The impact of violating assumptions about livestock movements in disease simulation models should be evaluated for diseases with different characteristics to more accurately assess the influence on model outcomes.

5.7. Acknowledgements

We are grateful to all farmers who participated in our previous interview studies for their general time and discussions, which significantly contributed to this study. We deeply thank Chris Compton and Cord Heuer from EpiCentre, Massey University for their invaluable inputs, knowledge, and suggestions for this study. This project was partially supported by the

Research Project for Improving Food Safety and Animal Health of the Ministry of Agriculture, Forestry and Fisheries of Japan, the IVABS travel funding from Massey University, and Kathleen Spragg Agricultural Research Trust. AH was supported by the JASSO scholarship, Japan. The funding sources had no involvement in the study design, the collection, analysis and interpretation of data, and the content of and the decision to submit the article for publication.

Chapter 6

Influence of population sampling strategies on making phylodynamic inferences for livestock diseases

Arata Hidano[1], M. Carolyn Gates[1]

[1] EpiCentre, School of Veterinary Science, Massey University,
Palmerston North, New Zealand

6.1. Abstract

As the cost of sequencing pathogen genomes has continued to decrease, many more research studies are using genetic data to make inferences about the epidemiology of important livestock pathogens. Isolates for these studies are often collected through contact tracing investigations or convenience sampling, which do not represent the whole infected population. Nevertheless, there is little knowledge about how phylodynamic inferences are influenced due to biased sampling when disease spreads through complex livestock population and contact structures. To fill this knowledge gap, we carried out a theoretical study to test the influence of different sampling strategies on estimates of the time to the most recent common ancestor (TMRCA) as an indicator of time elapsed since disease introduction. We used a novel individual-based disease simulation model that simulated the spread and mutation of a theoretical pathogen through New Zealand dairy contact networks during an outbreak. The resulting population of herds was then sampled using five different strategies: (1) random selection, (2) uniform probability selection based on time and geographical areas, (3) proportional to the number of detected herds in each geographical area, (4) proportional to the number of herds in each geographical area, and (5) all isolates. The pathogen isolates were analysed in BEAST v1.8.4 using the Extended Bayesian Skyline Coalescent model as a tree prior. The TMRCA was compared against the known time since introduction to estimate model accuracy. The results showed that there were no significant gains in accuracies of estimates when incorporating all potentially available samples compared with sampling only a subset of the population and within each sampling strategy, the model estimates varied greatly depending on various epidemiological characteristics of the epidemic including detection rate of infected cases, initial conditions of the epidemic, and contact structures of infected populations. Subsampling by stratifying based on geographical locations of samples needs a cautious consideration because geographical locations may not necessarily represent pathogen population structures when disease spreads through complex networks. These suggest that researchers may want to apply various subsampling strategies when making phylodynamic inferences.

6.2. Introduction

During infectious disease outbreaks, policy-makers need accurate epidemiological information such as the time of disease introduction, basic reproduction number, and transmission chains in order to make appropriate disease control decisions. These parameter estimates can be challenging to obtain from traditional epidemiological survey data collected during outbreak investigations, which are often biased by the sampling strategies and poor population coverage. To overcome this problem, researchers are increasingly using phylodynamic methods to make epidemiological inferences particularly for, but not limited to (Croucher and Didelot, 2015), fast-evolving pathogens such as RNA viruses (Pybus and Rambaut, 2009) based on genetic diversities among samples that are collected during the outbreak.

The term ‘phylodynamics’ was first coined by Grenfell and colleagues in 2004 to describe research studies that integrate immunodynamics, epidemiology, and evolutionary biology to better understand the link between epidemic processes and pathogen evolution (Grenfell et al., 2004). Within context of infectious disease epidemiology, phylodynamic studies can be roughly categorised into two groups based on their purposes, the scale of the study, and availability of data. One is to understand so-called ‘who infected whom’, which is to describe the transmission chain among infected individuals. When relatively rich epidemiological information is available, this approach has a potential to identify under-detected infections (Mollentze et al., 2014), risk factors associated with transmissions (Gardy et al., 2011), and the index case in the outbreak. The other group focuses on making population-level inferences based on phylogeny or genealogy such as the historical changes in the effective population size (Stadler et al., 2013), the time of the initial disease introduction (Gire et al., 2014), and geographical spread patterns (De Maio et al., 2015).

Many infectious disease studies that make population-level inferences use a Bayesian and population genetics framework, which offers a great advantage of accounting for uncertainties in evolutionary parameters and phylogenies—it is well known that many different phylogenetic tree topologies can equally explain the observed genetic data and hence it is difficult to determine a single topology based on data at hand (Lemey et al., 2009). The development of user-friendly software such as BEAST (Drummond and Rambaut, 2007), BEAST2 (Bouckaert et al., 2014b), and MrBayes (Ronquist et al., 2012) has also provided researchers with an access to sophisticated methods (Nascimento et al., 2017). With this background, Bayesian evolutionary analysis has been frequently applied to many of recent important human disease outbreaks. For instance, Scarpino and colleagues analysed the genetic sequence data from the 2014 Sierra Leone Ebola outbreak using both the Bayesian evolutionary analysis based on BEAST and transmission chain analysis (Scarpino et al., 2015). The authors reported the estimated starting date of the outbreak were similar in both analyses. Boskova et al., (2018) conducted a comprehensive analysis on genetic sequence data from ZIKA virus outbreak in Brazil and USA using Bayesian and non-Bayesian methods (Boskova et al., 2018). In particular, the authors performed a parallel analysis using two different tree priors—the coalescent skyline plot (Drummond et al., 2005) and birth-death skyline model (Stadler et al., 2013)—identifying that phylodynamic inferences including the time to the most recent common ancestor (TMRCA) may be unreliable if genetic sequence data contain few variations.

Specifying a tree prior is an essential part of Bayesian evolutionary analysis because the parameter space for possible tree topologies is simply too large. The use of population genetic models as a prior also enables researchers to quantify the relationship between the demographics of genetic population and genetic diversities observed in the population. In this context, population refers to pathogen population. and not the host population. The historical fluctuation of genetic population size, so called effective population size (ESS), has been often of epidemiological interest because it may infer how disease prevalence has changed over time in the past; note that, however, a study has shown that ESS does not equate to disease prevalence (Volz, 2012).

Both of the commonly used population genetic models (coalescent models and birth-death models) make several key assumptions about the underlying host population. First, the models assume that the population is panmictic, meaning that individuals have a homogenous mixing contact structure. This assumption, however, is almost always violated in reality and several studies have shown that this violation can lead to an erroneous inference such as false bottleneck signals

(Chikhi et al., 2010) and false decline in ESS towards the present (Heller et al., 2013). Based on a simulation exercise of pathogen sequence mutations in six structured populations between which pathogens migrate, Hall and colleagues showed that abovementioned biases do also occur in viral phylodynamic studies when the homogeneous assumption is violated (Hall et al., 2016). This assumption can be relaxed by using a structured coalescent (Vaughan et al., 2014) or a structured birth-death model (Kühnert et al., 2016), both of which require a specification of population structures (e.g. which sequence belongs to which population) and migration rates between populations. In the context of livestock diseases, this presents additional challenges related to the fact that is often unclear what the 'population' represents since animals are typically segregated into spatially isolated herds and the migrations between herds are typically sporadic and highly seasonal. Geographical areas may be a good proxy for a population in the context of species evolution or some human diseases; however, many livestock diseases transmit between farms via livestock movements, which do not necessarily correlate with geographical boundaries (Hidano et al., 2016). A network community structure based on livestock movements may be a better proxy, however, there are no studies that have investigated this so far.

Another key assumption made by these models is related to sampling of sequences. Coalescent models assume that genetic samples analysed are a small fraction of random samples that were collected from a large population. In contrast, a conventional birth-death model assumes that a whole population is sampled, but this assumption can be relaxed to account for incomplete sampling (Stadler, 2009). For example, Stadler and colleagues developed the birth-death skyline plot in which sampling proportions can vary over time (Stadler et al., 2013). Regardless, all these models still assume that samples were selected randomly from the background population of sequences during a given time period. For infectious disease studies, this assumption is rarely met because of imperfect test accuracy, incomplete epidemiological tracing due to imperfect movement records and problems with recall bias, and biased surveillance intensity.

Only a handful studies to date have investigated the performance of different sampling strategies on making phylodynamic inferences. A seminal study by Stack and colleagues identified that the temporal distribution of samples can have a substantial impact on estimating the past dynamics of the effective population size (Stack et al., 2010). Another study showed that an over-sampling of epidemiologically linked sequences leads to a significantly biased estimate (Silva et al., 2012). Karcher et al., (2016) also showed that effective population size estimates may be biased if sampling times are dependent on the effective population size itself—for instance if more samples are collected during the period in which more individuals are infected (Karcher et al., 2016). All these studies were, however, conducted assuming that the population is non-structured and panmictic. A more recent study, however, compared the performance of three subsampling strategies both in non-structured and simply structured populations (Hall et al., 2016). This study suggested that sampling sequences with uniform probability with respect to both time and population structure is likely to perform equal or better than sampling proportional to the effective population size at a given time in a given population.

These studies indicate that sampling strategies are likely to influence phylodynamic inferences for livestock diseases; livestock populations and their contact patterns are highly heterogeneous and structured. Bayesian evolutionary analysis has been already applied to various livestock diseases such as Foot-and-mouth disease (Nishi et al., 2017), Avian Influenza (Fourment and Holmes, 2015), and other RNA viruses. Sampling protocols of sequence data are, however, not sufficiently considered in the current literature. This is not surprising given that many conventional phylogenetic studies have been performed based on genetic data collected through a convenient sampling with little consideration of epidemiological representativeness (Foxman and Riley, 2001). This is somewhat inevitable given that genetic sequences are often obtained from samples that are submitted to laboratories, collected by surveillance, and contact tracing (Field et al., 2014). For instance, a Canadian study analysed 26 genetic sequence data of bovine viral diarrhoea virus that were obtained from a laboratory. Using epidemiological information attached to samples such as their estimated isolate date and locations, this study estimated various parameters including TMRCA, historic ESS, and geographical spread patterns (Chernick et al., 2014). The authors, however, noted that the model inference for the geographical spread was inconsistent to the existing knowledge and argued the need of more thorough sampling. Alkhamis and colleagues analysed 285 sequences of porcine reproductive and respiratory syndrome virus, which had been collected in a convenience sampling manner (Alkhamis et al., 2016). The authors acknowledged that this sampling may have biased their inferences, while noting that they included all available sequence data hence their results be the best representation of the available knowledge.

Despite the growing recognition of biases caused by subsampling techniques, there are still no established guidelines for phylodynamic studies that make inference about livestock diseases that spread through complex livestock population and contact structures. The advent of high-throughput technology to obtain genetic information about infectious disease agents in a timely and inexpensive manner has made a substantial amount of pathogen sequence data readily available. This will undoubtedly lead to an increase of the number of phylodynamic applications to livestock diseases. The importance of performing a subsampling will also increase given that available genetic samples are always collected in a somewhat biased manner. The objective of this study is therefore to compare different subsampling strategies of genetic sequences for their accuracy and precision of inferences. We conducted a large-scale simulation study using an individual-based disease model to simulate the spread and genetic mutation of a fast mutating virus through a typical cattle movement network. Genetic samples were obtained under five subsampling strategies, which were analysed using BEAST to infer the TMRCA. Further analyses were also conducted to determine whether the epidemiological characteristics of the outbreak influenced the accuracy and precision of inferences.

6.3. Materials and methods

6.3.1. An overview of study

To explore the impact of different inclusion strategies for genetic sequences on phylodynamic inferences in a complex livestock disease transmission system, we generated a “true” dataset where all epidemiological and evolutional parameters are known. This dataset was obtained by developing a novel spatially explicit disease simulation model in an individual-based model framework. The model has three components: (1) a demographic component describing various demographic events including between-herd livestock movement, births, deaths, and culling of animals, (2) a disease transmission component describing pathogen transmission between individuals within a herd and between herds, and (3) a genetic mutation component describing nucleotide substitutions over time.

6.3.2. Demographic component

The New Zealand dairy farm network was used as a sample population for the simulation framework. Data used to generate this network were extracted from the New Zealand Dairy Industry Good Database (DIGAD), which contains demographic, movement, and production records for approximately 70% of commercial dairy herds in New Zealand over the past thirty years (Anon, 2001). For the purpose of this study, we used records from 1st July 2000 to 31st June 2010. Herds in a given year were defined as a group of animals that were present between 1st July and 31st June, which is a typical production season in New Zealand (Hidano and Gates, 2018). The contact networks were based on the actual observed patterns of movements during this time period. As of 1st July 2000, there were 3,624,420 individual animals present in the database and over the study period, there were records for a total of 16,534,951 individual animals across 45,965 unique herds. These included 9,667,100 parturition records, 6,165,850 culling records, and 8,841,850 between-herd movements of individual animals. In 18,717 occasions, missing data for between-herd animal movements were suspected because the destination herd of a movement does not match to the source herd of an immediate next movement for the same animal. On these occasions, we simply added a movement from the destination herd to the source herd, sampling a random date between these two movements, assuming there is only one missing record in each occasion.

To replicate a realistic cattle population structure in New Zealand, we modelled the population based on a metapopulation scheme with two layers: within-herd and between-herd dynamics. Within-herd dynamic concerns individual animals within a herd. In this study, we grouped individuals into three age groups within a herd; calves that are <=12 months, heifers <12 and <=24 months, and adults >24 months. We also assumed that each age group does not mix with other groups and hence there is no disease transmission between different age groups except in the following situation. Calves and heifers move to a heifer and an adult age group, respectively, within the same herd on 10th July every year, which can spread disease if moved animals are infected. The last assumption is reasonable given that majority of dairy herds in New Zealand run a spring calving system, where calving usually occurs between June and September each year.

Each herd is connected via livestock movements. When animals move between herds via a livestock movement, they join the corresponding age group; for instance, adult animals from one farm only move to the adult age group of the destination farm. Locations of herd were divided into 16 geographical regions as previously described (Hidano et al., 2016).

We also explicitly considered other demographic events of animals—birth, death and culling. Birth events are those new calves join the calf group of a given herd. On a death or culling, animals were removed from a given herd. These demographic events (birth, death, culling, and between-herd movement) were modelled as a non-Markov process using the actual events recorded in the DIGAD database.

6.3.3. Disease transmission component

In this study, we modelled a highly contagious, rapidly mutating virus (such as Foot-and-mouth disease virus; FMD) that provides a relatively large number of genetic mutations in a limited time duration. In the simulation, pathogen spreads between individuals within an age group via a direct contact and between herds via a movement of infected individuals. For simplicity, we assumed that pathogen spreads only between cattle farms through cattle movements because this study was not intended to make inferences about pathogen transmission, which also occurs through local spread and involves other ungulate species (Green et al., 2006).

Individuals were classified into one of the following four mutually-exclusive disease statuses: susceptible (S), exposed (E), infectious (I), and immune (R). Where there is at least one infectious animal, each individual in the same age group is assumed to receive a force of infection λ, which is described as follow:

$$\lambda_{ut} = Ce \times I_{ut} \div N_{ut}$$

where Ce, I_{ut}, and N_{ut} represent the number of effective contacts made by each individual per unit time, the number of infected individuals, and the number of total individuals at a given time t in a given animal group u (i.e. a given age group in a given herd). Individuals in E status subsequently move to I status after a latent period and move from I status to R status after an infectious period. Each newly infected animal acquires the same genetic sequence as the one that is randomly selected from all genetic sequences carried by infected animals in the same age group in a given herd at the time when the new infection happened.

In the disease transmission component, disease transmission within a herd was modelled as a continuous-time Markov process. On the other hand, transitions of individual animal disease status (i.e. transition from E to I and transition from I to R) were modelled as a non-Markov event. This is to avoid an assumption that a probability that animals transit to a next disease status is independent of the period the animal had spent in the current disease status, which is unlikely to hold for diseases we use in this study. Values for latent and infectious period were randomly sampled from distributions described in Table 6.1.

Table 6-1 Parameter values used in the disease simulation.

Parameter	**Distribution**
Ce (no. effective contacts)	5 / day (fixed)
Latent period (from E to I)	Uniform (0, 100) days
Infectious period (from I to R)	Uniform (60, 300) days
Substitution rate	0.012 /site/year (fixed)
Baseline probability of detection (P_{base})	0.01 / day (fixed)
Elevated probability of detection (P_{inc})	0.02 / day (fixed)

6.3.4. Genetic mutation component

As previously explained, each infected animal carries one genetic sequence. For the sequence, we used the VP1 region of FMD virus with a length of 633 nucleotides. Once an animal acquires a sequence, a nucleotide substitution starts occurring over time. This mutation was assumed to occur following Jukes-Cantor model with a substitution rate of 0.012 per site per year (Di Nardo et al., 2014) in individuals that have either status E or I. The mutation was modelled as a continuous-time Markov process. We assumed there is no recombination between sequences and individuals can be infected by only one virus. As a genetic sequence for the initial case, the sequence of Accession number FJ785304_1 was used.

6.3.5. Disease detection and control component

Disease detection was assumed to occur only via clinical detection when the animal entered the infectious (I) period and not by other means such as a slaughterhouse inspection. Each herd had a probability to be detected positive as soon as there was at least one animal in infectious (I) status present in either of three age groups. When there are no farms that are detected positive, we assumed that all infected herds have the same baseline probability to be detected (P_{base}). To mimic the real situation of disease outbreaks in which resources are concentrated into the areas where detected herds exist (and hence preferential sampling occurs), we assumed that infected herds in the geographical regions which have detected herds have an elevated probability of being detected (P_{inc}). Once a herd is detected positive, several events occur. First is the genetic sampling. One animal (hence one genetic sequence) was randomly sampled from all animals that were in infectious (I) status on the herd, only if this herd was never detected positive before. The identification number of a sampled animal, the herd identification number of the herd this animal existed, and the time of detection were recorded. Second is a placement of a movement restriction; the detected herd is banned for moving off any animals for 180 days. We did not consider any additional epidemiological investigations triggered by a detection such as a backward- and forward-contact tracing. After 180 days of a movement restriction, all animals in either E or I status were assumed to move into R status. Values used for each parameter can be found in Table 6-1.

6.3.6. Simulation algorithm and conditions

We used the modified Gillespie algorithm that has previously been implemented by other studies (Conlan et al., 2015; Widgren et al., 2016); whereas non-Markov events occurred at a given scheduled day, Markov events could occur at any given continuous time. As already explained, there are three Markov events in this simulation; disease transmission, disease detection, and genetic mutation. For Markov events, the distribution of inter-event time τ (i.e. time to a next Markov event: t_{MARKOV}) can be described as $q(\tau) = 1 - e^{-\mathrm{B}t}$, where B_t represents the sum of the rate of all Markov events at a given time t. By randomly sampling a value from this distribution, we can calculate the t_{MARKOV} at the time t. On the other hand, for non-Markov events, the time at which these events occur are pre-defined, therefore, we also know the time to a next non-Markov event, $t_{NON\text{-}MARKOV}$. We calculated t_{MARKOV} every time when any Markov and non-Markov events occurred. Calculated t_{MARKOV} was then compared to $t_{NON\text{-}MARKOV}$, and if $t_{NON\text{-}MARKOV} < t_{MARKOV}$ the non-Markov event was carried out and we again calculated the time for the next Markov event: otherwise, the Markov event was implemented and the simulation day was updated.

During the preliminary simulations, we often observed a disease spread stopped before a sufficient number of genetic sequences for the phylodynamic analysis to be collected. It was therefore necessary to adjust the initial conditions to ensure the outbreaks were of sufficient duration and magnitude to permit the different sampling strategies to be evaluated. First, as the infection seed on Day 0 (which is 1st July 2000), we chose one herd randomly from farms that moved at least one adult animal off to at least seven other farms within the first year. This is to ensure that disease spreads to a large enough number of farms and hence a sufficient number of genetic sequences could be collected. There were 109 farms with 12,213 adult animals that met this condition. At the chosen herd, one animal from each of three age groups was randomly chosen to be infected. Their disease statuses were set to I and allocated the genetic sequence that was described in Section 2.7. Furthermore, we assumed that the first disease detection would not occur by Day 365, mimicking the situation where a disease spreads silently. This may or may not be the expected scenario in New Zealand, however, again we emphasise that the objective of this study was not to replicate the realistic FMD outbreak in New Zealand: it is to evaluate which sample inclusion strategies are optimal for a disease that spreads over a complex livestock movement network.

At each iteration, we recorded following information of all infected animals; animal identification number, genetic sequence identification number, identification number of sequence that infected this animal, infection date, identification number of farm at which this animal was infected, whether or not this sequence was isolated, and nucleotide sequence if isolated. Isolated sequences were then subsampled according to sampling strategies described below and exported as a NEXUS file, which is the required format for BEAST analysis. The simulation model is coded in the C language and simulation codes are rigorously checked for its validity. The simulation code can be found at the first author's repository (https://github.com/arata-hidano/Molecular_simulation/blob/master/Disease_molecular_simulation_TB2.c). In each iteration, a simulation was run for 10 years, or until the disease dies out, or a total of 150 genetic sequences were obtained. Results of iterations were only analysed when at least 50 genetic sequences were obtained—this is to avoid a sample size being too small after the subsampling process (described in Section 4) for the Phylodynamic analysis. A simulation was performed until a total of 100 eligible iterations were obtained.

6.3.7. Sample inclusion strategy

At the end of each iteration, there was a list of genetic sequences that were collected through a disease spread as described in Section 2.5. These sequences were further subsampled based on different sampling strategies. These strategies include: (1) uniform probability sampling with respect to both time and geographical areas (hereafter referred to as 'equal' strategy), (2) random sampling ('random' strategy) that does not account for time and geographical areas of samples, (3) random sampling proportional to the number of infected herds in each geographical area ('herd infected' strategy), (4) random sampling proportional to the number of herds in each geographical area ('herd' strategy) so that consistent proportion of herds are collected across regions, and (5) no selection and use all available sequences ('full' strategy). In 'equal' strategy, samples were stratified based on the time and locations and one sample was included from each stratum as follows. First, it is necessary to divide the sampling period into intervals. This sampling period (T_{samp}) corresponds to the period between the earliest and last sampling time. To determine how many sampling intervals (n_i) we create, the number of geographical regions (n_r) that had at least one sequence-isolated herd was obtained. Based on the total number of genetic sequences collected (n_t), n_i was determined as $n_i = n_t/n_r + 1$. This was to maximise the number of samples included in each geographical area while ensuring the probability of selecting a sample in each region is the same. Then, the width of the sampling interval can be described as T_{samp}/n_i. We selected one sample per interval per geographical area, which resulted in the sample size of k. For strategies 'random', 'herd infected', and 'herd', we collected k samples to compare the effectiveness of each strategy given a sample size. For strategy 'random', k samples were subsampled without replacement from all collected sequences without any regards to time or geographical regions. For strategy 'herd', samples were selected from each geographical region proportional to the number of herds in each region. As the number of herds, we excluded herds that were not involved in any livestock movements (inactive herds) because they could not be infected in this simulation framework. Given the majority of eligible iterations stopped before Day 1095 (i.e. 3 years from the start), we defined inactive herds as those had no animals that moved in or moved off from the herd between Day 0 and 1095. For 'herd infected' strategy, samples were selected from each geographical region proportional to the number of farms that were detected positive in each region. Note that the number of farms detected positive is not necessarily the number of farms infected in each region, but it is often the information available in the real disease outbreak situation.

6.3.8. Phylodynamic analysis

For each sampling strategy, the obtained genetic sequences were analysed using BEAST version 1.8.4 to estimate the time to the most recent common ancestor (TMRCA). In our study, TMRCA approximates to the time between Day 0 and the last sampling date (T_{last}) because we used a single sequence as a seed; at the start (Day 0) there were not multiple sequences that had a common ancestor before Day 0. We used an Extended Bayesian Skyline Coalescent model as a tree prior (Heled and Drummond, 2008). The date that each sequence was obtained was used as a tip date. The Jukes-Cantor model was used as substitution model assuming a strict molecular clock, and a prior for the substitution rate was set to have a uniform distribution of 0.00002 and 0.00005 per site per day as a lower and upper limit, respectively. We used this narrow prior to identify the mere influence of sampling strategies on phylodynamic inferences rather than accounting for an uncertainty around the substitution rate. Other parameters were set to the BEAST default. Markov chain Monte Carlo (MCMC) chain was run over 50,000,000 iterations, sampling at every 1,000 iterations, and the first 10% was discarded as a burn-in period. Results obtained by BEAST were summarised using Tracer version 1.7 (Rambaut et al., 2018). The effective sample size (ESS) for each parameter was ensured to be larger than 200; otherwise, MCMC chain was run longer.

6.3.9. Evaluation of accuracy and precision of phylodynamic inferences

The median estimated value for the tree height (i.e. TMRCA) was compared to the corresponding T_{last} in each iteration. Following the previous study (Hall et al., 2016), we calculated following statistics for each sampling strategies:

$$\text{Percent bias} = 100 \times \frac{Tlast - estimated\ TMRCA}{Tlast}$$

which represents in which direction and how much the estimate is biased. The negative of this statistic indicates that the TMRCA is overestimated and vice-versa.

$$\text{Percent error} = 100 \times \frac{|Tlast - estimated\ TMRCA|}{Tlast}$$

which represents the divergence of the estimated median value from the true value.

We also obtained the size of highest posterior density (HPD) as an indicator of the precision of the estimate,

$$\text{HPD size} = 100 \times \frac{TMRCA_{97.5} - TMRCA_{2.5}}{Tlast}$$

where $\text{TMRCA}_{97.5}$ and $\text{TMRCA}_{2.5}$ represent the upper and lower limit of its 95% HPD. Above three statistics were obtained for each of 100 eligible iterations, which were then used to estimate kernel density distributions, as was done by the previous study (Hall et al., 2016). The bandwidth was chosen using the method described by (Sheather and Jones, 1991). We then calculated the coefficient of overlap using the OVL_4 estimator proposed by (Schmid and Schmidt, 2006), which represents the agreement between two distributions. The coefficient takes value between 0 and 1, where 0 indicates that two distributions are completely distinct and 1 indicates that the two are identical. The coefficients and their 95% bootstrap confidence interval were obtained using the R package overlap (Ridout, 2018). A post-hoc non-parametric all-pairs Kruskal rank comparison test (Nemenyi test) was performed for each of above three statistics to test the null hypothesis whether all distributions were drawn from the same distribution, using the R package (Pohlert, 2018). In addition, we calculated two additional statistics: coverage and convergence rate. The former represents the proportion of iterations (out of 100 iterations) in which the HPD contained the true T_{last}. The latter represents the proportion of iterations in which 200 ESS was obtained for parameters of interest within 50,000,000 MCMC iterations as a crude measure of the efficiency of mixing.

6.3.10. Associations between phylodynamic inferences and disease outbreak characteristics

Associations between the performance of the BEAST inferences in each strategy and simulation characteristics were investigated. For this, a generalised additive model was fitted using the R package gam (Hastie, 2018). The seven simulation characteristics analysed were T_{last}, the number of total genetic samples collected, the proportion of genetic samples that were subsampled, the number of animals that were infected, the proportion of infected farms that were sampled out of the total number of infected farms, the proportion of maximum number of samples that were collected from single geographical region, and the proportion of maximum number of samples that were collected from single network community structure. A multivariable model was then built by keeping variables that showed p-value < 0.05 by ANOVA test. A community structure is a common feature in many network, where nodes (in our case, farms) are tightly connected (Girvan and Newman, 2002). We identified community memberships of farms for each of the 10 years between 2000 and 2009, aggregating all livestock movements in one year into one static network. This means the community membership of farms can vary each year. Each genetic sample was allocated the membership of the farm in a given year from which the sample was isolated. We used the greedy algorithm (Clauset et al., 2004) to identify a community membership for each farm using the R package igraph (*igraph*, 2018). The desired number of communities were set to 100. All the statistical analyses were performed using R version 3.4.2.

6.4. Results

6.4.1. Disease simulation results

During the 10-year period, there were 5,487,015 distinct animals had moved between farms, consisting a total of 8,860,567 moves. Among these, 1,480,616 (16.7%) were calves, 2,000,401 (22.6%) were heifers, and 5,379,550 (60.7%) were adult animals. There were 29,223 distinct source farms and 25,613 destination farms. Minimum, median, and maximum number of geographical regions that disease spread were 4, 9, and 13, respectively. Minimum, median, and maximum number of communities that disease spread were 1, 2, and 5, respectively. There was no significant association between the number of regions disease spread and the number of communities disease spread (Pearson coefficient = 0.14, p = 0.13).

Figure 6-1 shows the distributions of seven statistics. The median value for the last day of sampling (which was equivalent to the time to the most recent common ancestor (TMRCA) in this study) was 904 days (minimum 500 days and maximum 1758 days). The median numbers of infected animals and farms were 23,925 and 78, respectively. The median number of genetic sequences sampled was 72.5; these numbers indicate the sample size used for the BEAST analysis in 'full' strategies.

In other strategies, we subsampled the same number of sequences for each strategy; the sample size ranged between 16 and 79, with the median being 26. Figure 6-1 (F) shows the maximum sample size that was obtained from one geographical region. The proportion of maximum number of samples collected from one geographical region ranged between 30.2% and 90.6%. The proportion of maximum number of samples collected from one community structure ranged between 75.3% and 100%.

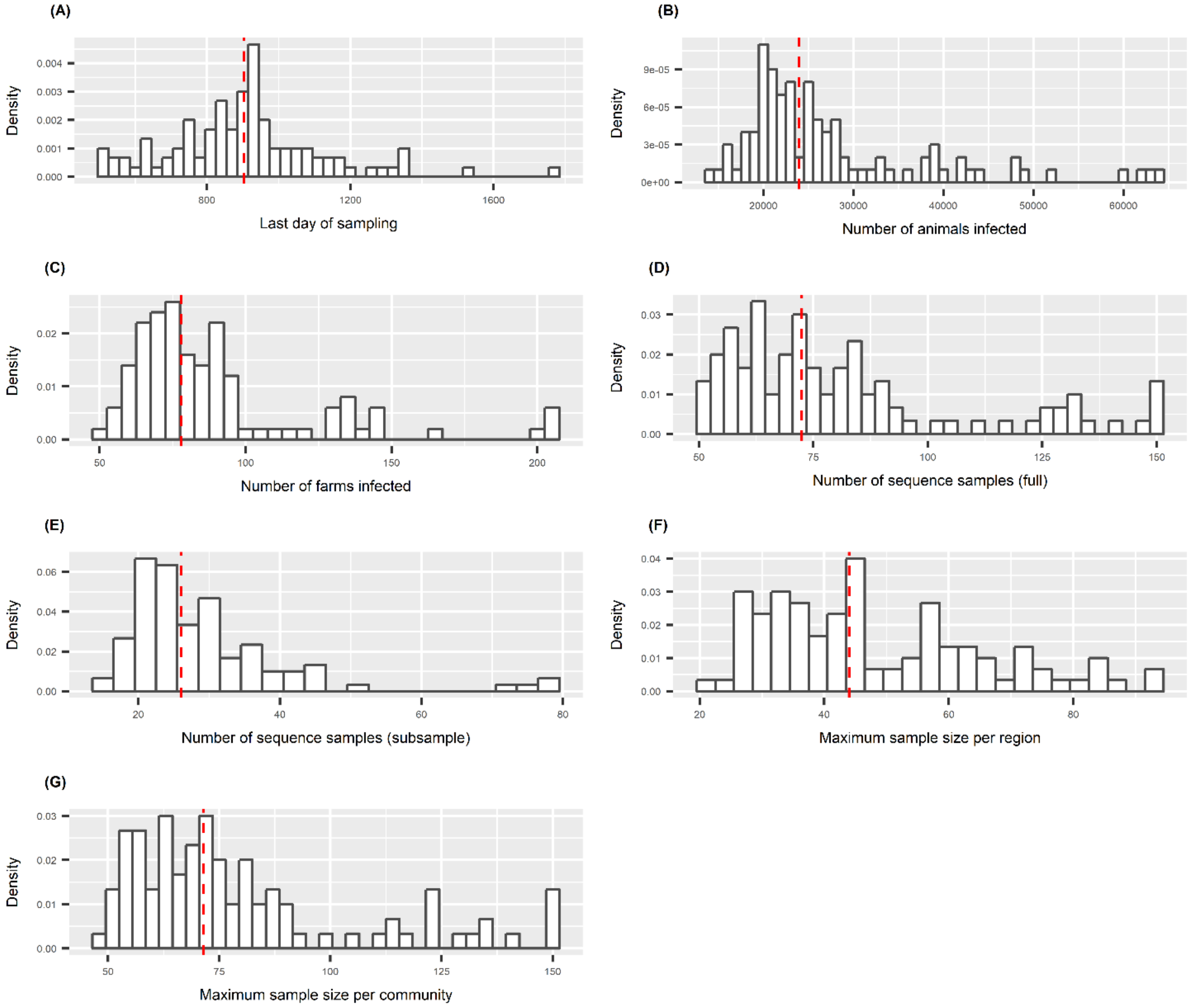


Figure 6-1 Descriptive statistics of results obtained from individual-based disease simulation models over 100 iterations.

The red dashed lines represent the median values for each statistic.

6.4.2. Comparisons of accuracy and precision of TMRCA estimations between sampling strategies

As shown in Figure 6-2 and Table 6-2, distributions of percent bias were similar between 'equal' and other four strategies. None of post-hoc significant tests showed evidence that percent bias values from each strategy were drawn from different distributions.

Table 6-2 The coefficients of overlap and their bootstrap 95% confidence interval for the kernel density estimates of percent bias between each of five sampling strategies.

	Equal	Random	Herd infected	Herd
Random	0.90 (0.81, 0.96)			
Herd infected	0.85 (0.75, 0.93)	0.91 (0.84, 0.96)		
Herd	0.87 (0.77, 0.94)	0.88 (0.79, 0.94)	0.87 (0.77, 0.94)	
Full	0.89 (0.80, 0.95)	0.86 (0.76, 0.94)	0.87 (0.77, 0.93)	0.89 (0.81, 0.94)

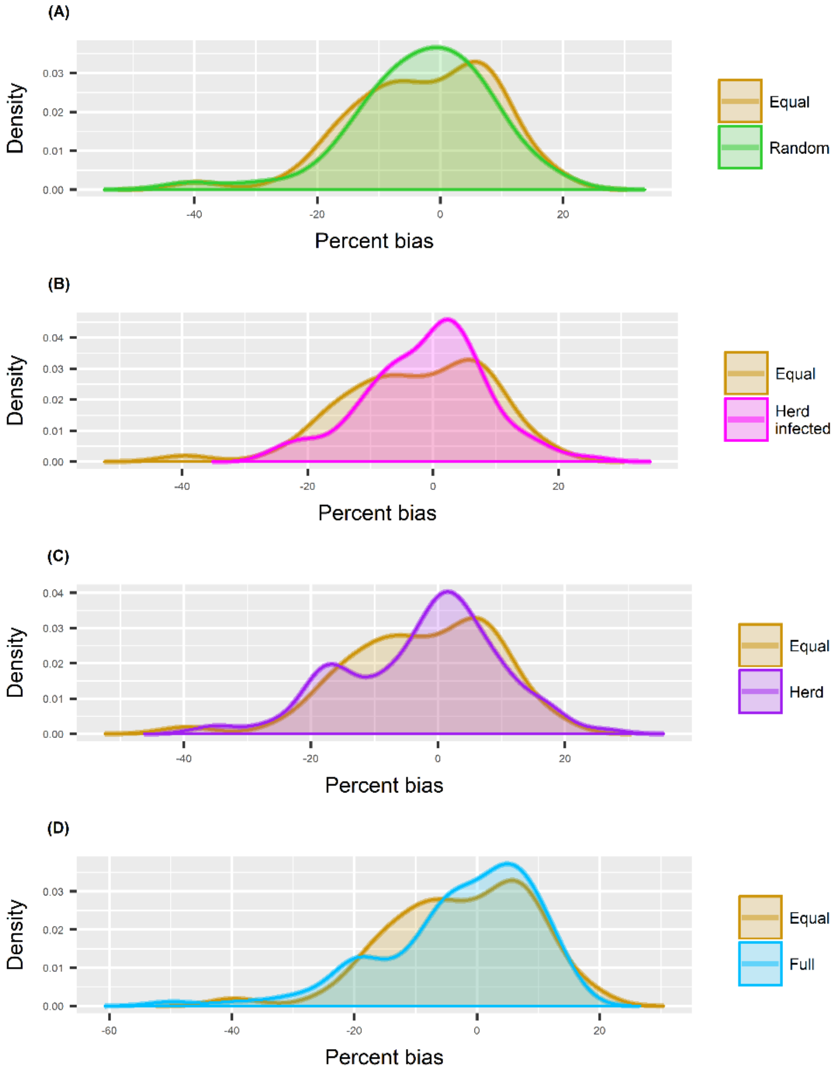


Figure 6-2 Kernel density estimates for percent bias in TMRCA estimates from five sampling strategies.

Similarly, Figure 6-3 and Table 6-3 show that there was a substantial overlap in distributions of percent error between each of five sampling strategies. There was, however, marginal evidence that percent bias values from ‘equal’ and ‘herd infected’ strategies were drawn from different distributions (p =0.12).

Table 6-3 The coefficients of overlap and their bootstrap 95% confidence interval for the kernel density estimates of percent error between each of five sampling strategies.

	Equal	Random	Herd infected	Herd
Random	0.89 (0.80, 0.95)			
Herd infected	0.82 (0.70, 0.91)	0.89 (0.81, 0.95)		
Herd	0.81 (0.69, 0.90)	0.84 (0.74, 0.92)	0.85 (0.76, 0.92)	
Full	0.88 (0.79, 0.94)	0.93 (0.87, 0.96)	0.91 (0.83, 0.95)	0.84 (0.73, 0.92)

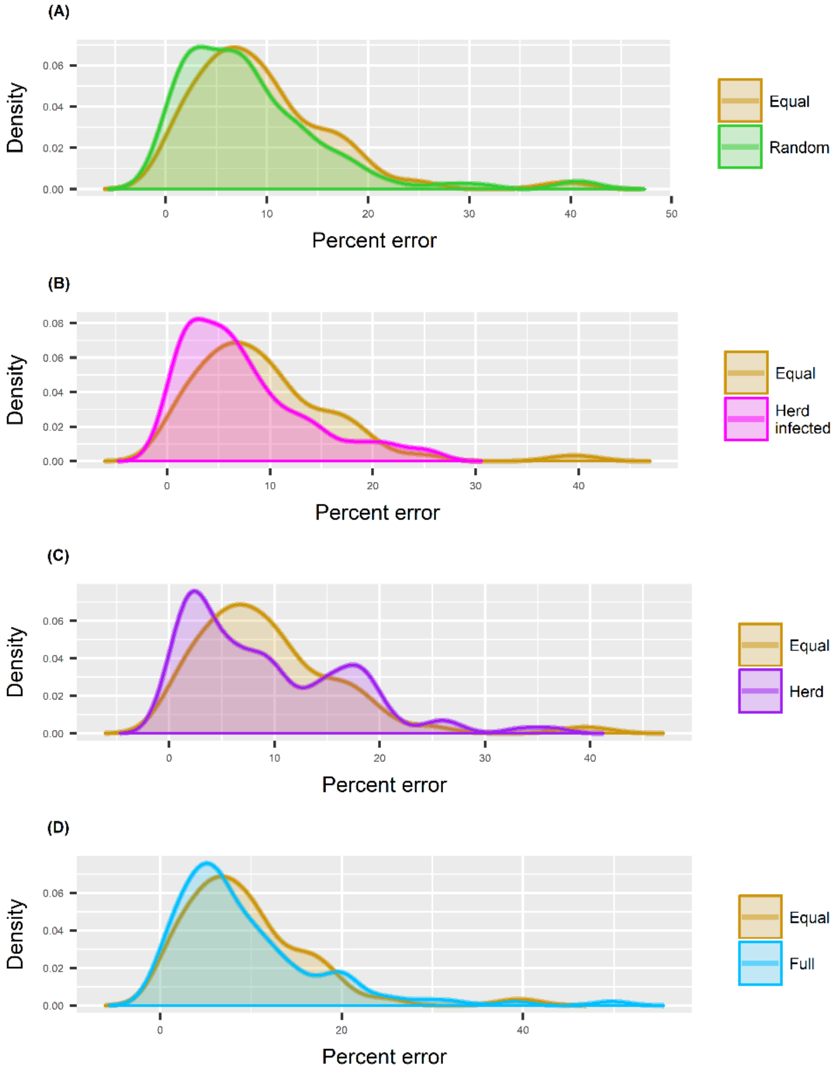


Figure 6-3 Kernel density estimates for percent error in TMRCA estimates from five sampling strategies.

Figure 6-4 represents the distributions of HPD size for ‘equal’ and other four strategies. Table 6-4 displays the coefficients for overlap between given two distributions. Distributions of ‘equal’, ‘random’, ‘herd infected’, and ‘herd’ were all similar. There was less overlap between following pairs: ‘equal’ and ‘full’, ‘random’ and ‘full’, ‘herd infected’ and ‘full’, and ‘herd’ and ‘full’. For each of these pair, post-hoc tests showed a strong evidence that HPD size values used to estimate each distribution were drawn from different distributions (p <0.001, respectively). Again, there was marginal evidence that HPD size values from ‘equal’ and ‘herd’ infected’ strategies were drawn from different distributions (p = 0.1).

Table 6-4 The coefficients of overlap and their bootstrap 95% confidence interval for the kernel density estimates of 95% HPD size between each of five sampling strategies.

	Equal	Random	Herd infected	Herd
Random	0.86 (0.75, 0.93)			
Herd infected	0.84 (0.73, 0.93)	0.92 (0.86, 0.97)		
Herd	0.88 (0.79, 0.94)	0.89 (0.79, 0.95)	0.93 (0.87, 0.97)	
Full	0.74 (0.62, 0.85)	0.59 (0.48, 0.71)	0.58 (0.47, 0.70)	0.63 (0.52, 0.75)

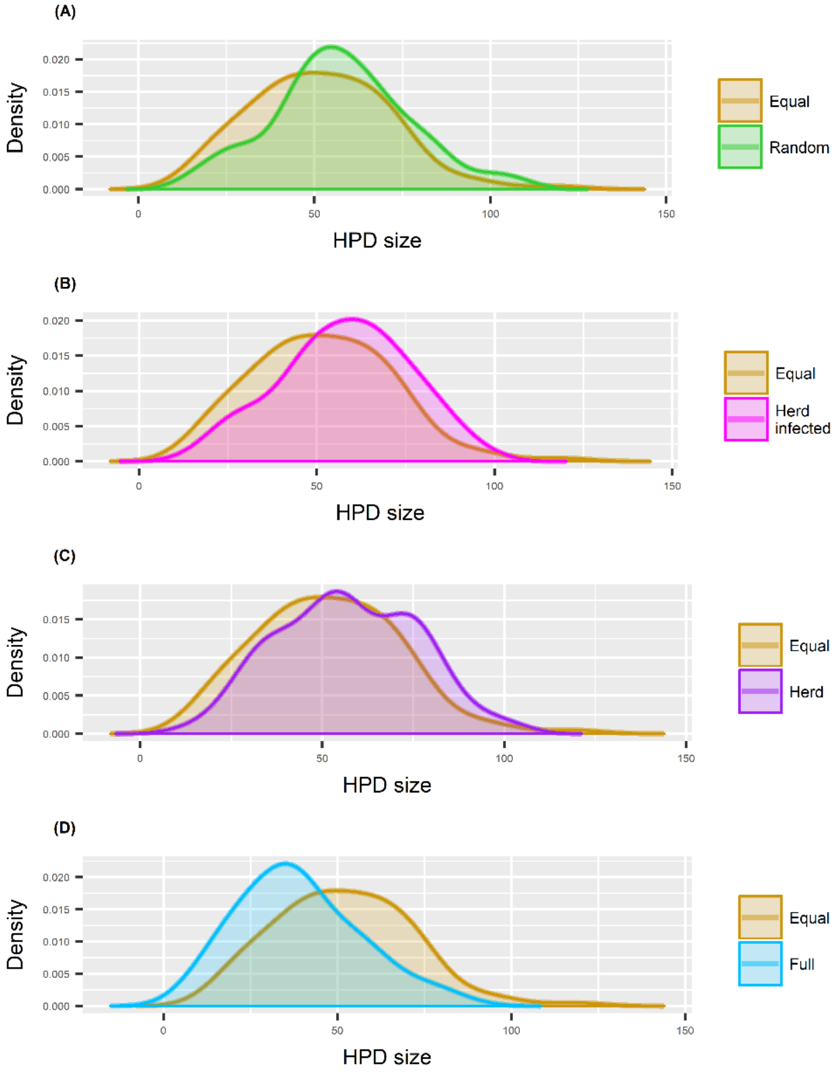


Figure 6-4 Kernel density estimates for HPD size in TMRCA estimates from five sample selection strategies.

Table 6-5 displays the results of coverage and convergence statistics. HPDs from all but ‘full’ strategies contained the true value almost all the time. In more than 90%, MCMC converged within 50,000,000 iterations for all but ‘full’ strategy.

Table 6-5 Coverage and convergence statistics for each sampling strategy.

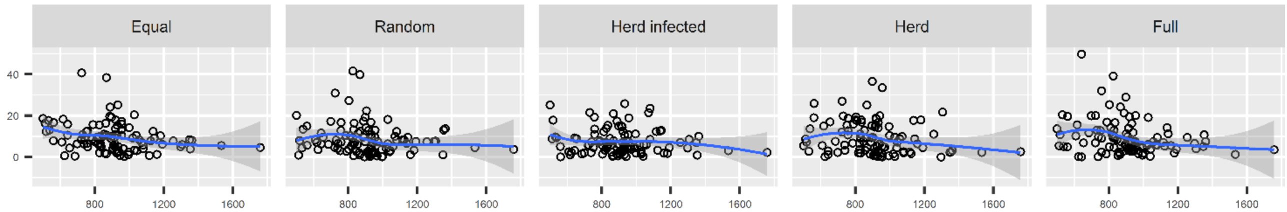
Coverage indicates the proportion of 95% HPD that contained the true value of the last sampling time out of 100 iterations. Convergence metric is a rudimentary measure to indicate how many iterations failed to converge within 50,000,000 MCMC iterations.

Strategy	Coverage (%)	Convergence (%)
Equal	98	93
Random	99	92
Herd infected	99	95
Herd	99	93
Full	91	59

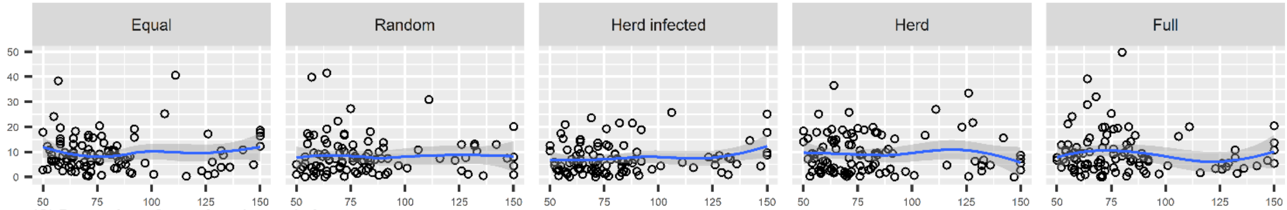
6.4.3. Association between phylodynamic inferences and simulation characteristics

Figure 6-5 shows the association between percent error and each of seven simulation characteristics. There was a significant negative association between percent error and the last sampling day in ‘equal’, ‘herd’, and ‘full’ sampling strategy. This association in ‘equal’ strategy, however, became insignificant after adjusting by other factors. There were no significant associations between percent error and the number of genetic sequences collected. The proportion of genetic sequences subsampled was associated with percent error in ‘herd infected’, which was the only factor remained significant in the final model for ‘herd infected’ strategy (Table S1 in Appendix 3). The number of animals infected was associated with percent error in ‘equal’ and ‘full’ strategies, but these became insignificant after adjusting by other factors. For only ‘equal’ strategy, there was a negative significant association between percent error and the proportion of infected farms sampled, which remained significant in the final model (Table S1 in Appendix 3). There was no significant association between percent error and the proportion of maximum sample size collected from one region. Positive associations between percent error and the proportion of maximum sample size collected from one community were found for all but ‘equal’ strategy; all of which, however, became insignificant after adjusting by other factors.

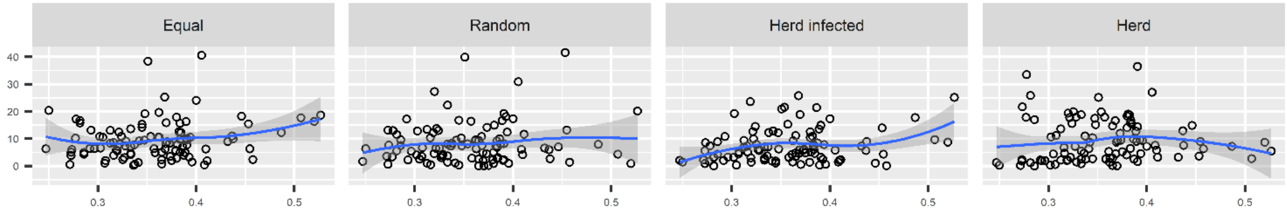
(A) Last sampling day



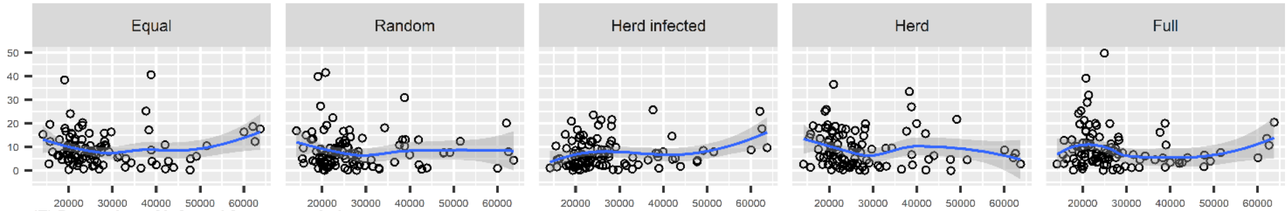
(B) Full sample size



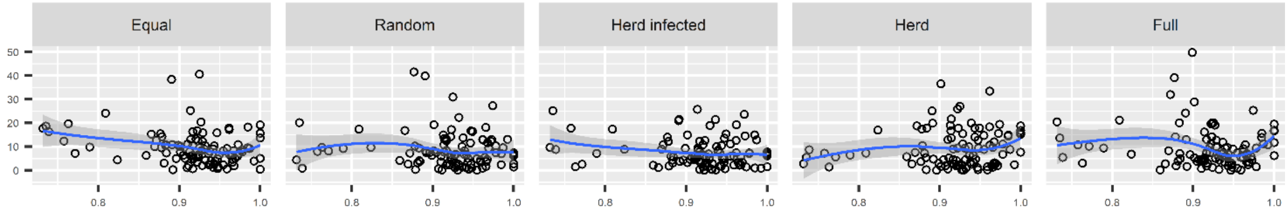
(C) Proportion of sample subsampled



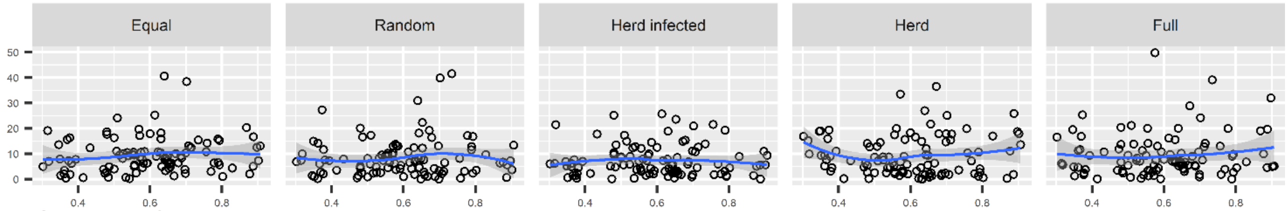
(D) Number of animals infected



(E) Proportion of infected farms sampled



(F) Proportion of max sample size per region



(G) Proportion of max sample size per community

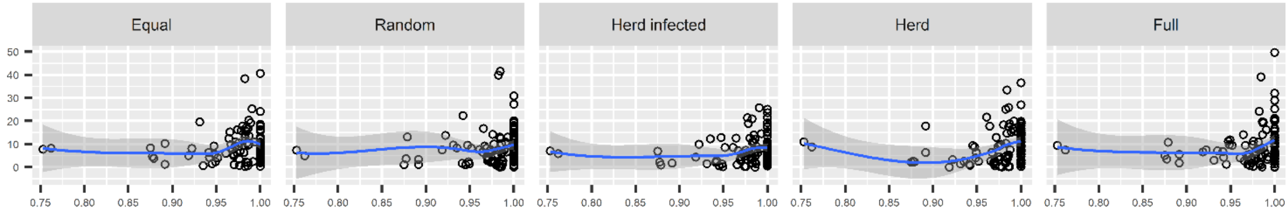


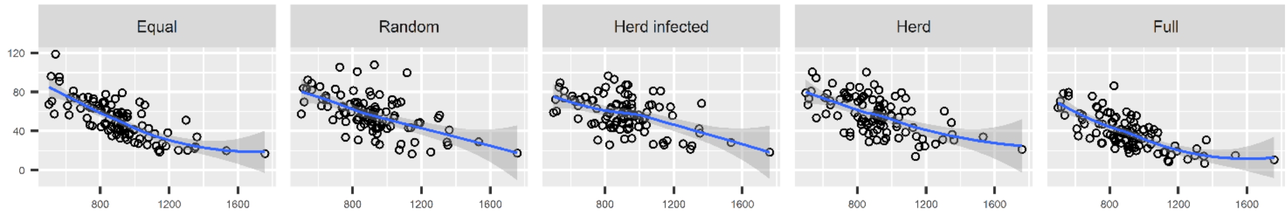
Figure 6-5 Scatter plots showing the associations between percent error (shown in Y-axis) and each simulation characteristic (shown in X-axis).

Simulation characteristics include (A) the last day of sampling, (B) the total number of genetic sequences sampled, (C) the proportion of genetic sequences subsampled out of all samples, (D) the number of animals infected, (E) the proportion of infected farms sampled to the total number of infected farms, (F) the proportion of maximum sample size per geographical region to the total sample size, and (G) the proportion of maximum sample size per community to the total sample size. Blue line and grey areas represent the non-parametric loess smooth curves and their 95% confidence intervals.

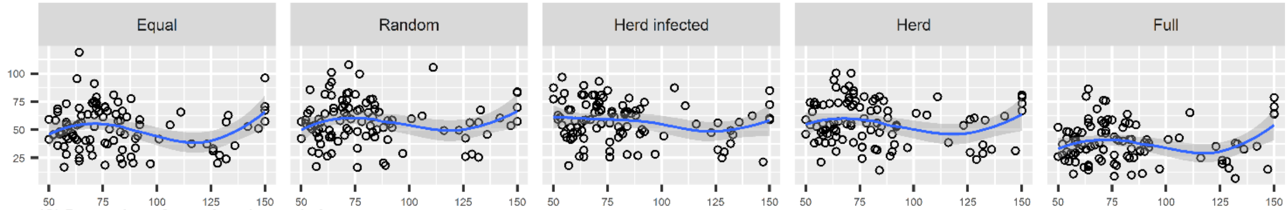
For percent bias, the proportion of maximum sample size collected from one community was identified significant in the final model for ‘equal’, ‘herd’, and ‘full’ strategies (Figure S1 and Table S1 in Appendix 3). No simulation characteristics were significantly associated with percent bias in ‘random’ strategy. Similar to percent error, the proportion of genetic sequences subsampled was significantly associated with percent bias, which was the only variable remained in the final model (Table S1 and Figure S2B in Appendix 3).

The associations between HPD size and each statistic are shown in Figure 6-6. There were strong negative associations between HPD size and the last sampling day, which remained strongly significant in the final model for all strategies (Table S1 in Appendix 3). In addition, the proportion of maximum sample size collected from one geographical region was significantly associated with HPD size in ‘equal’ strategy (Table S1, Figure S2C in Appendix 3).

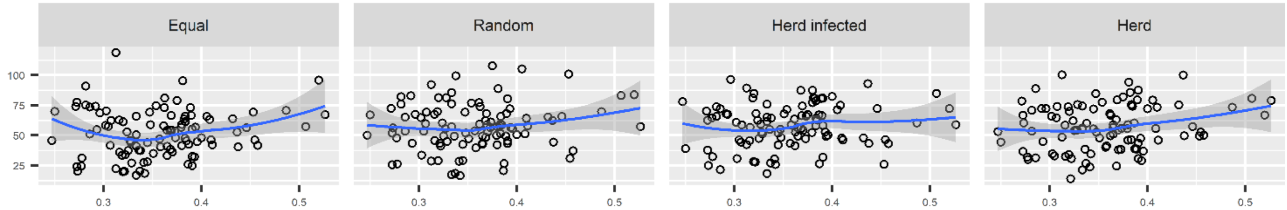
(A) Last sampling day



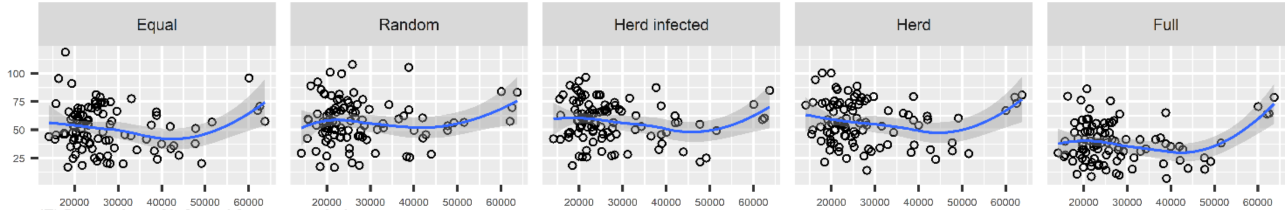
(B) Full sample size



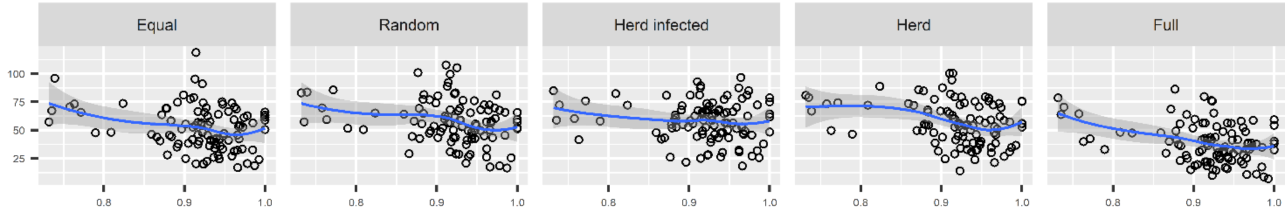
(C) Proportion of sample subsampled



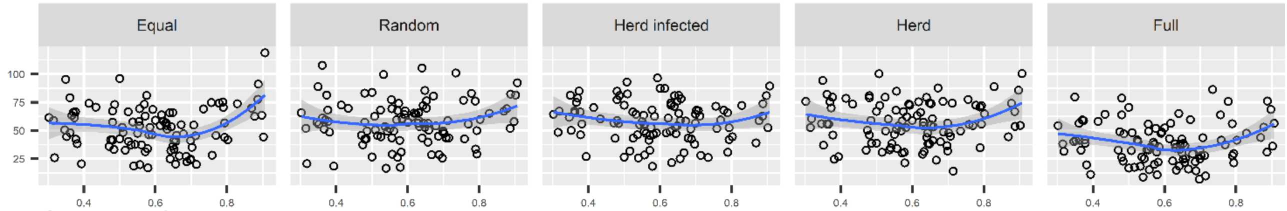
(D) Number of animals infected



(E) Proportion of infected farms sampled



(F) Proportion of max sample size per region



(G) Proportion of max sample size per community

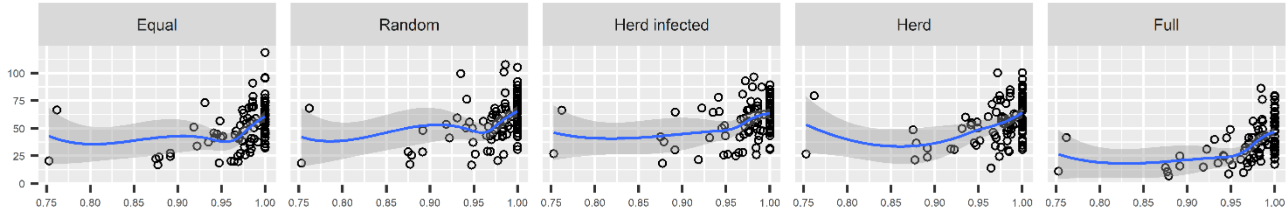


Figure 6-6 Scatter plots showing the associations between HPD size (shown in Y-axis) and each simulation characteristic (shown in X-axis).

Simulation characteristics include (A) the last day of sampling, (B) the total number of genetic sequences sampled, (C) the proportion of genetic sequences subsampled out of all samples, (D) the number of animals infected, (E) the proportion of infected farms sampled to the total number of infected farms, (F) the proportion of maximum sample size per geographical region to the total sample size, and (G) the proportion of maximum sample size per community to the total sample size. Blue line and grey areas represent the non-parametric loess smooth curves and their 95% confidence intervals.

6.5. Discussion

To the best of our knowledge, this is the first study to investigate how different sampling strategies perform when making phylodynamic inferences for diseases spreading through complex livestock contact and population structures. We should note that the results are from an individual-based simulation model, which are highly stochastic and contain a lot of noise. Therefore, the regression analyses merely provide potential correlations and do not suggest conclusive relationships between disease outbreak characteristics and their phylodynamic inferences. Moreover, these results may not be generalisable to other systems such as different country, different production systems, and slow mutation diseases. Nevertheless, our results provide several important considerations when designing sampling strategies for phylodynamic inferences for livestock diseases that spread through animal movements.

Firstly, there was no evidence to suggests that including all possible samples offer a remarkable advantage in the inference over subsampling a smaller proportion of the population. The precision of estimates was improved but a resulting narrower highest posterior density estimation excluded the true value in the 10% of iterations. Moreover, mixing in MCMC iterations was often poor for 'full' strategy, consuming extra computation time. The lack of superiority of 'full' strategy is not surprising. Because every infected farm had a probability to be detected, genetic samples were collected more frequently in a given time when there were a larger number of infected farms. This is often what happens during a disease outbreak. However, this is exactly the situation referred to as 'preferential sampling', which is known to induce biased inferences in geostatistics (Diggle et al., 2010) and more recently in phylodynamics (Karcher et al., 2016). Karcher and colleagues suggested that the estimate of effective population size is biased and less precise if sampling frequency depends on the effective population size and when this is not accounted for (Karcher et al., 2016). Should the detection of infected cases be biased, which is often the case, using all available genetic samples naturally leads to the analysis based on preferentially-sampled samples.

Of the remaining four subsampling strategies, the performance of all strategies was relatively accurate, although there was a tendency to overestimate the TMRCA; that is, models often estimated the date of disease introduction more prior than the true value. Nevertheless, the HPD included the true value almost all the time, despite the violation of the homogeneous mixing assumption. Clearly, the narrow prior we used for the clock rate contributed to this and other phylodynamic parameters such as the effective population size may be more sensitive to the subsampling strategy and violations of assumptions. Although we found no clear evidence of a superiority of one subsampling strategy over another, there was a marginal evidence that suggests some strategies may perform better in certain situations. We found ‘equal’ strategy—the uniform probability sampling with respect to time and geographical areas—did not substantially perform better in our case. We somewhat expected ‘equal’ strategy would perform better than others because this can reduce the effect of preferential sampling—samples were selected in ‘equal’ strategy irrespective of the number of infected farms in a given time in a given region. There are potential explanations for this result, however. Clearly, a geographical region does not represent true epidemiological populations in which a disease likely spreads; livestock movements are not necessarily correlated with geographical areas. Therefore, preferential sampling can still occur in this strategy if such an epidemiological population is very different from populations defined by geographical areas. This may occur when an infection gives rise to a farm that moves animals to many other farms in different geographical areas in a short time.

We also found the higher proportion of infected farms sampled was associated with lower percent error in estimates in ‘equal’ strategy. When the detection coverage is low, infected farms that were not detected were likely to be from regions that had a lower number of infected farms. This results in a biased sampling in terms of geographical areas. These together therefore indicate that 'equal' strategy performs better when the coverage of sampling is less biased. This emphasises the importance of representative sampling of sequences from the population. The non-linear association observed between percent bias and the proportion of maximum sample size from a community ($P_{community}$) in 'equal' strategy is difficult to explain. The higher $P_{community}$ was, however, associated with a smaller value in percent bias in 'herd' and 'full' strategies. An additional analysis suggested that percent error was significantly smaller when disease spread across more than two communities compared to when it spread only within one community in 'herd infected', 'herd' and 'full', but not 'equal' and 'random' strategy (Figure S3 in Appendix 3). This may be because samples from the same region are less epidemiologically linked when disease spread across multiple communities; given that multiple communities can exist in one geographical region, samples from one region may be less epidemiologically linked because they can come from different communities. On the other hand, samples from one region are likely to be more epidemiologically linked when disease spreads within only one community. Further studies are required to identify how disease spread patterns in network communities influence phylodynamic inferences.

A higher proportion of samples that were subsampled seems to deteriorate the accuracy of the TMRCA estimates in 'herd infected' strategy. In our subsampling scheme, a larger subsample size is obtained when sequence isolations did not much cluster in time relative to the length of disease spread period—given we subsampled only one sequence per each sampling interval (see Methods section), clustered samples in each interval result in a smaller proportion of samples subsampled. This can occur, for instance, between-farm disease transmissions occurred through movements of latently infected animals and hence intervals between disease detections were relatively long. In this circumstance, if a disease suddenly spread across farms in one (or few) region this would lead to an unbalance where many samples were collected in a given short time period. Because 'herd infected' strategy does not consider time, this again results in preferential sampling—more epidemiologically linked samples were subsampled. This might explain why this strategy performs worse when a larger proportion of samples was subsampled. However, again, our limited analyses do not provide any conclusive mechanisms to understand this, which should be further investigated in future studies.

It was clearly suggested that the longer duration of a disease spread, which is represented by a longer time of last sampling date, increases the precision of the TMRCA estimates. A longer disease spread period directly reflects more diversities in genetic sequences in our study as a mutation rate stays constant over time. A recent study indeed highlighted that both estimates for clock rate and the TMRCA would be biased if analysed genetic sequence data have little information because these two parameters have a strong correlation (Boskova et al., 2018). In particular, our results suggested that this bias occurred in 'herd' and 'full' strategies, although we set a strong restriction on the clock rate prior; that is, the estimate of TMRCA can be biased and imprecise if sequence data contains little information even when there is good prior knowledge on the pathogen mutation rate if all available samples were included for analyses.

One of the limitations of our study is that we did not use tree priors that have been developed for structured populations (De Maio et al., 2015; Kühnert et al., 2016; Müller et al., 2017; Vaughan et al., 2014). It is well known that violations of panmictic (i.e. homogeneous mixing) assumptions can lead to an erroneous estimation of the past population demographics such as spurious bottleneck signals (Chikhi et al., 2010) and false population declines towards the present (Heller et al., 2013). A theoretical study also showed that TMRCA is influenced by the structure of populations and the intensity of migrations between populations (Notohara and Umeda, 2006). Livestock movement patterns are known to be highly heterogeneous and sporadic and so the panmictic assumption is almost always violated when making phylodynamic inferences for livestock diseases. Two key questions therefore arise around which network structural components are actually important in making phylodynamic inferences and what defines the population structure of pathogens. Network community structure is clearly one of important features that influence disease spreads (Salathé and Jones, 2010). We indeed showed some evidence that the community structure may influence phylodynamic inferences—however, this influence is likely to be complex and non-intuitive. This complexity is reasonable given that disease spread patterns are highly influenced by initial conditions of the epidemic (Bajardi et al., 2012). This complex interrelationship between network characteristics, initial epidemic conditions, and resulting heterogeneous disease spread patterns makes it difficult to identify a situation in which the violation of the homogeneous mixing assumption does not seriously bias the inference.

In terms of the population structure for pathogens, the original coalescent theory proposed by Kingman, population (or deme) refers to a group of genes where any two genetic sequences have the uniform probability to share a common ancestor (Kingman, 1982). In many biological phenomenon, geographical divisions may be a good proxy for the population structure. In such cases, although computational expensive, it is relatively straightforward to use structured models as a tree prior by specifying geographical areas as population and gene migration rates between populations. This approach may be reasonable for livestock diseases that predominantly spread locally with sporadic jumps to other areas, as long as reasonable prior parameters can be specified. Nevertheless, for diseases that spread over complex livestock movements, defining both population structures and migration rates between populations is not straightforward. In particular, the latter is difficult to quantify; it is the migration rate of pathogen genes hence the rate of infection spreads between populations. This parameter is likely to vary over time for livestock diseases; the infection pressure depends on the size of infected individuals in one population and frequencies of livestock movements, both of which dynamically change over time. Currently available structured population models implemented in BEAST can only account for relatively simple demographic models and constant migration rates over time. In reality, these assumptions are likely to be violated. Further studies are therefore warranted to understand which violation of assumptions is critical so that better inferences can be made using available tools.

We also note that we could run only a relatively short simulation because disease spreads died out. This was to some extent inevitable because we needed to limit the disease spreads to occur only through livestock movements. The limited eligible initial seeding conditions that were required for a sufficient disease spread may also have limited the potential variations in phylodynamic inferences; however, as we argued, it is anyway difficult to obtain a reliable phylodynamic inference when disease spread is limited and there are only few variations in the sequence. The use of our simulation conditions was not intended to replicate the real disease spread. Given that livestock movement patterns in New Zealand are known to be highly heterogeneous over time and space (Hidano et al., 2016), our results may not be applicable to other more stable contact systems.

6.6. Conclusion

Sampling genetic sequences for phylodynamic inferences is an important but complex problem. We should be aware that there are at least two layers of sampling issues that can lead to biased phylodynamic estimates—the detection of infected cases from population and subsampling of collected samples for analysis. We showed that bias can still arise even though all infected farms are isolated if we use inappropriate subsampling methods. The performance of subsampling strategies seems to be interdependent on various factors such as the characteristics of the detection of infected cases, initial conditions of the epidemic, population contact structures, and how disease spreads over network communities. Nevertheless, based on our and previous study results, we do not recommend using all available sequences and various subsampling strategies be used to make phylodynamic inferences. Given geographical area does not necessarily represent pathogen population structures when disease spreads through complex networks, subsampling by stratifying based on geographical areas needs a cautious consideration. As applications of phylodynamic techniques are expected to increase for livestock diseases, further studies are required to understand how complex livestock population and contact structures influence phylodynamic inferences.

6.7. Acknowledgements

We thank Chris Jewell for his tremendous technical supports and knowledge for building an individual-based simulation model used in this study. We are also grateful to participants of 'Taming the BEAST in the South Pacific' held in 2017 at New Zealand for useful discussions. Special thanks goes to Tim Vaughan for his constructive suggestions. This study could have never been done without all the knowledge and supports I received at 'Mathematical modelling for infectious diseases' held in 2016 at Welcome Trust, Cambridge, UK. This project was partially supported by the Research Project for Improving Food Safety and Animal Health of the Ministry of Agriculture, Forestry and Fisheries of Japan, the IVABS travel funding from Massey University, Welcome Trust, University of Auckland, and Kathleen Spragg Agricultural Research Trust. AH was supported by the JASSO scholarship, Japan. The funding sources had no involvement in the study design, the collection, analysis and interpretation of data, and the content of and the decision to submit the article for publication.

Chapter 7

General discussion

7.1. Aim of Thesis

The five main aims of the Thesis were to:

1. Identify limitations in the use of new epidemiological data streams available in livestock production systems with particular emphasis on livestock movement data and high-resolution genetic sequence data
2. Make inferences about farmers' livestock purchasing practice using historical livestock movement data
3. Validate the findings from national movement data ('big data') analysis by conducting in-depth qualitative interviews with farmers ('small data')
4. Explore additional hypotheses generated by qualitative interviews ('small data') using national livestock demographic data ('big data')
5. Investigate how different sampling strategies for collecting pathogen genetic sequence data from livestock farms may bias phylodynamic analyses

7.2. Overview of key findings

Although livestock movements play a significant role in spreading infectious diseases between farms, our literature review highlighted that there is very limited knowledge available about how farmers make decisions around buying and selling livestock. There are a few potential reasons that explain this lack of understanding. First, the majority of published studies have focused on describing the statistical properties of livestock movement networks and there has been little interest in investigating the underlying mechanisms that create the observed movement patterns. Second, although there have been an increasing number of studies exploring the reasons why farmers may or may not engage in specific farm management practices, these studies categorise livestock trading practice as one of biosecurity practices and typically look at only one aspect of livestock trading such as whether or not farmers verify disease status of animals they are purchasing. This framing may not, however, be, optimal given that livestock trading decisions involve much more than biosecurity— for instance, farmers may need to purchase livestock to maintain their business or purchase in additional cattle to improve the genetic profile of their herd, which are decisions made independently of biosecurity planning. Therefore, framing livestock trading as a biosecurity issue has not provided sufficient information as to why, how, and when farmers make selling and purchasing decisions. This issue of gaps in framing farm practices between farmers and scientists has been highlighted in Chapter 2.

The next step was to analyse livestock movement records to make inferences about farmers' trading practices in New Zealand. In Chapter 3, we investigated how regionalisation established for bovine tuberculosis (bTB) may influence livestock movement patterns in New Zealand. Regionalisation divides geographical areas into different risk groups. Our hypothesis was that farmers may avoid purchasing livestock from high risk areas to minimize the risk of introducing disease to their farms. To answer this question, we developed a simulation model that constructs livestock movement patterns in various scenarios where farmers consider the disease risk of a source farm to different extent. Using this model, we showed that the observed livestock movement pattern can be explained by the majority of, but not all, farmers avoiding purchasing cattle from high disease risk areas. This study therefore suggested that farmers may be acutually conscious of disease risk when purchasing livestock.

We then attempted to verify this finding from 'big data' analysis. One of the best approach to understanding the underlying mechanism of livestock movement patterns was to conduct in-depth qualitative interviews with farmers since these can explore complex phenomena not amenable to quantitative research (Pope and Mays, 1995). In Chapter 4, we therefore performed a qualitative interview study recruiting farmers from both low bTB risk and high bTB risk areas as well as additional farmers that previously experienced a herd breakdown with bTB —this was to understand the potential difference in livestock trading behaviours between these populations. The study results were somewhat contradictory to that of above simulation study. We found that although farmers do not want to get bTB infections, they do not necessarily consider bTB risk status of the source farms when purchasing livestock for different reasons. This is further discussed below; however, this study exemplifies how relying solely on 'big data' analysis could result in an erroneous conclusion and shows the importance of doing qualitative research to verify the findings.

Our qualitative study provided several additional interesting insights on farmers' livestock trading behaviours. Among these, it was of particular interest in how farmers choose to sell their livestock. Many interviewed farmers indicated that the decision of whether a given cow is going to be sold or culled often depends on the decision of stock agents acting as middle men in the trade. As we discuss later, stock agents seem to have a significant role in dairy cattle trading in New Zealand. This finding therefore naturally leads to the question of whether there are any characteristic differences between animals being culled versus sold. To answer this question, we then in Chapter 5 analysed both livestock movement and animal production records that fortunately co-exist in New Zealand. Our results clearly showed that culled and sold animals have distinct demographic and production characteristics. As we have discussed in Chapter 5, this finding has an important implication for animal disease simulation models that usually assume that animals sold are randomly chosen from herds.

In Chapter 6, we moved onto another new data stream we have identified important—genetic sequence data. Genetic sequence data have been increasingly used to make an epidemiological inference in various ways. Bayesian evolutionary analysis using phylogenetic tree is one of these approaches, which is often referred to as phylodynamic analysis. To the best of our knowledge, however, there are no studies that have evaluated how different sampling strategies of genetic sequences influence inferences made by these analyses for livestock diseases. Using a simulation exercise, we showed that using all available samples can lead to an erroneous inference. More importantly, our results suggested that the performance of different sampling strategies would substantially vary depending on the characteristics of disease epidemics such as livestock contact network structures on which disease spread. This has several important implications we discuss below; in short, however, we contend that appropriate epidemiological knowledge is necessary even when genetic sequence data are solely used to make an epidemiological inference.

7.3. Farmers' livestock trading practice

New Zealand dairy farmers often need to purchase livestock for various reasons. First of all, there has been a substantial conversion from beef and sheep to dairy farms over the last two decades. New dairy farms often start with cows that have suboptimal production capacities and farmers try to improve the overall production capacity of farms over time by introducing cows that are better performers than ones currently in their herds (Chapter 4). Second, the pastured-based dairy system in New Zealand is largely influenced by environmental conditions such as weather and natural disaster. These environmental factors directly influence the growth of pasture and hence fodder availability. Lack of fodder availability means that farmers may need to send their livestock off to grazing-off—grazing their livestock at someone else's paddock, which can be costly depending on the number of animals and geographical distance between their farms and grazing sites. Alternatively, farmers may sell their animals to reduce the fodder requirement. However, as a result of this, farmers may need to purchase cows in the next season to increase their milk production. Third, farmers sometimes need to sell their animals to deal with a decrease in the market milk price (Chapter 5). Dairy production (in fact many other agricultural production systems in New Zealand) heavily depends on the international market, where milk produced in New Zealand accounts for more than one-third of the total milk traded in the world (Anon, 2014). The resulting huge fluctuation in milk price creates a need for dairy farmers to sell and purchase livestock to mitigate the impacts of market volatility.

With these specific New Zealand dairy farm production systems, farmers need to be able to sell and purchase cattle frequently, quickly, and sporadically. We speculate this demand has at least partially contributed to the New Zealand dairy trading system heavily relying on stock agents who are perceived to have up-to-date knowledge on cattle that are available for trade at any given time. All interviewed farmers in our qualitative study mentioned that they use stock agents for livestock trading. This trading system has various important implications for livestock trading practices among New Zealand dairy farmers as we discuss below.

7.3.1. Do farmers consider disease risk status of source farms when purchasing livestock?

This is the question asked in Chapter 3. As already mentioned, however, there was a disparity between findings from Chapter 3 and Chapter 4; 'big data' analysis suggested that farmers avoid purchasing livestock from high bTB risk areas, whereas the qualitative study indicated that farmers do not necessarily consciously consider bTB risk. Dairy cattle livestock trading system that has been revealed in Chapter 4 provides some explanations for this disparity. Farmers often demonstrated that they delegate the purchasing decision making to stock agents. Farmers trust that stock agents would not try to sell problematic cows to them because this would not benefit stock agents either (Chapter 4); stock agents earn profits when they successfully match livestock buyers and sellers, therefore, stock agents need farmers that repeatedly come back to them. Maintaining a good reputation is very important for stock agents. Farmers know that one distrustful deal can heavily damage agents' reputations therefore stock agents would avoid it for their best interest. Also, farmers often value a long-lasting relationship with one agent, although some farmers seem to deal with a couple of stock agents simultaneously. A long-lasting relationship is valued not only because the longer they deal the more farmers can trust agents but also because stock agents come to understand each individual farmer's management style, preference, need, and hence what kind of cattle they are looking for. Purchasing practices then become somewhat an auto-pilot practice, which considerably saves time and effort of farmers. Moreover, some farmers consider that stock agents are experts in disease epidemiology and market price because many stock agent companies have a nation-wide network in which agents exchange information on disease and cattle price. All these together bring a substantial benefit to farmers.

Finding source farms from which cattle are purchased is therefore largely influenced by stock agents' decisions. Stock agents work locally, therefore, source farms are usually found locally too. This facilitates a local livestock trading pattern observed in Chapter 3. Stock agents, however, sometimes cannot find sufficient numbers of cattle locally when many farmers are purchasing livestock in the area for various reasons such as a rise in market milk price. In these circumstances, stock agents ask other agents in other areas to find potential source farms. This facilitates a long-distance livestock trading. Stock agents could theoretically find animals from high disease risk areas, however, this seems unlikely to happen because, again, stock agents need to maintain their reputations and do not want to take a risk of selling diseased livestock to their customers. Moreover, particularly for bTB, farms in high bTB risk areas are often small and they do not have a large number of surplus cattle. This means it is bothersome for agents to find a required number of cattle in high bTB risk areas because they need to gather livestock from multiple farms. These together likely to contribute to the apparent farmers' risk-averse trading behaviours when in fact farmers are not necessarily concerned with bTB disease risk. The role of stock agents and the relationship between agents and farmers warrant a further study.

We contend that this was a good example where a mixed method approach is necessary to avoid making an erroneous inference.

7.3.2. Why farmers do not necessarily verify the disease status of source farms?

Many interviewed farmers clearly indicated that they do not want to introduce serious livestock diseases onto farms. A question then naturally arises: Why do farmers not necessarily verify disease status of animals they are purchasing as recommended by government and animal health experts? The animal health literature often discusses that farmers do not perform this practice because this is not feasible for many farmers (Little et al., 2017a; Ritter et al., 2017) and information asymmetry exists between sellers and buyers (Shortall et al., 2017b). While this may partially explain the results, our qualitative interview study suggests another explanation: A framing gap of a behaviour between farmers and animal health experts, discussed in Chapter 2. Disease status recorded on paper may not necessarily represent animals' health status for farmers for various reasons.

First, interestingly, many interviewed farmers mentioned that they care about clinical status of animals when purchasing livestock. Some farmers even indicated that they visit source farms with stock agents and see animals' conditions in their own eyes (Chapter 4). For farmers, it is their experiences of raising, observing, and interacting with livestock that they can rely on. It is also their instinct that guides them whether or not they can trust source farmers are 'genuine' or good farmers. On the other hand, the disease status recorded on paper can be incomplete and unreliable (Chapter 4). Animals may have originally been born in a high risk area and then moved to the source farm in a low risk area sometime later. Therefore, information available on paper such as farm disease status and risk areas do not convey much information to some farmers. Animal health experts, including us epidemiologist, however, believe that these information are essential for good biosecurity practices, leading to asking the aforementioned question: Why do farmers not necessarily verify disease status of animal they are purchasing? Therefore, verifying disease status on paper is likely framed differently between farmers and animal health experts. A similar framing gap between farmers and experts was highlighted for animal welfare record keeping practices by a recent paper (Escobar and Demeritt, 2017).

It is therefore important to understand that not implementing the biosecurity practices we believe are good does not mean that farmers are not making an effort to improve their on-farm biosecurity. Rather, farmers have their own ways of doing things, which might have developed naturally through their farming life (Chapter 4). We argue that improving on-farm biosecurity requires an understanding of why farmers are currently doing what they are doing and then developing tailored recommendations for practice that can improve their on-farm biosecurity levels.

7.3.3. Selling livestock–is it as simple as selecting random animals?

Our qualitative interview study suggested that some farmers clearly have criteria when deciding which livestock to sell. This was further confirmed in Chapter 5, where we analysed livestock movement and production records. This is another good example of a mixed method approach, which validates qualitative findings in a quantitative manner.

Our quantitative analysis on livestock selling practices showed that various demographic and production characteristics influence the probability of animals being sold (Chapter 5). Nevertheless, based on our observations from the qualitative study (Chapter 4), farmers' livestock selling practices are much more complex than this. Notably, selling animals can influence farms' reputations and, therefore, there is a possibility that farmers may try to maintain a negative herd disease status for specific diseases to protect their reputation as sellers. One farmer in fact demonstrated that he started testing his herd for bovine viral diarrhoea (BVD) virus because some of purchasing farmers requested him to prove the disease status of animals. BVD is assumed to be prevalent among New Zealand farms and farmers may not implement a specific control measure against it, which was the case for this farmer. This specific scenario can result in an interesting situation where some purchasing farmers receive BVD free animals because they request the disease status, and some receive a mixture of BVD negative and positive, or in the extreme case a collection of BVD positive animals because they are not concerned about the disease. We therefore emphasise that selling practices may be interdependent on other farmers' purchasing practices, and that this potential interdependencies can largely vary depending on purchasing farmers' behaviours, awareness and attitude towards diseases.

Farmers' culling practices have been long studied by production scientists and veterinary epidemiologists (Compton et al., 2017). Nevertheless, there is very limited knowledge for livestock selling practices. Many studies did not differentiate selling from culling, or they only looked at culling practices. This is reasonable given their objectives were to identify reasons why cattle are removed from a herd. However, for disease simulation modelling studies, the distinction between these two practices is very important because culled and sold populations can each affect inferences from disease models as we discuss below.

Abattoir inspection data, which are essentially the characteristics of culled animals, are often used to evaluate efficacy of abattoir surveillance (Pascual-Linaza et al., 2017), which can also influence other key epidemiological parameters such as disease transmission rate and diagnostic test accuracies if they are simultaneously estimated (Conlan et al., 2012). Therefore, if the probability of animals being culled is associated with the probability of animals being infected with disease, which can be true for many important production diseases, estimates for these parameters can be biased if this association is not considered.

Similarly, livestock movement data are essentially the characteristics of sold animals. Livestock movement patterns are one of the most important determinants of infectious disease spread patterns. Again, if the probability of animals being sold is associated with the probability of animals being infected with disease, which can be also true for production diseases, this can introduce a new layer of heterogeneity in disease spread patterns.

Nevertheless, these two aspects have not gained attentions so far. As we discussed in Chapter 5, these aspects may be trivial for fast-spreading diseases such as foot-and-mouth disease because all animals in a herd get infected relatively quickly and subsequently get culled. However, for slow-spreading disease, the association between the probability of being infected with disease and the probability of being sold or culled likely gets stronger and larger bias can occur in the inference if these aspects are ignored. The evaluation of the extent of this bias was beyond the scope of this study. It is, however, important for future studies to identify when and to which extent these associations can be safely ignored.

7.4. Phylodynamic inferences for livestock diseases

A freely available and user friendly software for phylodynamic analysis such as BEAST (Bouckaert et al., 2014a; Drummond and Rambaut, 2007) and Mr.Bayes (Ronquist et al., 2012) has increasingly enabled researchers to apply phylodynamic methods to livestock disease pathogen data. However, only a small number of studies have sufficiently examined the validity of various assumptions made by these methods. As already highlighted in Chapter 6, these assumptions include homogeneous contact structure of pathogens and random sampling. In reality, these assumptions are always violated, therefore, we contend that applications of phylodynamic approaches for livestock diseases should not be continued without evaluating the impact of these violations.

It may be intuitive to believe that more genetic samples would bring more information, improving the phylodynamic inferences. We, however, showed that this is not the case when a coalescent model is used as tree prior (Chapter 6). We also showed that network community structure may be an important factor that influences phylodynamic inferences. We did not, however, evaluate a sampling strategy that reflects community structures. This is because we cannot always have information on the livestock movement community structure when sampling pathogens in reality, unless when nearly perfect real time livestock movement records are available. Nevertheless, as we discuss below, such a sampling strategy should be evaluated in future studies to understand how community structure influences phylodynamic inferences.

Our study results suggested that there may be no ‘silver-bullet’ sampling strategies for making a phylodynamic inference for livestock diseases because the characteristics of epidemics are likely to affect different sampling performances. This conveys the most important message from this study—epidemiological knowledge about a given epidemic is very important for phylodynamic analyses to be properly applied even though the analyses can be carried out with minimal epidemiological information. Users of these methods need to be aware of this fact and epidemiologists may have a responsibility to deliver this message given our long history of studying sampling design and bias.

7.5. Methodologies

A configuration-wiring algorithm was the core of the simulation model developed in Chapter 3. As already discussed in Chapter 3, this method has a great advantage to preserve the key temporal and spatial network characteristics that were observed in the real network. This allowed us to replicate the observed distance distribution of dairy cattle movements in New Zealand, which was important because our study objective was to understand whether or not farmers avoid purchasing from high risk areas given a movement distance is fixed. Nevertheless, our method has various limitations. First, we did not consider the size of each livestock movement when we were connecting source and destination farms. This may have resulted in allowing trades to occur in our simulations between source and destination farms that could not have happened in reality because the number of animals source farms were selling and the number of animals destination farms were needing did not match. Second, this model cannot predict future livestock movements because a movement network is constructed based on past network characteristics that are likely to change as we discussed throughout this thesis.

To overcome these limitations, our next step is to develop an algorithm to predict (1) timing and size of livestock selling and purchasing for each farm, and (2) connections between source and destination farms. Regarding (1), Chapter 5 already provides some information that can be used to predict farmers' selling practices. We need a similar analysis to Chapter 5 for farmers' purchasing practices. Regarding (2), Chapter 4 provides some useful information. We now know that dairy cattle movements may be often facilitated by stock agents that work locally, source and destination farms will continue to be locally connected based on the number of animals each farm is selling or purchasing. Chapter 4 also suggested that long distance movements are likely to occur for special reasons including the lack of cattle supply in regions and market price of livestock. Further studies therefore need to consider livestock price data in the models.

Chapter 4 used a qualitative interview method. We used a modified biographical narrative interpretive method (BNIM), starting each interview with a single question. As we discussed in Chapter 2, we contend BNIM is a very powerful approach to extract true information on farmers' behaviours. Nevertheless, the implementation of this method can be improved for various aspects. For instance, the interview approach can be improved. In BNIM, respondents are encouraged to speak for as long as they wish and interviewers are required to make as little input as possible (Burke, 2014). By analysing transcriptions, I feel I can reduce my verbal comments during the interview to reduce my influence on the participants response. Interview approach is a skill, which can be improved by performing more interviews. Also as already mentioned in Chapter 4, the study sample size can be improved in future studies.

Chapter 5 used a mixed-effect logistic regression model to identify risk factors associated with the fate of animals given they are going to be removed from herds. This approach was somewhat different from traditional approaches —many previous studies for livestock culling practices have used survival analysis instead. In fact, we could have used a variation of survival analysis such as multi-state model (Meira-Machado et al., 2009) that can account for different outcomes; culling and selling, in our example. It was, however, questionable whether the survival analysis framework fits to the study objective—time to an event may not an appropriate outcome when analysing selling practices. Our qualitative interview study suggested that the timing of livestock selling is largely governed by the trading system in New Zealand (Chapter 4). For instance, there is a general rule that heifers move between farms through sales or off-grazing at the first week of May every year. The timing of cows being sold is also likely to depend on environmental conditions, rather than any biological phenomena of cows. These together suggest that it is not meaningful to analyses the associations between risk factors and the time to selling practices. The other potential improvement can be done by using a multinomial regression model that can identify the associations between risk factors and all potential outcomes—remain in the herd, culled, or sold. This analysis will allow us to calculate a probability of each of three outcomes for each individual animal. Performing a mixed-effect multinomial regression model with our large data can be, however, too computationally expensive. Given our objective was to show that culled and sold animals have different production and demographic characteristics, this extra analysis was considered unnecessary.

In Chapter 6, it would have been ideal to use a structured coalescent model as tree prior. Nevertheless, this was nearly infeasible for three reasons. First, the computational time required for this model was enormous. This issue could have been mitigated by using a super-computer. However, we could not set up the use of such systems due to limited project time. Second, a structured coalescent model requires users to specify which populations each genetic sample belongs to. This was not straightforward because it is unclear what population really means for livestock diseases that spread through complex network. Third, a structural coalescent model also requires users to specify prior for migration rates between populations. This was a further complicated issue because migration rate is translated into an infection pressure from one population to the other, and it is questionable if such information can be obtained. The structured coalescent models that are currently available also assume that migration rate is constant over time, which is likely to be untrue for many livestock diseases because an infection pressure between populations will change over the volume and frequency of livestock movement and the number of infected farms in the population. With these reasons, we speculate that studies will continue to use a non-structured coalescent model as tree prior for livestock diseases. Therefore, we investigated how different sampling strategies perform when a non-structured coalescent model is used. Nevertheless, future studies should account for livestock population structures and we need to understand when non-structured coalescent models can be safely used to make a phylodynamic inference.

7.6. Future opportunities

7.6.1. Better understanding on livestock movement patterns

Better understanding on livestock movement patterns is essential for developing a livestock disease simulation model that may be reliably used for evaluating effectiveness of disease control strategies. Based on findings from Chapter 3, 4, and 5, we can summarise key areas to be further studied in immediate future.

First, we argue that analysis on livestock movement patterns needs to consider a wider range of factors that can influence livestock movement patterns, shifting from merely describing movement patterns using social network analyses. For instance, we identified that the market milk price may have changed the number of farms that sell livestock (Chapter 5). As already discussed, this factor can be only specific to New Zealand. However, other countries and regions may have their own specific factors that are influential for livestock movement patterns.

Second, given that livestock trading can involve more actors than farmers, behaviours and influences of these actors need to be understood. These actors may include stock agents, farm advisors, and veterinarians, all of which can have a substantial impact on forming livestock trading practices.

A mixed method similar to our studies can be powerful to achieve this objective; qualitative studies can first identify essential components of livestock trading system, then quantitative studies can calibrate parameters of these components to reconstruct and predict livestock movement patterns that can be modified according to the change in the system.

7.6.2. Use of disease simulation models to identify influential human behavioural changes

We developed an individual-based disease model that simulates both within and between herd disease transmission (Chapter 6). We also developed a framework in which livestock movement patterns can be reconstructed reflecting the extent to which farmers avoid purchasing from high bTB risk areas. This framework can be relatively easily extended to model other potential farmers' behavioural changes in trading and biosecurity practices. By incorporating this framework into the developed individual-based simulation model, we can investigate how such behavioural changes influence disease spread patterns. This will allow us to identify critical behavioural changes that are influential for disease spreads, which in turn helps us to determine which behaviours we need to understand more. This is important given that there is currently very limited knowledge on how farmers change behaviours in response to various disease factors and there are too many behaviours to study. Identified important behaviours can then be further studied using qualitative and quantitative research.

7.6.3. Use of disease simulation models to examine the validity of phylodynamic analysis for livestock diseases

We developed a model that simulates disease spread and genetic mutations in Chapter 6. This model can be immediately used in theoretical studies to answer various questions that exist in phylodynamic applications.

First, the immediate extension of Chapter 6 is to evaluate more closely how network community structures influence phylodynamic inferences. We can conduct a very similar study to Chapter 6, but simulating disease spreads on theoretical networks rather than the real complex network. Various theoretical networks will be simulated by modifying the characteristics of community structure such as modularity, the strength of intercommunity edges, and the number of communities (Girvan and Newman, 2002). This kind of study may provide direct relationships between community characteristics and phylodynamic inferences.

Second, this simulation model can be used to fit to a real disease outbreak data to make inferences about how genetic mutation rates are influenced by various biological factors such as host disease status, environmental conditions, and transmission modes. There are several studies that suggested that mutation rates are influenced by host disease status such as latent period (Ford et al., 2011), dormancy and endspores (Didelot et al., 2012; He et al., 2013), hyper mutations (Köser et al., 2012). Information on inconstant mutation rates can then be used as useful prior for further phylodynamic analyses.

Third, this model can be used to evaluate the impact of within-host evolution and multiple infection of pathogens on phylodynamic inferences. Within-host evolution of pathogens has been increasingly known to be an important factor that can hinder making a valid epidemiological inference if not considered (Worby et al., 2014). Various methods have been developed to account for within-host evolution for human disease (Dialdestoro et al., 2016; Volz et al., 2017), however, its impact for livestock disease is still unknown. Similarly, multiple infection of the same pathogen, or ‘super infection’, can be another confounding factor for phylodynamic inferences (Maio et al., 2016). This potential confounding effect can be also examined in our disease simulation framework by allowing multiple infection of pathogens.

7.7. Concluding comments

Despite all the development in science and technologies, we have been still observing serious disease outbreaks, threatening public and animal health and economies. Disease information spreads faster than ever, creating confusion, fear, and pressures on decision makers to implement an appropriate disease control measure. Obtaining reliable epidemiological information in a timely manner is therefore essential for the decision making process.

New data streams including ‘big data’ can play a significant role in providing key epidemiological information that may not be otherwise obtained. In particular, large scale livestock movement records and high resolution genetic sequence data have been increasingly used by researchers to make a better epidemiological inference. Big data are, however, not silver bullets. Bias, validity, and representativeness—which epidemiologists have been working for the last many years—are the central concern for big data. Moreover, how we use big data can lead to making an erroneous epidemiological inference. This was the key coherent message this thesis aimed to deliver.

New data streams will continue to emerge. Likewise, new methods will continue to be developed to analyse these data. I argue that it is important for epidemiologists to keep up with new methods not only for making a better epidemiological inference but also, and more importantly, for examining the validity of these emerging methods. Scrutinising data quality and understanding potential bias that exists in data are important works for epidemiologists. The increasing interdisciplinary research also provides an excellent opportunity for epidemiologists to use qualitative approaches, which I believe are vital for understanding the underlying mechanism of human behaviours and potential bias in big data. For these reasons, I conclude this thesis with my personal belief that developing an epidemiologically critical thinking should continue to be the central theme of educations for young epidemiologists even under this big data era.

Reference

Abramowitz, S., McKune, S.L., Fallah, M., Monger, J., Tehoungue, K., Omidian, P.A., 2017. The Opposite of Denial: Social Learning at the Onset of the Ebola Emergency in Liberia. Journal of Health Communication 22, 59–65. https://doi.org/10.1080/10810730.2016.1209599

Adkin, A., Brouwer, A., Downs, S.H., Kelly, L., 2016. Assessing the impact of a cattle risk-based trading scheme on the movement of bovine tuberculosis infected animals in England and Wales. Preventive Veterinary Medicine 123, 23–31. https://doi.org/10.1016/j.prevetmed.2015.11.021

Ajzen, I., 2015. The theory of planned behaviour is alive and well, and not ready to retire: a commentary on Sniehotta, Presseau, and Araújo-Soares. Health Psychology Review 9, 131–137. https://doi.org/10.1080/17437199.2014.883474

Ajzen, I., 2005. Attitudes, personality, and behavior. [electronic resource], Mapping social psychology. Maidenhead, Berkshire, England ; New York : Open University Press, 2005.

Ajzen, I., Fishbein, M., 1980. Understanding attitudes and predicting social behavior. Englewood Cliffs, N.J. : Prentice-Hall, c1980.

Akerlof, G.A., 1970. The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. Q J Econ 84, 488–500. https://doi.org/10.2307/1879431

Alarcon, P., Wieland, B., Mateus, A.L.P., Dewberry, C., 2014. Pig farmers’ perceptions, attitudes, influences and management of information in the decision-making process for disease control. Preventive Veterinary Medicine, Special Issue: SVEPM 2013 – a truly global conference2013 Society of Veterinary Epidemiology and Preventive Medicine conference 116, 223–242. https://doi.org/10.1016/j.prevetmed.2013.08.004

Aldous, D.J., 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statist. Sci. 16, 23–34. https://doi.org/10.1214/ss/998929474

Alkhamis, M.A., Perez, A.M., Murtaugh, M.P., Wang, X., Morrison, R.B., 2016. Applications of Bayesian Phylodynamic Methods in a Recent U.S. Porcine Reproductive and Respiratory Syndrome Virus Outbreak. Front Microbiol 7. https://doi.org/10.3389/fmicb.2016.00067

Alvåsen, K., Jansson Mörk, M., Hallén Sandgren, C., Thomsen, P.T., Emanuelson, U., 2012. Herd-level risk factors associated with cow mortality in Swedish dairy herds. Journal of Dairy Science 95, 4352–4362. https://doi.org/10.3168/jds.2011-5085

Ambler, original by G., Benner, modified by A., 2015. mfp: Multivariable Fractional Polynomials.

Anneberg, I., Østergaard, S., Ettema, J.F., Kudahl, A.B., 2016. Economic figures in herd health programmes as motivation factors for farmers. Preventive Veterinary Medicine 134, 170–178. https://doi.org/10.1016/j.prevetmed.2016.10.007

Anon, 2018a. Economic Values [WWW Document]. URL https://www.dairynz.co.nz/animal/animal-evaluation/interpreting-the-info/economic-values/ (accessed 5.22.18).

Anon, 2018b. Farmers urged to read NAIT report [WWW Document]. The Beehive. URL http://www.beehive.govt.nz/release/farmers-urged-read-nait-report (accessed 5.16.18).

Anon, 2017a. New Zealand Dairy Statistics 2016-17 [WWW Document]. URL https://www.dairynz.co.nz/publications/dairy-industry/new-zealand-dairy-statistics-2016-17/ (accessed 4.29.18).

Anon, 2017b. Analysis of drivers and barriers to land use change: A Report prepared for the Ministry for Primary Industries [WWW Document]. MPI. URL https://www.mpi.govt.nz/dmsdocument/23056/loggedIn (accessed 5.29.18).

Anon, 2016. New Zealand Dairy Statistics 2015-16 [WWW Document]. URL https://www.dairynz.co.nz/publications/dairy-industry/new-zealand-dairy-statistics-2015-16/ (accessed 4.18.17).

Anon, 2014. Global production and price volatility [WWW Document]. URL https://www.dairynz.co.nz/news/latest-news/global-production-and-price-volatility/ (accessed 6.29.18).

Anon, 2011. New Zealand Dairy Statistics 2010-11.

Anon, 2001. Dairy Industry (Herd Testing and New Zealand Dairy Core Database) Regulations 2001 (SR 2001/373) (as at 04 June 2015) – New Zealand Legislation [WWW Document]. URL http://www.legislation.govt.nz/regulation/public/2001/0373/latest/whole.html (accessed 7.4.16).

Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B., 2002. Estimating Divergence Times from Molecular Data on Phylogenetic and Population Genetic Timescales. Annual Review of Ecology and Systematics 33, 707–740. https://doi.org/10.1146/annurev.ecolsys.33.010802.150500

Bajardi, P., Barrat, A., Natale, F., Savini, L., Colizza, V., 2011. Dynamical patterns of cattle trade movements. PloS one 6, e19869.

Bajardi, P., Barrat, A., Savini, L., Colizza, V., 2012. Optimizing surveillance for livestock disease spreading through animal movements. Journal of The Royal Society Interface 9, 2814–2825. https://doi.org/10.1098/rsif.2012.0289

Bandura, A., 1978. Self-efficacy: Toward a unifying theory of behavioral change. Advances in Behaviour Research and Therapy, Perceived Self-Efficacy: Analyses of Bandura's Theory of Behavioural Change 1, 139–161. https://doi.org/10.1016/0146-6402(78)90002-4

Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., Viboud, C., 2016. Big Data for Infectious Disease Surveillance and Modeling. J Infect Dis 214, S375–S379. https://doi.org/10.1093/infdis/jiw400

Baraniuk, R.G., 2011. More Is Less: Signal Processing and the Data Deluge. Science 331, 717–719. https://doi.org/10.1126/science.1197448

Barnes, A.P., Moxey, A.P., Vosough Ahmadi, B., Borthwick, F.A., 2015. The effect of animal health compensation on 'positive' behaviours towards exotic disease reporting and implementing biosecurity: A review, a synthesis and a research agenda. Preventive Veterinary Medicine 122, 42–52. https://doi.org/10.1016/j.prevetmed.2015.09.003

Bates, T.W., Thurmond, M.C., Carpenter, T.E., 2001. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. American Journal of Veterinary Research 62, 1121–1129. https://doi.org/10.2460/ajvr.2001.62.1121

Bell, G., Hey, T., Szalay, A., 2009. Beyond the Data Deluge. Science 323, 1297–1298. https://doi.org/10.1126/science.1170411

Benjamin, L.A., Fosgate, G.T., Ward, M.P., Roussel, A.J., Feagin, R.A., Schwartz, A.L., 2010. Attitudes towards biosecurity practices relevant to Johne's disease control on beef cattle farms. Preventive Veterinary Medicine 94, 222–230. https://doi.org/10.1016/j.prevetmed.2010.01.001

Bennett, R.M., 2009. Farm costs associated with premovement testing for bovine tuberculosis. Veterinary Record 164, 77–79. https://doi.org/10.1136/vr.164.3.77

Bensemann, J., Shadbolt, N., 2015. Farmers' choice of marketing strategy: A Study of New Zealand Lamb Producers. International Food and Agribusiness Management Review 18, 211–243.

Bergeå, H., Roth, A., Emanuelson, U., Agenäs, S., 2016. Farmer awareness of cow longevity and implications for decision-making at farm level. Acta Agriculturae Scandinavica, Section A — Animal Science 66, 25–34. https://doi.org/10.1080/09064702.2016.1196726

Berger-Wolf, T.Y., Saia, J., 2006. A Framework for Analysis of Dynamic Social Networks, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06. ACM, New York, NY, USA, pp. 523–528. https://doi.org/10.1145/1150402.1150462

Berry, D.P., Harris, B.L., Winkelman, A.M., Montgomerie, W., 2005. Phenotypic Associations Between Traits Other than Production and Longevity in New Zealand Dairy Cattle. Journal of Dairy Science 88, 2962–2974. https://doi.org/10.3168/jds.S0022-0302(05)72976-3

Bertels, F., Silander, O.K., Pachkov, M., Rainey, P.B., Nimwegen, E. van, 2014. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. Mol Biol Evol 31, 1077–1088. https://doi.org/10.1093/molbev/msu088

Blunden, G., Moran, W., Bradly, A., 1997. 'Archaic' Relations of Production in Modern Agricultural Systems: The Example of Sharemilking in New Zealand. Environ Plan A 29, 1759–1776. https://doi.org/10.1068/a291759

Boskova, V., Bonhoeffer, S., Stadler, T., 2014. Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models. PLoS Comput Biol 10, e1003913. https://doi.org/10.1371/journal.pcbi.1003913

Boskova, V., Stadler, T., Magnus, C., 2018. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. Virus Evol 4. https://doi.org/10.1093/ve/vex044

Bostrom, A., 2008. Lead is like mercury: risk comparisons, analogies and mental models. Journal of Risk Research 11, 99–117. https://doi.org/10.1080/13669870701602956

Botterill, L.E.C., Mazur, N., 2004. Risk & risk perception : a literature review : a report for the Rural Industries Research and Development Corporation, RIRDC publication ; 04/043. Rural Industries Research and Development Corporation, Canberra.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014a. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology 10, e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014b. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology 10, 1–6. https://doi.org/10.1371/journal.pcbi.1003537

Brennan, M.L., Christley, R.M., 2012. Biosecurity on Cattle Farms: A Study in North-West England. PLOS ONE 7, e28139. https://doi.org/10.1371/journal.pone.0028139

Brennan, M.L., Wright, N., Wapenaar, W., Jarratt, S., Hobson-West, P., Richens, I.F., Kaler, J., Buchanan, H., Huxley, J.N., O'Connor, H.M., 2016. Exploring Attitudes and Beliefs towards Implementing Cattle Disease Prevention and Control Measures: A Qualitative Study with Dairy Farmers in Great Britain. Animals 6, 61. https://doi.org/10.3390/ani6100061

Brewer, N.T., Chapman, G.B., Gibbons, F.X., Gerrard, M., McCaul, K.D., Weinstein, N.D., 2007. Meta-analysis of the relationship between risk perception and health behavior:

the example of vaccination. Health Psychol 26, 136–145. https://doi.org/10.1037/0278-6133.26.2.136

Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature 511, 228–231. https://doi.org/10.1038/nature13529

Broughan, J.M., Maye, D., Carmody, P., Brunton, L.A., Ashton, A., Wint, W., Alexander, N., Naylor, R., Ward, K., Goodchild, A.V., Hinchliffe, S., Eglin, R.D., Upton, P., Nicholson, R., Enticott, G., 2016. Farm characteristics and farmer perceptions associated with bovine tuberculosis incidents in areas of emerging endemic spread. Preventive Veterinary Medicine 129, 88–98. https://doi.org/10.1016/j.prevetmed.2016.05.007

Buddle, B., de Lisle, G., Griffin, J., Hutchings, S., 2015. Epidemiology, diagnostics, and management of tuberculosis in domestic cattle and deer in New Zealand in the face of a wildlife reservoir. New Zealand Veterinary Journal 1–9. https://doi.org/10.1080/00480169.2014.929518

Burke, C.T., 2014. Biographical Narrative Interview Method: Tracing Graduates' Futures. SAGE Publications, Ltd., 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom. https://doi.org/10.4135/978144627305013510257

Bürkner, P.-C., 2018. brms: Bayesian Regression Models using "Stan."

Bürkner, P.-C., 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software. https://doi.org/10.18637/jss.v080.i01

Burton, R.J.F., 2014. The influence of farmer demographic characteristics on environmental behaviour: A review. Journal of Environmental Management 135, 19–26. https://doi.org/10.1016/j.jenvman.2013.12.005

Burton, R.J.F., 2006. An alternative to farmer age as an indicator of life-cycle stage: The case for a farm family age index. Journal of Rural Studies 22, 485–492. https://doi.org/10.1016/j.jrurstud.2006.02.005

Burton, R.J.F., 2004. Seeing Through the 'Good Farmer's' Eyes: Towards Developing an Understanding of the Social Symbolic Value of 'Productivist' Behaviour. Sociologia Ruralis 44, 195–215. https://doi.org/10.1111/j.1467-9523.2004.00270.x

Burton, R.J.F., Peoples, S., Cooper, M.H., 2012. Building 'cowshed cultures': A cultural perspective on the promotion of stockmanship and animal welfare on dairy farms. Journal of Rural Studies, Rural Realities in the Post-Socialist Space 28, 174–187. https://doi.org/10.1016/j.jrurstud.2011.12.003

Cardwell, J.M., Van Winden, S., Beauvais, W., Mastin, A., De Glanville, W.A., Hardstaff, J., Booth, R.E., Fishwick, J., Pfeiffer, D.U., 2016. Assessing the impact of tailored biosecurity advice on farmer behaviour and pathogen presence in beef herds in England and Wales. Preventive Veterinary Medicine 135, 9–16. https://doi.org/10.1016/j.prevetmed.2016.10.018

Chen, M., Mao, S., Liu, Y., 2014. Big Data: A Survey. Mobile Netw Appl 19, 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chernick, A., Godson, D.L., van der Meer, F., 2014. Metadata beyond the sequence enables the phylodynamic inference of bovine viral diarrhea virus type 1a isolates from Western Canada. Infection, Genetics and Evolution 28, 367–374. https://doi.org/10.1016/j.meegid.2014.01.003

Chikhi, L., Sousa, V.C., Luisi, P., Goossens, B., Beaumont, M.A., 2010. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. Genetics 186, 983–995. https://doi.org/10.1534/genetics.110.118661

Christley, R.M., Pinchbeck, G.L., Bowers, R.G., Clancy, D., French, N.P., Bennett, R., Turner, J., 2005. Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals. Am J Epidemiol 162, 1024–1031. https://doi.org/10.1093/aje/kwi308

Christley, R.M., Robinson, S.E., Moore, B., Setzkorn, C., Donald, I., 2011. Responses of farmers to introduction in England and Wales of pre-movement testing for bovine tuberculosis. Preventive Veterinary Medicine, Special Issue: SVEPM 20102010 Society of Veterinary Epidemiology and Preventive Medicine conference 100, 126–133. https://doi.org/10.1016/j.prevetmed.2011.02.005

Cialdini, R.B., Goldstein, N.J., 2004. Social Influence: Compliance and Conformity. Annual Review of Psychology 55, 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

Clauset, A., Newman, M.E.J., Moore, C., 2004. Finding community structure in very large networks. Physical Review E 70. https://doi.org/10.1103/PhysRevE.70.066111

Clegg, T.A., Duignan, A., More, S.J., 2015. The relative effectiveness of testers during field surveillance for bovine tuberculosis in unrestricted low-risk herds in Ireland. Preventive Veterinary Medicine 119, 85–89. https://doi.org/10.1016/j.prevetmed.2015.02.005

Compton, C.W.R., Heuer, C., Thomsen, P.T., Carpenter, T.E., Phyn, C.V.C., McDougall, S., 2017. Invited review: A systematic literature review and meta-analysis of mortality and culling in dairy cattle. Journal of Dairy Science 100, 1–16. https://doi.org/10.3168/jds.2016-11302

Conlan, A.J.K., Brooks Pollock, E., McKinley, T.J., Mitchell, A.P., Jones, G.J., Vordermeier, M., Wood, J.L.N., 2015. Potential Benefits of Cattle Vaccination as a Supplementary Control for Bovine Tuberculosis. PLOS Computational Biology 11, e1004038. https://doi.org/10.1371/journal.pcbi.1004038

Conlan, A.J.K., McKinley, T.J., Karolemeas, K., Pollock, E.B., Goodchild, A.V., Mitchell, A.P., Birch, C.P.D., Clifton-Hadley, R.S., Wood, J.L.N., 2012. Estimating the Hidden Burden of Bovine Tuberculosis in Great Britain. PLoS Comput Biol 8, e1002730. https://doi.org/10.1371/journal.pcbi.1002730

Conner, M., 2015. Extending not retiring the theory of planned behaviour: a commentary on Sniehotta, Presseau and Araújo-Soares. Health Psychology Review 9, 141–145. https://doi.org/10.1080/17437199.2014.899060

Conner, M., Norman, P., 2017. Health behaviour: Current issues and challenges. Psychology & Health 32, 895–906. https://doi.org/10.1080/08870446.2017.1336240

Corner, R.A., 2002. Two studies on the control of wildlife-derived tuberculosis : farmer views and model validation : a thesis presented in partial fulfilment of the requirements for the degree of Master of Veterinary Studies in Epidemiology at Massey University, Palmerston North, New Zealand (Thesis). Massey University.

Croucher, N.J., Didelot, X., 2015. The application of genomics to tracing bacterial pathogen transmission. Current Opinion in Microbiology, Host–microbe interactions: bacteria • Genomics 23, 62–67. https://doi.org/10.1016/j.mib.2014.11.004

Curry-Roper, J.M., 2000. Embeddedness in Place: Its Role in the Sustainability of a Rural Farm Community in Iowa. Space and Culture 3, 204–222. https://doi.org/10.1177/120633120000300402

Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V., Vernon, M.C., 2011. Networks and the Epidemiology of Infectious Disease. Interdisciplinary Perspectives on Infectious Diseases 2011, 1–28. https://doi.org/10.1155/2011/284909

Dawson, P.M., Werkman, M., Brooks-Pollock, E., Tildesley, M.J., 2015. Epidemic predictions in an imperfect world: modelling disease spread with partial data. Proc. R. Soc. B 282, 20150205. https://doi.org/10.1098/rspb.2015.0205

De Maio, N., Wu, C.-H., O'Reilly, K.M., Wilson, D., 2015. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. PLoS Genet 11, e1005421. https://doi.org/10.1371/journal.pgen.1005421

DeGroot, B.J., Keown, J.F., Van Vleck, L.D., Marotz, E.L., 2002. Genetic Parameters and Responses of Linear Type, Yield Traits, and Somatic Cell Scores to Divergent Selection for Predicted Transmitting Ability for Type in Holsteins1. Journal of Dairy Science 85, 1578–1585. https://doi.org/10.3168/jds.S0022-0302(02)74227-6

Delabouglise, A., Antoine-Moussiaux, N., Phan, T.D., Dao, D.C., Nguyen, T.T., Truong, B.D., Nguyen, X.N.T., Vu, T.D., Nguyen, K.V., Le, H.T., Salem, G., Peyre, M., 2016. The Perceived Value of Passive Animal Health Surveillance: The Case of Highly Pathogenic Avian Influenza in Vietnam. Zoonoses Public Health 63, 112–128. https://doi.org/10.1111/zph.12212

Delabouglise, A., Dao, T.H., Truong, D.B., Nguyen, T.T., Nguyen, N.T.X., Duboz, R., Fournié, G., Antoine-Moussiaux, N., Grosbois, V., Vu, D.T., Le, T.H., Nguyen, V.K., Salem, G., Peyre, M., 2015. When private actors matter: Information-sharing network and surveillance of Highly Pathogenic Avian Influenza in Vietnam. Acta Tropica 147, 38–44. https://doi.org/10.1016/j.actatropica.2015.03.025

Di Nardo, A., Knowles, N.J., Wadsworth, J., Haydon, D.T., King, D.P., 2014. Phylodynamic reconstruction of O CATHAY topotype foot-and-mouth disease virus epidemics in the Philippines. Veterinary Research 45, 90. https://doi.org/10.1186/s13567-014-0090-y

Dialdestoro, K., Sibbesen, J.A., Maretty, L., Raghwani, J., Gall, A., Kellam, P., Pybus, O.G., Hein, J., Jenkins, P.A., 2016. Coalescent Inference Using Serially Sampled, High-Throughput Sequencing Data from Intrahost HIV Infection. Genetics 202, 1449–1472. https://doi.org/10.1534/genetics.115.177931

Didelot, X., Eyre, D.W., Cule, M., Ip, C.L., Ansari, M.A., Griffiths, D., Vaughan, A., O'Connor, L., Golubchik, T., Batty, E.M., Piazza, P., Wilson, D.J., Bowden, R., Donnelly, P.J., Dingle, K.E., Wilcox, M., Walker, A.S., Crook, D.W., Peto, T.E.A., Harding, R.M., 2012. Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. Genome Biology 13, R118. https://doi.org/10.1186/gb-2012-13-12-r118

Diggle, P.J., Menezes, R., Su, T., 2010. Geostatistical inference under preferential sampling. Journal of the Royal Statistical Society: Series C (Applied Statistics) 59, 191–232. https://doi.org/10.1111/j.1467-9876.2009.00701.x

Dohoo, I., Martin, W., Stryhn, H., 2009. Veterinary Epidemiologic Research, 2nd edition. ed. VER Inc., Charlottetown, CA.

Dorjee, S., Revie, C.W., Poljak, Z., McNab, W.B., Sanchez, J., 2013. Network analysis of swine shipments in Ontario, Canada, to support disease spread modelling and risk-based disease management. Preventive Veterinary Medicine 112, 118–127. https://doi.org/10.1016/j.prevetmed.2013.06.008

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biol 4, e88. https://doi.org/10.1371/journal.pbio.0040088

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214. https://doi.org/10.1186/1471-2148-7-214

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol 22, 1185–1192. https://doi.org/10.1093/molbev/msi103

Dubé, C., Ribble, C., Kelton, D., 2010. An analysis of the movement of dairy cattle through 2 large livestock markets in the province of Ontario, Canada. Can Vet J 51, 1254–1260.

Dubé, C., Ribble, C., Kelton, D., McNab, B., 2009. A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. Transboundary and Emerging Diseases 56, 73–85. https://doi.org/10.1111/j.1865-1682.2008.01064.x

Dumas, D., Alexander, P.A., Grossnickle, E.M., 2013. Relational Reasoning and Its Manifestations in the Educational Context: a Systematic Review of the Literature. Educ Psychol Rev 25, 391–427. https://doi.org/10.1007/s10648-013-9224-4

Dyall, T., Miron, D.J., Donald, G.E., Garner, M.G., 2010. Survey of beef cattle producers to better understand factors influencing cattle movements from an exotic disease modelling perspective. Anim. Prod. Sci. 50, 497–502. https://doi.org/10.1071/AN09166

Eames, K., Bansal, S., Frost, S., Riley, S., 2015. Six challenges in measuring contact networks for use in modelling. Epidemics 10, 72–77. https://doi.org/10.1016/j.epidem.2014.08.006

Edmunds, W.J., O'callaghan, C.J., Nokes, D.J., 1997. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. Proceedings of the Royal Society of London B: Biological Sciences 264, 949–957. https://doi.org/10.1098/rspb.1997.0131

Elbers, A.R.W., de Koeijer, A.A., Scolamacchia, F., van Rijn, P.A., 2010. Questionnaire survey about the motives of commercial livestock farmers and hobby holders to vaccinate their animals against Bluetongue virus serotype 8 in 2008–2009 in the Netherlands. Vaccine 28, 2473–2481. https://doi.org/10.1016/j.vaccine.2010.01.047

Ensoy, C., Faes, C., Welby, S., Van der Stede, Y., Aerts, M., 2014. Exploring cattle movements in Belgium. Preventive Veterinary Medicine 116, 89–101. https://doi.org/10.1016/j.prevetmed.2014.05.003

Enticott, G., 2016. Market instruments, biosecurity and place-based understandings of animal disease. Journal of Rural Studies 45, 312–319. https://doi.org/10.1016/j.jrurstud.2016.04.008

Enticott, G., 2014. Relational distance, neoliberalism and the regulation of animal health. Geoforum 52, 42–50. https://doi.org/10.1016/j.geoforum.2013.12.004

Enticott, G., 2012a. The local universality of veterinary expertise and the geography of animal disease. Transactions of the Institute of British Geographers 37, 75–88. https://doi.org/10.1111/j.1475-5661.2011.00452.x

Enticott, G., 2012b. Regulating animal health, gender and quality control: A study of veterinary surgeons in Great Britain. Journal of Rural Studies, Growing Old in Rural Places 28, 559–567. https://doi.org/10.1016/j.jrurstud.2012.05.004

Enticott, G., 2011. Techniques of neutralising wildlife crime in rural England and Wales. Journal of Rural Studies 27, 200–208. https://doi.org/10.1016/j.jrurstud.2011.01.005

Enticott, G., 2008. The ecological paradox: social and natural consequences of the geographies of animal health promotion. Transactions of the Institute of British Geographers 33, 433–446. https://doi.org/10.1111/j.1475-5661.2008.00321.x

Enticott, G., Maye, D., Carmody, P., Naylor, R., Ward, K., Hinchliffe, S., Wint, W., Alexander, N., Elgin, R., Ashton, A., Upton, P., Nicholson, R., Goodchild, T., Brunton, L., Broughan, J., 2015a. Farming on the edge: farmer attitudes to bovine tuberculosis in newly endemic areas. Veterinary Record 177, 439–439. https://doi.org/10.1136/vr.103187

Enticott, G., Maye, D., Carmody, P., Naylor, R., Ward, K., Hinchliffe, S., Wint, W., Alexander, N., Elgin, R., Ashton, A., Upton, P., Nicholson, R., Goodchild, T., Brunton, L., Broughan, J., 2015b. Farming on the edge: farmer attitudes to bovine tuberculosis in newly endemic areas. Veterinary Record 177, 439–439. https://doi.org/10.1136/vr.103187

Enticott, G., Maye, D., Fisher, R., Ilbery, B., Kirwan, J., 2014. Badger Vaccination: Dimensions of Trust and Confidence in the Governance of Animal Disease. Environ Plan A 46, 2881–2897.

Enticott, G., Wilkinson, K., 2013. Biosecurity: Whose Knowledge Counts?, in: Dobson, A., Barker, K., Taylor, S. (Eds.), Biosecurity: The Socio-Politics of Invasive Species and Infectious Diseases. Routledge, London and New York, pp. 91–104.

Escobar, M.P., Demeritt, D., 2017. Paperwork and the decoupling of audit and animal welfare: The challenges of materiality for better regulation. Environment and Planning C: Politics and Space 35, 169–190. https://doi.org/10.1177/0263774X16646771

Evans, M.R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C.M., Merz, M., O’Malley, M.A., Orzack, S.H., Weisberg, M., Wilkinson, D.J., Wolkenhauer, O., Benton, T.G., 2013. Do simple models lead to generality in ecology? Trends in Ecology & Evolution 28, 578–583. https://doi.org/10.1016/j.tree.2013.05.022

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17, 368–376. https://doi.org/10.1007/BF01734359

Ferguson, N., 2007. Capturing human behaviour. Nature 446, 733–733. https://doi.org/10.1038/446733a

Ferrer, R.A., Klein, W.M., 2015. Risk perceptions and health behavior. Current Opinion in Psychology, Health behavior 5, 85–89. https://doi.org/10.1016/j.copsyc.2015.03.012

Field, N., Cohen, T., Struelens, M.J., Palm, D., Cookson, B., Glynn, J.R., Gallo, V., Ramsay, M., Sonnenberg, P., MacCannell, D., Charlett, A., Egger, M., Green, J., Vineis, P., Abubakar, I., 2014. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. The Lancet Infectious Diseases 14, 341–352. https://doi.org/10.1016/S1473-3099(13)70324-4

Firestone, S.M., Lewis, F.I., Schemann, K., Ward, M.P., Toribio, J.-A.L.M.L., Taylor, M.R., Dhand, N.K., 2014. Applying Bayesian network modelling to understand the links between on-farm biosecurity practice during the 2007 equine influenza outbreak and horse managers’ perceptions of a subsequent outbreak. Preventive Veterinary Medicine, Special Issue: SVEPM 2013 – a truly global conference 116, 243–251. https://doi.org/10.1016/j.prevetmed.2013.11.015

Fischer, E.A.J., Verwaart, T., Valeeva, N.I., Hagenaars, T.J., 2017. The feedback loop between endemic disease and pig farmer decision making: An agent-based simulation study, in: Proceedings SVEPM Conference. Inverness, Scotland, pp. 225–237.

Fish, R., Austin, Z., Christley, R., Haygarth, P.M., Heathwaite, L.A., Latham, S., Medd, W., Mort, M., Oliver, D.M., Pickup, R., Wastling, J.M., Wynne, B., 2011. Uncertainties in the governance of animal disease: an interdisciplinary framework for analysis. Philosophical Transactions of the Royal Society of London B: Biological Sciences 366, 2023–2034. https://doi.org/10.1098/rstb.2010.0400

Fishbein, M., Ajzen, I., 2010. Predicting and changing behavior : the reasoned action approach. New York : Psychology Press, c2010.

Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T.R., Sacchettini, J.C., Lipsitch, M., Flynn, J.L., Fortune, S.M., 2011. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat Genet 43, 482–486. https://doi.org/10.1038/ng.811

Fourment, M., Holmes, E.C., 2015. Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry. BMC Evol Biol 15. https://doi.org/10.1186/s12862-015-0410-5

Fournié, G., Guitian, J., Desvaux, S., Cuong, V.C., Dung, D.H., Pfeiffer, D.U., Mangtani, P., Ghani, A.C., 2013. Interventions for avian influenza A (H5N1) risk management in live bird market networks. PNAS 110, 9177–9182. https://doi.org/10.1073/pnas.1220815110

Foxman, B., Riley, L., 2001. Molecular Epidemiology: Focus on Infection. Am. J. Epidemiol. 153, 1135–1141. https://doi.org/10.1093/aje/153.12.1135

Funk, S., Bansal, S., Bauch, C.T., Eames, K.T.D., Edmunds, W.J., Galvani, A.P., Klepac, P., 2015. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. Epidemics, Challenges in Modelling Infectious DIsease Dynamics 10, 21–25. https://doi.org/10.1016/j.epidem.2014.09.005

Funk, S., Gilad, E., Jansen, V.A.A., 2010. Endemic disease, awareness, and local behavioural response. Journal of Theoretical Biology 264, 501–509. https://doi.org/10.1016/j.jtbi.2010.02.032

Funk, S., Gilad, E., Watkins, C., Jansen, V.A.A., 2009. The spread of awareness and its impact on epidemic outbreaks. PNAS 106, 6872–6877. https://doi.org/10.1073/pnas.0810762106

Funk, Sebastian, Salathé, M., Jansen, V.A.A., 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of The Royal Society Interface 7, 1247–1256. https://doi.org/10.1098/rsif.2010.0142

Gardy, J.L., Johnston, J.C., Sui, S.J.H., Cook, V.J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., others, 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. New England Journal of Medicine 364, 730–739.

Garforth, C., 2015. Livestock Keepers' Reasons for Doing and Not Doing Things Which Governments, Vets and Scientists Would Like Them to Do. Zoonoses Public Health 62, 29–38. https://doi.org/10.1111/zph.12189

Garforth, C., Rehman, T., McKemey, K., Tranter, R., Cooke, R., Yates, C., Park, J., Dorward, P., 2004. Improving the Design of Knowledge Transfer Strategies by Understanding Farmer Attitudes and Behaviour. Journal of Farm Management 12, 17–32.

Garforth, C.J., Bailey, A.P., Tranter, R.B., 2013. Farmers' attitudes to disease risk management in England: A comparative analysis of sheep and pig farmers. Preventive Veterinary Medicine 110, 456–466. https://doi.org/10.1016/j.prevetmed.2013.02.018

Garner, M.G., Hamilton, S.A., 2011. Principles of epidemiological modelling. Rev. - Off. Int. Epizoot. 30, 407–416.

Gates, M.C., 2014. Characteristics of replacement breeding cattle trade in Great Britain. Veterinary Record 175, 67–67. https://doi.org/10.1136/vr.102351

Gates, M.C., Holmstrom, L.K., Biggers, K.E., Beckham, T.R., 2015. Integrating Novel Data Streams to Support Biosurveillance in Commercial Livestock Production Systems in Developed Countries: Challenges and Opportunities. Front. Public Health 3. https://doi.org/10.3389/fpubh.2015.00074

Gates, M.C., Volkova, V.V., Woolhouse, M.E.J., 2013. Impact of changes in cattle movement regulations on the risks of bovine tuberculosis for Scottish farms.

Preventive Veterinary Medicine 108, 125–136. https://doi.org/10.1016/j.prevetmed.2012.07.016

Gates, M.C., Woolhouse, M.E.J., 2015. Controlling infectious disease through the targeted manipulation of contact network structure. Epidemics. https://doi.org/10.1016/j.epidem.2015.02.008

Gavryushkina, A., Welch, D., Stadler, T., Drummond, A.J., 2014. Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration. PLoS Comput Biol 10, e1003919. https://doi.org/10.1371/journal.pcbi.1003919

Geraghty, T., Graham, D.A., Mullowney, P., More, S.J., 2014. A review of bovine Johne's disease control activities in 6 endemically infected countries. Preventive Veterinary Medicine 116, 1–11. https://doi.org/10.1016/j.prevetmed.2014.06.003

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., Park, D.J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., others, 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345, 1369–1372.

Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99, 7821–7826.

Glanz, K., Rimer, B., Viswanath, K., 2015. Health Behavior: Theory, Research, and Practice, 5th Edition.

Goldlust, S., Lee, E.C., Haran, M., Rohani, P., Bansal, S., 2017. Assessing the distribution and determinants of vaccine underutilization in the United States. bioRxiv 113043. https://doi.org/10.1101/113043

Goodreau, S.M., 2007. Advances in exponential random graph (p*) models applied to a large social network. Social Networks 29, 231–248. https://doi.org/10.1016/j.socnet.2006.08.001

Green, D.., Kiss, I.., Kao, R.., 2006. Modelling the initial spread of foot-and-mouth disease through animal movements. Proceedings of the Royal Society B: Biological Sciences 273, 2729–2735. https://doi.org/10.1098/rspb.2006.3648

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L.N., Daly, J.M., Mumford, J.A., Holmes, E.C., 2004. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. Science 303, 327–332. https://doi.org/10.1126/science.1090727

Griffiths, R.C., Tavare, S., 1994. Sampling Theory for Neutral Alleles in a Varying Environment. Philosophical Transactions of the Royal Society of London B: Biological Sciences 344, 403–410. https://doi.org/10.1098/rstb.1994.0079

Grisi-Filho, J.H.H., Amaku, M., Ferreira, F., Dias, R.A., Neto, J.S.F., Negreiros, R.L., Ossada, R., 2013. Detecting livestock production zones. Preventive Veterinary Medicine 110, 304–311. https://doi.org/10.1016/j.prevetmed.2012.12.013

Guest, G., Bunce, A., Johnson, L., 2006. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods 18, 59–82. https://doi.org/10.1177/1525822X05279903

Guimera, R., Sales-Pardo, M., Amaral, L.A., 2007. Classes of complex networks defined by role-to-role connectivity profiles. Nature physics 3, 63–69.

Gunarathne, A., Kubota, S., Kumarawadu, P., Karunagoda, K., Kon, H., 2015. Is Hiding Foot and Mouth Disease Sensitive Behavior for Farmers? A Survey Study in Sri Lanka, Is Hiding Foot and Mouth Disease Sensitive Behavior for Farmers? A Survey Study in Sri Lanka. Asian-Australas J Anim Sci 29, 280–287. https://doi.org/10.5713/ajas.15.0241

Gutnik, L.A., Hakimzada, A.F., Yoskowitz, N.A., Patel, V.L., 2006. The role of emotion in decision-making: A cognitive neuroeconomic approach towards understanding sexual risk behavior. Journal of Biomedical Informatics 39, 720–736. https://doi.org/10.1016/j.jbi.2006.03.002

Guy, S., Kashima, Y., Walker, I., O'Neill, S., 2013. Comparing the atmosphere to a bathtub: effectiveness of analogy for reasoning about accumulation. Climatic Change 121, 579–594. https://doi.org/10.1007/s10584-013-0949-3

Hadley, G.L., Wolf, C.A., Harsh, S.B., 2006. Dairy Cattle Culling Patterns, Explanations, and Implications. Journal of Dairy Science 89, 2286–2296. https://doi.org/10.3168/jds.S0022-0302(06)72300-1

Hagger, M.S., Chatzisarantis, N.L.D., 2014. An Integrated Behavior Change Model for Physical Activity: Exercise and Sport Sciences Reviews 42, 62–69. https://doi.org/10.1249/JES.0000000000000008

Hagger, M.S., Gucciardi, D.F., Chatzisarantis, N.L.D., 2017. On Nomological Validity and Auxiliary Assumptions: The Importance of Simultaneously Testing Effects in Social Cognitive Theories Applied to Health Behavior and Some Guidelines. Front Psychol 8. https://doi.org/10.3389/fpsyg.2017.01933

Haine, D., Cue, R., Sewalem, A., Wade, K., Lacroix, R., Lefebvre, D., Rushton, J., Arsenault, J., Bouchard, É., Dubuc, J., 2017. Culling from the actors' perspectives—Decision-making criteria for culling in Québec dairy herds enrolled in a veterinary preventive medicine program. Preventive Veterinary Medicine 148, 1–9. https://doi.org/10.1016/j.prevetmed.2017.09.015

Halasa, T., Dürr, S., 2017. Editorial: Modeling Disease Spread and Control. Front. Vet. Sci. 4. https://doi.org/10.3389/fvets.2017.00199

Hall, A., 1996. Editorial: What is molecular epidemiology? Tropical Medicine & International Health 1, 407–408. https://doi.org/10.1046/j.1365-3156.1996.d01-96.x

Hall, M.D., Woolhouse, M.E.J., Rambaut, A., 2016. The effects of sampling strategy on the quality of reconstruction of viral population dynamics using Bayesian skyline family coalescent methods: A simulation study. Virus Evolution 2, vew003. https://doi.org/10.1093/ve/vew003

Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., Lencastre, H. de, Parkhill, J., Peacock, S.J., Bentley, S.D., 2010. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. Science 327, 469–474. https://doi.org/10.1126/science.1182395

Hastie, T., 2018. gam: Generalized Additive Models.

He, M., Miyajima, F., Roberts, P., Ellison, L., Pickard, D.J., Martin, M.J., Connor, T.R., Harris, S.R., Fairley, D., Bamford, K.B., D'Arc, S., Brazier, J., Brown, D., Coia, J.E., Douce, G., Gerding, D., Kim, H.J., Koh, T.H., Kato, H., Senoh, M., Louie, T., Michell, S., Butt, E., Peacock, S.J., Brown, N.M., Riley, T., Songer, G., Wilcox, M., Pirmohamed, M., Kuijper, E., Hawkey, P., Wren, B.W., Dougan, G., Parkhill, J., Lawley, T.D., 2013. Emergence and global spread of epidemic healthcare-associated Clostridium difficile. Nat Genet 45, 109–113. https://doi.org/10.1038/ng.2478

Heckhausen, H., Gollwitzer, P.M., 1987. Thought contents and cognitive functioning in motivational versus volitional states of mind. Motiv Emot 11, 101–120. https://doi.org/10.1007/BF00992338

Hedlund, J., 2000. Risky business: safety regulations, risk compensation, and individual behavior. Injury Prevention 6, 82–89. https://doi.org/10.1136/ip.6.2.82

Hein, J., Schierup, M.H., Wiuf, C., 2005. Gene Genealogies, Variation and Evolution : A Primer in Coalescent Theory. Oxford University Press, Oxford.

Heled, J., Drummond, A.J., 2008. Bayesian inference of population size history from multiple loci. BMC Evolutionary Biology 8, 289. https://doi.org/10.1186/1471-2148-8-289

Heller, R., Chikhi, L., Siegismund, H.R., 2013. The Confounding Effect of Population Structure on Bayesian Skyline Plot Inferences of Demographic History. PLOS ONE 8, e62992. https://doi.org/10.1371/journal.pone.0062992

Hennessy, D.A., 2007. Biosecurity and Spread of an Infectious Animal Disease. Am. J. Agr. Econ. 89, 1226–1231. https://doi.org/10.1111/j.1467-8276.2007.01088.x

Hennink, M.M., Kaiser, B.N., Marconi, V.C., 2017. Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? Qual Health Res 27, 591–608. https://doi.org/10.1177/1049732316665344

Hidano, A., Carpenter, T.E., Gates, M.C., Stevenson, M.A., Enticott, G., 2017a. “Big data” and “small data”: towards a better understanding of disease data by modelling human behaviour, in: 3rd International Conference on Animal Health Surveillance. Presented at the 3rd International Conference on Animal Health Surveillance, Rotorua, New Zealand.

Hidano, A., Carpenter, T.E., Stevenson, M.A., Gates, M.C., 2016. Evaluating the efficacy of regionalisation in limiting high-risk livestock trade movements. Preventive Veterinary Medicine 133, 31–41. https://doi.org/10.1016/j.prevetmed.2016.09.015

Hidano, A., Enticott, G., Christley, R.M., Gates, M.C., 2018a. Modelling dynamic human behavioural changes in animal disease models: challenges and opportunities for addressing bias. Front. Vet. Sci. 5. https://doi.org/10.3389/fvets.2018.00137

Hidano, A., Gates, M.C., 2018. Why sold, not culled? Analysing farm and animal characteristics associated with livestock selling practices. Submitted.

Hidano, A., Gates, M.C., Carpenter, T.E., Enticott, G., 2017b. Farmer decision-making around livestock trading practices: the multifactorial roles of stock agents, in: Inaugural Meeting of ISESSA 2017. Presented at the International Society for Economics and Social Sciences of Animal Health, International Society for Economics and Social Sciences of Animal Health, Aviemore, Scotland, p. 91.

Hidano, A., Gates, M.C., Enticott, G., 2018b. Understanding farmers’ livestock trading decision making: A qualitative study. In preparation.

Higgins, H.M., Mouncey, J., Nanjiani, I., Cook, A.J.C., 2017. Understanding how new evidence influences practitioners’ beliefs regarding dry cow therapy: A Bayesian approach using probabilistic elicitation. Preventive Veterinary Medicine, SVEPM 2016 - Annual Conference of The Society for Veterinary Epidemiology and Preventive Medicine 139, 115–122. https://doi.org/10.1016/j.prevetmed.2016.08.012

Ho, S.Y.W., Duchêne, S., 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol 23, 5947–5965. https://doi.org/10.1111/mec.12953

Ho, S.Y.W., Larson, G., 2006. Molecular clocks: when timesare a-changin’. Trends in Genetics 22, 79–83. https://doi.org/10.1016/j.tig.2005.11.006

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Genet 4, 275–284. https://doi.org/10.1038/nrg1044

Holmes, C.W., Brookes, I.M., Garrick, D.J., Mackenzie, D.D.S., Parkinson, T.J., Wilson, G.F., 2002. Milk Production from Pasture, 2nd ed. Massey University.

Hoscheit, P., Geeraert, S., Beaunée, G., Monod, H., Gilligan, C.A., Filipe, J.A.N., Vergu, E., Moslonka-Lefebvre, M., 2017. Dynamical network models for cattle trade: towards economy-based epidemic risk assessment. J Complex Netw 5, 604–624. https://doi.org/10.1093/comnet/cnw026

Huelsenbeck, J.P., Suchard, M.A., 2007. A Nonparametric Method for Accommodating and Testing Across-Site Rate Variation. Syst Biol 56, 975–987. https://doi.org/10.1080/10635150701670569

igraph: Network Analysis and Visualization, 2018.

Jackman, S., Tahk, with contributions from A., Zeileis, A., Maimone, C., Meers, J.F. and Z., 2017. pscl: Political Science Computational Laboratory.

Jackson, C.H., 2008. Displaying Uncertainty With Shading. The American Statistician 62, 340–347. https://doi.org/10.1198/000313008X370843

Jewell, C.P., Andel, M. van, Vink, W.D., McFadden, A.M.J., 2016. Compatibility between livestock databases used for quantitative biosecurity response in New Zealand. New Zealand Veterinary Journal 64, 158–164. https://doi.org/10.1080/00480169.2015.1117955

Kao, R.R., Haydon, D.T., Lycett, S.J., Murcia, P.R., 2014. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends in Microbiology 22, 282–291. https://doi.org/10.1016/j.tim.2014.02.011

Karafillakis, E., Larson, H.J., 2017. The benefit of the doubt or doubts over benefits? A systematic literature review of perceived risks of vaccines in European populations. Vaccine 35, 4840–4850. https://doi.org/10.1016/j.vaccine.2017.07.061

Karcher, M.D., Palacios, J.A., Bedford, T., Suchard, M.A., Minin, V.N., 2016. Quantifying and Mitigating the Effect of Preferential Sampling on Phylodynamic Inference. PLOS Comput Biol 12, e1004789. https://doi.org/10.1371/journal.pcbi.1004789

Keeling, M.J., 2005. Models of foot-and-mouth disease. Proceedings of the Royal Society B: Biological Sciences 272, 1195–1202. https://doi.org/10.1098/rspb.2004.3046

Kelly, P., O'Sullivan, K., Berry, D., More, S., Meaney, W., O'Callaghan, E., O'Brien, B., 2009. Farm management factors associated with bulk tank somatic cell count in Irish dairy herds. Ir Vet J 62, S45–S51. https://doi.org/10.1186/2046-0481-62-S4-S45

Khoury, M.J., Lam, T.K., Ioannidis, J.P.A., Hartge, P., Spitz, M.R., Buring, J.E., Chanock, S.J., Croyle, R.T., Goddard, K.A., Ginsburg, G.S., Herceg, Z., Hiatt, R.A., Hoover, R.N., Hunter, D.J., Kramer, B.S., Lauer, M.S., Meyerhardt, J.A., Olopade, O.I., Palmer, J.R., Sellers, T.A., Seminara, D., Ransohoff, D.F., Rebbeck, T.R., Tourassi, G., Winn, D.M., Zauber, A., Schully, S.D., 2013. Transforming Epidemiology for 21st Century Medicine and Public Health. Cancer Epidemiol Biomarkers Prev 22, 508–516. https://doi.org/10.1158/1055-9965.EPI-13-0146

King, G., 2011. Ensuring the Data-Rich Future of the Social Sciences. Science 331, 719–721. https://doi.org/10.1126/science.1197872

Kingman, J.F.C., 1982. The coalescent. Stochastic Processes and their Applications 13, 235–248. https://doi.org/10.1016/0304-4149(82)90011-4

Kirman, A., Livet, P., Teschl, M., 2010. Rationality and emotions. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 215–219. https://doi.org/10.1098/rstb.2009.0194

Klein, W.M.P., Zajac, L.E., Monin, M.M., 2009. Worry as a Moderator of the Association Between Risk Perceptions and Quitting Intentions in Young Adult and Adult Smokers. ann. behav. med. 38, 256–261. https://doi.org/10.1007/s12160-009-9143-2

Kobayashi, M., Melkonyan, T., 2011. Strategic Incentives in Biosecurity Actions: Theoretical and Empirical Analyses. Journal of Agricultural and Resource Economics 36, 242–262.

Koh, D., Takahashi, K., Lim, M.-K., Imai, T., Chia, S.-E., Qian, F., Ng, V., Fones, C., 2005. SARS Risk Perception and Preventive Measures, Singapore and Japan. Emerg Infect Dis 11, 641–642. https://doi.org/10.3201/eid1104.040765

Köser, C.U., Holden, M.T.G., Ellington, M.J., Cartwright, E.J.P., Brown, N.M., Ogilvy-Stuart, A.L., Hsu, L.Y., Chewapreecha, C., Croucher, N.J., Harris, S.R., Sanders, M., Enright, M.C., Dougan, G., Bentley, S.D., Parkhill, J., Fraser, L.J., Betley, J.R., Schulz-Trieglaff, O.B., Smith, G.P., Peacock, S.J., 2012. Rapid Whole-Genome

Sequencing for Investigation of a Neonatal MRSA Outbreak. New England Journal of Medicine 366, 2267–2275. https://doi.org/10.1056/NEJMoa1109910

Kristensen, E., Jakobsen, E.B., 2011. Challenging the myth of the irrational dairy farmer; understanding decision-making related to herd health. New Zealand Veterinary Journal 59, 1–7. https://doi.org/10.1080/00480169.2011.547162

Kristensen, Erling, Jakobsen, E.B., 2011. Danish dairy farmers' perception of biosecurity. Preventive Veterinary Medicine 99, 122–129. https://doi.org/10.1016/j.prevetmed.2011.01.010

Krivitsky, P.N., 2012. Exponential-family random graph models for valued networks. Electron. J. Statist. 6, 1100–1128. https://doi.org/10.1214/12-EJS696

Kuehne, G., Llewellyn, R., Pannell, D.J., Wilkinson, R., Dolling, P., Ouzman, J., Ewing, M., 2017. Predicting farmer uptake of new agricultural practices: A tool for research, extension and policy. Agricultural Systems 156, 115–125. https://doi.org/10.1016/j.agsy.2017.06.007

Kühnert, D., Stadler, T., Vaughan, T.G., Drummond, A.J., 2016. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. Mol Biol Evol 33, 2102–2116. https://doi.org/10.1093/molbev/msw064

Kumar, S., 2005. Molecular clocks: four decades of evolution. Nat Rev Genet 6, 654–662. https://doi.org/10.1038/nrg1659

Kwasnicka, D., Dombrowski, S.U., White, M., Sniehotta, F., 2016. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychology Review 10, 277–296. https://doi.org/10.1080/17437199.2016.1151372

Lambie, N.R., 1998. An application of dynamic microeconomic theory to bovine tuberculosis control in New Zealand cattle (Thesis). Lincoln University.

Lazer, D., Kennedy, R., King, G., Vespignani, A., 2014. The Parable of Google Flu: Traps in Big Data Analysis. Science 343, 1203–1205. https://doi.org/10.1126/science.1248506

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Alstyne, M.V., 2009. Computational Social Science. Science 323, 721–723. https://doi.org/10.1126/science.1167742

Lemey, P., Salemi, M., Vandamme, A.-M., 2009. The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press.

Lindström, T., Sisson, S.A., Lewerin, S.S., Wennergren, U., 2011. Bayesian analysis of animal movements related to factors at herd and between herd levels: Implications for disease spread modeling. Preventive Veterinary Medicine 98, 230–242. https://doi.org/10.1016/j.prevetmed.2010.11.005

Lindström, T., Sisson, S.A., Nöremark, M., Jonsson, A., Wennergren, U., 2009. Estimation of distance related probability of animal movements between holdings and implications for disease spread modeling. Preventive Veterinary Medicine 91, 85–94. https://doi.org/10.1016/j.prevetmed.2009.05.022

Little, R., Wheeler, K., Edge, S., 2017a. Developing a risk-based trading scheme for cattle in England: farmer perspectives on managing trading risk for bovine tuberculosis. Veterinary Record 180, 148–148. https://doi.org/10.1136/vr.103522

Little, R., Wheeler, K., Edge, S., 2017b. Developing a risk-based trading scheme for cattle in England: farmer perspectives on managing trading risk for bovine tuberculosis. Veterinary Record 180, 148–148. https://doi.org/10.1136/vr.103522

Livingstone, P., Hancox, N., Nugent, G., de Lisle, G., 2015a. Toward eradication: the effect of *Mycobacterium bovis* infection in wildlife on the evolution and future direction of bovine tuberculosis management in New Zealand. New Zealand Veterinary Journal 1–15. https://doi.org/10.1080/00480169.2014.971082

Livingstone, P., Hancox, N., Nugent, G., Mackereth, G., Hutchings, S., 2015b. Development of the New Zealand strategy for local eradication of tuberculosis from wildlife and livestock. New Zealand Veterinary Journal 1–10. https://doi.org/10.1080/00480169.2015.1013581

Livingstone, P.G., Ryan, T.J., Hancox, N.G., Crews, K.B., Bosson, M.A.J., Knowles, G.J.E., McCook, W., 2006. Regionalisation: A strategy that will assist with bovine tuberculosis control and facilitate trade. Veterinary Microbiology, 4th International Conference on Mycobacterium bovis4th International Conference on Mycobacterium bovis 112, 291–301. https://doi.org/10.1016/j.vetmic.2005.11.016

Lowe, P., Phillipson, J., 2009. Barriers to Research Collaboration across Disciplines: Scientific Paradigms and Institutional Practices. Environ Plan A 41, 1171–1184. https://doi.org/10.1068/a4175

Lupton, D., 2013. Risk and emotion: towards an alternative theoretical perspective. Health, Risk & Society 15, 634–647. https://doi.org/10.1080/13698575.2013.848847

Lusher, D., Koskinen, J., Robins, G., 2012. Exponential Random Graph Models for Social Networks. Cambridge University Press.

Madin, B., 2011. Understanding and predicting the influence of animal movement on the spread of transboundary animal diseases (phd). Murdoch University.

Maio, N.D., Wu, C.-H., Wilson, D.J., 2016. SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent. PLOS Comput Biol 12, e1005130. https://doi.org/10.1371/journal.pcbi.1005130

Manabe, T., Hanh, T.T., Lam, D.M., Van, D.T.H., Thuy, P.T.P., Thi Thanh Huyen, D., Thi Mai Phuong, T., Minh, D.H., Takasaki, J., Chau, N.Q., Toan, L.Q., Kudo, K., 2012. Knowledge, Attitudes, Practices and Emotional Reactions among Residents of Avian Influenza (H5N1) Hit Communities in Vietnam. PLoS One 7. https://doi.org/10.1371/journal.pone.0047560

Mancy, R., Brock, P.M., Kao, R.R., 2017. An Integrated Framework for Process-Driven Model Construction in Disease Ecology and Animal Health. Front. Vet. Sci. 4. https://doi.org/10.3389/fvets.2017.00155

Mankad, A., 2016. Psychological influences on biosecurity control and farmer decision-making. A review. Agron. Sustain. Dev. 36, 40. https://doi.org/10.1007/s13593-016-0375-9

Marquetoux, N., Stevenson, M.A., Wilson, P., Ridler, A., Heuer, C., 2016. Using social network analysis to inform disease control interventions. Preventive Veterinary Medicine 126, 94–104. https://doi.org/10.1016/j.prevetmed.2016.01.022

Martínez-López, B., Perez, A.M., Sánchez-Vizcaíno, J.M., 2009. Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine. Transboundary and Emerging Diseases 56, 109–120. https://doi.org/10.1111/j.1865-1682.2009.01073.x

Maye, D., Enticott, G., Fisher, R., 2015. Neoliberalising Nature: A Longitudinal Study of Badger Vaccination., in: Places of Possibility? Rural Societies in a Neoliberal World. Presented at the XXVI European Society for Rural Sociology Congress, James Hutton Institute, Aberdeen, Scotland, p. 365.

Maye, D., Enticott, G., Naylor, R., 2017. Using scenario-based influence mapping to examine farmers' biosecurity behaviour. Land Use Policy 66, 265–277. https://doi.org/10.1016/j.landusepol.2017.04.026

McAloon, C.G., Macken-Walsh, Á., Moran, L., Whyte, P., More, S.J., O'Grady, L., Doherty, M.L., 2017. Johne's disease in the eyes of Irish cattle farmers: A qualitative narrative research approach to understanding implications for disease management. Preventive Veterinary Medicine 141, 7–13. https://doi.org/10.1016/j.prevetmed.2017.04.001

McConnel, C.S., Lombard, J.E., Wagner, B.A., Garry, F.B., 2008. Evaluation of Factors Associated with Increased Dairy Cow Mortality on United States Dairy Operations. Journal of Dairy Science 91, 1423–1432. https://doi.org/10.3168/jds.2007-0440

McCormick, B.J.J., Breda, L.K., Ward, M.P., 2017. Bayesian Network analysis of piglet scours. Scientific Reports 7, 6202. https://doi.org/10.1038/s41598-017-06399-2

McDougall, S., Bryan, M.A., Tiddy, R.M., 2009. Effect of treatment with the nonsteroidal antiinflammatory meloxicam on milk production, somatic cell count, probability of re-treatment, and culling of dairy cows with mild clinical mastitis. Journal of Dairy Science 92, 4421–4431. https://doi.org/10.3168/jds.2009-2284

Meira-Machado, L., Uña-Álvarez, J. de, Cadarso-Suárez, C., Andersen, P.K., 2009. Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 18, 195–222. https://doi.org/10.1177/0962280208092301

Mellado, M., Reséndiz, D., Martínez, A.M., Santiago, M.A. de, Véliz, F.G., García, J.E., 2015. Milk yield and reproductive performance of Holstein cows testing positive for bovine tuberculosis. Trop Anim Health Prod 47, 1061–1066. https://doi.org/10.1007/s11250-015-0828-1

Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., Vespignani, A., 2011. Modeling human mobility responses to the large-scale spreading of infectious diseases. Scientific Reports 1, 62. https://doi.org/10.1038/srep00062

Millman, C., Christley, R., Rigby, D., Dennis, D., O'Brien, S.J., Williams, N., 2017. "Catch 22": Biosecurity awareness, interpretation and practice amongst poultry catchers. Preventive Veterinary Medicine 141, 22–32. https://doi.org/10.1016/j.prevetmed.2017.04.002

Miranda-de la Lama, G.C., Villarroel, M., María, G.A., 2014. Livestock transport from the perspective of the pre-slaughter logistic chain: a review. Meat Science 98, 9–20. https://doi.org/10.1016/j.meatsci.2014.04.005

Misztal, I., Lawlor, T.J., Short, T.H., VanRaden, P.M., 1992. Multiple-Trait Estimation of Variance Components of Yield and Type Traits Using an Animal Model. Journal of Dairy Science 75, 544–551. https://doi.org/10.3168/jds.S0022-0302(92)77791-1

Mitchell, A., Bourn, D., Mawdsley, J., Wint, W., Clifton-Hadley, R., Gilbert, M., 2005. Characteristics of cattle movements in Britain – an analysis of records from the Cattle Tracing System. Animal Science 80, 265–273. https://doi.org/10.1079/ASC50020265

Mollentze, N., Nel, L.H., Townsend, S., Roux, K. le, Hampson, K., Haydon, D.T., Soubeyrand, S., 2014. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proceedings of the Royal Society of London B: Biological Sciences 281, 20133251. https://doi.org/10.1098/rspb.2013.3251

Molloy, M., Reed, B., 1998. The Size of the Giant Component of a Random Graph with a Given Degree Sequence. Combinatorics, Probability and Computing 7, 295–305. https://doi.org/null

Mooney, S.J., Westreich, D.J., El-Sayed, A.M., 2015. Commentary: Epidemiology in the era of big data. Epidemiology 26, 390–394. https://doi.org/10.1097/EDE.0000000000000274

Moslonka-Lefebvre, M., Pautasso, M., Jeger, M.J., 2009. Disease spread in small-size directed networks: Epidemic threshold, correlation between links to and from nodes, and clustering. Journal of Theoretical Biology 260, 402–411. https://doi.org/10.1016/j.jtbi.2009.06.015

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J., 2008. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLOS Medicine 5, e74. https://doi.org/10.1371/journal.pmed.0050074

Müller, N.F., Rasmussen, D.A., Stadler, T., 2017. The Structured Coalescent and Its Approximations. Mol Biol Evol 34, 2970–2981. https://doi.org/10.1093/molbev/msx186

Murray, A.G., 2014. A game theory based framework for assessing incentives for local area collaboration with an application to Scottish salmon farming. Preventive Veterinary Medicine 115, 255–262. https://doi.org/10.1016/j.prevetmed.2014.03.023

Nascimento, F.F., dos Reis, M., Yang, Z., 2017. A biologist’s guide to Bayesian phylogenetic analysis. Nat Ecol Evol 1, 1446–1454. https://doi.org/10.1038/s41559-017-0280-x

Naylor, R., Hamilton-Webb, A., Little, R., Maye, D., 2018. The ‘Good Farmer’: Farmer Identities and the Control of Exotic Livestock Disease in England. Sociologia Ruralis 58, 3–19. https://doi.org/10.1111/soru.12127

Nekouei, O., VanLeeuwen, J., Stryhn, H., Kelton, D., Keefe, G., 2016. Lifetime effects of infection with bovine leukemia virus on longevity and milk production of dairy cows. Preventive Veterinary Medicine 133, 1–9. https://doi.org/10.1016/j.prevetmed.2016.09.011

Nguyen, T.T.T., Fearnley, L., Dinh, X.T., Tran, T.T.A., Tran, T.T., Nguyen, V.T., Tago, D., Padungtod, P., Newman, S.H., Tripodi, A., 2017. A Stakeholder Survey on Live Bird Market Closures Policy for Controlling Highly Pathogenic Avian Influenza in Vietnam. Front. Vet. Sci. 4. https://doi.org/10.3389/fvets.2017.00136

Nishi, T., Yamada, M., Fukai, K., Shimada, N., Morioka, K., Yoshida, K., Sakamoto, K., Kanno, T., Yamakawa, M., 2017. Genome variability of foot-and-mouth disease virus during the short period of the 2010 epidemic in Japan. Veterinary Microbiology 199, 62–67. https://doi.org/10.1016/j.vetmic.2016.12.025

Northern Ireland Audit Office, 2009. The Control of Bovine Tuberculosis in Northern Ireland [WWW Document]. Report by the Comptroller and Auditor General NIA 92/08-09, 18 March 2009. URL https://www.niauditoffice.gov.uk/publication/control-bovine-tuberculosis-northern-ireland (accessed 1.25.18).

Notohara, M., Umeda, T., 2006. The coalescence time of sampled genes in the structured coalescent model. Theoretical Population Biology 70, 289–299. https://doi.org/10.1016/j.tpb.2006.05.005

Oraby, T., Thampi, V., Bauch, C.T., 2014. The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. Proc. R. Soc. B 281, 20133172. https://doi.org/10.1098/rspb.2013.3172

Orbell, S., Sheeran, P., 1998. ‘Inclined abstainers’: A problem for predicting health-related behaviour. British Journal of Social Psychology 37, 151–165. https://doi.org/10.1111/j.2044-8309.1998.tb01162.x

Oreszczyn, S., Lane, A., Carr, S., 2010. The role of networks of practice and webs of influencers on farmers’ engagement with and learning about agricultural innovations. Journal of Rural Studies 26, 404–417. https://doi.org/10.1016/j.jrurstud.2010.03.003

OSPRI, 2015. Annual Report 2014/2015.

Pananos, A.D., Bury, T.M., Wang, C., Schonfeld, J., Mohanty, S.P., Nyhan, B., Salathé, M., Bauch, C.T., 2017. Critical dynamics in population vaccinating behavior. PNAS 114, 13762–13767. https://doi.org/10.1073/pnas.1704093114

Pascual-Linaza, A.V., Gordon, A.W., Stringer, L.A., Menzies, F.D., 2017. Efficiency of slaughterhouse surveillance for the detection of bovine tuberculosis in cattle in Northern Ireland. Epidemiology & Infection 145, 995–1005. https://doi.org/10.1017/S0950268816003095

Paul, M., Baritaux, V., Wongnarkpet, S., Poolkhet, C., Thanapongtharm, W., Roger, F., Bonnet, P., Ducrot, C., 2013. Practices associated with Highly Pathogenic Avian Influenza spread in traditional poultry marketing chains: Social and economic perspectives. Acta Tropica 126, 43–53. https://doi.org/10.1016/j.actatropica.2013.01.008

Pohlert, T., 2018. PMCMRplus: Calculate Pairwise Multiple Comparisons of Mean Rank Sums Extended.

Pomeroy, L.W., Bansal, S., Tildesley, M., Moreno-Torres, K.I., Moritz, M., Xiao, N., Carpenter, T.E., Garabed, R.B., 2017. Data-Driven Models of Foot-and-Mouth Disease Dynamics: A Review. Transbound Emerg Dis 64, 716–728. https://doi.org/10.1111/tbed.12437

Pond, S.L.K., Frost, S.D.W., 2005. A Simple Hierarchical Approach to Modeling Distributions of Substitution Rates. Mol Biol Evol 22, 223–234. https://doi.org/10.1093/molbev/msi009

Pope, C., Mays, N., 1995. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 311, 42–45.

Prestwich, A., Webb, T.L., Conner, M., 2015. Using theory to develop and test interventions to promote changes in health behaviour: evidence, issues, and recommendations. Current Opinion in Psychology, Health behavior 5, 1–5. https://doi.org/10.1016/j.copsyc.2015.02.011

Pritchard, T.C., Coffey, M.P., Bond, K.S., Hutchings, M.R., Wall, E., 2017. Phenotypic effects of subclinical paratuberculosis (Johne’s disease) in dairy cattle. Journal of Dairy Science 100, 679–690. https://doi.org/10.3168/jds.2016-11323

Probert, W.J.M., Shea, K., Fonnesbeck, C.J., Runge, M.C., Carpenter, T.E., Dürr, S., Garner, M.G., Harvey, N., Stevenson, M.A., Webb, C.T., Werkman, M., Tildesley, M.J., Ferrari, M.J., 2016. Decision-making for foot-and-mouth disease control: Objectives matter. Epidemics 15, 10–19. https://doi.org/10.1016/j.epidem.2015.11.002

Pybus, O.G., 2006. Model Selection and the Molecular Clock. PLoS Biol 4, e151. https://doi.org/10.1371/journal.pbio.0040151

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet 10, 540–550. https://doi.org/10.1038/nrg2583

Raboisson, D., Cahuzac, E., Sans, P., Allaire, G., 2011. Herd-level and contextual factors influencing dairy cow mortality in France in 2005 and 2006. Journal of Dairy Science 94, 1790–1803. https://doi.org/10.3168/jds.2010-3634

Racicot, M., Venne, D., Durivage, A., Vaillancourt, J.-P., 2012a. Evaluation of strategies to enhance biosecurity compliance on poultry farms in Québec: Effect of audits and cameras. Preventive Veterinary Medicine 103, 208–218. https://doi.org/10.1016/j.prevetmed.2011.08.004

Racicot, M., Venne, D., Durivage, A., Vaillancourt, J.-P., 2012b. Evaluation of the relationship between personality traits, experience, education and biosecurity compliance on poultry farms in Québec, Canada. Preventive Veterinary Medicine 103, 201–207. https://doi.org/10.1016/j.prevetmed.2011.08.011

Racicot, M., Venne, D., Durivage, A., Vaillancourt, J.-P., 2011. Description of 44 biosecurity errors while entering and exiting poultry barns based on video surveillance in Quebec, Canada. Preventive Veterinary Medicine 100, 193–199. https://doi.org/10.1016/j.prevetmed.2011.04.011

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol. https://doi.org/10.1093/sysbio/syy032

Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.

Rasmussen, D.A., Volz, E.M., Koelle, K., 2014. Phylodynamic inference for structured epidemiological models. PLoS computational biology 10, e1003570.

Rat-Aspert, O., Fourichon, C., 2010. Modelling collective effectiveness of voluntary vaccination with and without incentives. Preventive Veterinary Medicine, SVEPM 2009Papers presented at the 2009 annual meeting of the Society for Veterinary Epidemiology and Preventive Medicine (SVEPM) 93, 265–275. https://doi.org/10.1016/j.prevetmed.2009.11.004

Reimus, K., Orro, T., Emanuelson, U., Viltrop, A., Mõtus, K., 2018. On-farm mortality and related risk factors in Estonian dairy cows. Preventive Veterinary Medicine 155, 53–60. https://doi.org/10.1016/j.prevetmed.2018.04.006

Relun, A., Grosbois, V., Alexandrov, T., Sánchez-Vizcaíno, J.M., Waret-Szkuta, A., Molia, S., Etter, E.M.C., Martínez-López, B., 2017. Prediction of Pig Trade Movements in Different European Production Systems Using Exponential Random Graph Models. Front Vet Sci 4. https://doi.org/10.3389/fvets.2017.00027

Rhodes, R.E., Janssen, I., Bredin, S.S.D., Warburton, D.E.R., Bauman, A., 2017. Physical activity: Health impact, prevalence, correlates and interventions. Psychology & Health 32, 942–975. https://doi.org/10.1080/08870446.2017.1325486

Richens, I.F., Hobson-West, P., Brennan, M.L., Hood, Z., Kaler, J., Green, M., Wright, N., Wapenaar, W., 2016. Factors influencing veterinary surgeons' decision-making about dairy cattle vaccination. Veterinary Record 179, 410–410. https://doi.org/10.1136/vr.103822

Ridout, M.M. and M., 2018. overlap: Estimates of Coefficient of Overlapping for Animal Activity Patterns.

Ritchie, M., Berthouze, L., House, T., Kiss, I.Z., 2014. Higher-order structure and epidemic dynamics in clustered networks. Journal of Theoretical Biology 348, 21–32. https://doi.org/10.1016/j.jtbi.2014.01.025

Ritter, C., Jansen, J., Roche, S., Kelton, D.F., Adams, C.L., Orsel, K., Erskine, R.J., Benedictus, G., Lam, T.J.G.M., Barkema, H.W., 2017. Invited review: Determinants of farmers' adoption of management-based strategies for infectious disease prevention and control. Journal of Dairy Science. https://doi.org/10.3168/jds.2016-11977

Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007. An introduction to exponential random graph (p*) models for social networks. Social Networks, Special Section: Advances in Exponential Random Graph (p*) Models 29, 173–191. https://doi.org/10.1016/j.socnet.2006.08.002

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenberg, N.A., Nordborg, M., 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet 3, 380–390. https://doi.org/10.1038/nrg795

Ryan, T., Cameron, C., Cato, A., 1997. Control of tuberculosis in cattle and deer in intensive production areas in New Zealand : effect of animal movements. ISVEE ISVEE 8: Proceedings of the 8th Symposium of the International Society for Veterinary Epidemiology and Economics, Paris, France (published as Epidemiologie et Sante Animale, Issues 31-32), 11.

Ryan, T.J., Livingstone, P.G., Ramsey, D.S.L., de Lisle, G.W., Nugent, G., Collins, D.M., Buddle, B.M., 2006. Advances in understanding disease epidemiology and implications for control and eradication of tuberculosis in livestock: The experience from New Zealand. Veterinary Microbiology 112, 211–219. https://doi.org/10.1016/j.vetmic.2005.11.025

Salathé, M., Jones, J.H., 2010. Dynamics and control of diseases in networks with community structure. PLoS computational biology 6, e1000736.

Sanson, R.L., 2005. A survey to investigate movements off sheep and cattle farms in New Zealand, with reference to the potential transmission of foot-and-mouth disease. New Zealand Veterinary Journal 53, 223–233. https://doi.org/10.1080/00480169.2005.36550

Santman-Berends, I.M.G.A., Buddiger, M., Smolenaars, A.J.G., Steuten, C.D.M., Roos, C.A.J., Van Erp, A.J.M., Van Schaik, G., 2014. A multidisciplinary approach to determine factors associated with calf rearing practices and calf mortality in dairy herds. Preventive Veterinary Medicine, Special Issue: SVEPM 2014 - supporting decision making on animal health through advanced and multidisciplinary methodologies, 2014 Society of Veterinary Epidemiology and Preventive Medicine conference 117, 375–387. https://doi.org/10.1016/j.prevetmed.2014.07.011

Sauerbrei, W., Meier-Hirmer, C., Benner, A., Royston, P., 2006. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis 50, 3464–3485. https://doi.org/10.1016/j.csda.2005.07.015

Sauter-Louis, C., 2001. The effectiveness of on-farm control programmes against wildlife-derived bovine tuberculosis in New Zealand: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy at Massey University.

Scarpino, S.V., Iamarino, A., Wells, C., Yamin, D., Ndeffo-Mbah, M., Wenzel, N.S., Fox, S.J., Nyenswah, T., Altice, F.L., Galvani, A.P., Meyers, L.A., Townsend, J.P., 2015. Epidemiological and Viral Genomic Sequence Analysis of the 2014 Ebola Outbreak Reveals Clustered Transmission. Clin Infect Dis. 60, 1079–1082. https://doi.org/10.1093/cid/ciu1131

Schemann, K., Taylor, M.R., Toribio, J.-A.L.M.L., Dhand, N.K., 2011. Horse owners' biosecurity practices following the first equine influenza outbreak in Australia. Preventive Veterinary Medicine 102, 304–314. https://doi.org/10.1016/j.prevetmed.2011.08.002

Schmid, F., Schmidt, A., 2006. Nonparametric estimation of the coefficient of overlapping—theory and empirical application. Computational Statistics & Data Analysis 50, 1583–1596. https://doi.org/10.1016/j.csda.2005.01.014

Segerdahl, P., 2007. Can natural behavior be cultivated? The farm as local human/animal culture. J Agric Environ Ethics 20, 167–193. https://doi.org/10.1007/s10806-006-9028-3

Serrano, M.A., Boguna, M., 2005. Weighted Configuration Model. arXiv:cond-mat/0501750 776, 101–107. https://doi.org/10.1063/1.1985381

Sheather, S.J., Jones, M.C., 1991. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. Journal of the Royal Statistical Society. Series B (Methodological) 53, 683–690.

Shortall, O., Green, M., Brennan, M., Wapenaar, W., Kaler, J., 2017a. Exploring expert opinion on the practicality and effectiveness of biosecurity measures on dairy farms in the United Kingdom using choice modeling. Journal of Dairy Science 100, 2225–2239. https://doi.org/10.3168/jds.2016-11435

Shortall, O., Green, M., Brennan, M., Wapenaar, W., Kaler, J., 2017b. Exploring expert opinion on the practicality and effectiveness of biosecurity measures on dairy farms in the United Kingdom using choice modeling. Journal of Dairy Science 100, 2225–2239. https://doi.org/10.3168/jds.2016-11435

Shortall, O., Ruston, A., Green, M., Brennan, M., Wapenaar, W., Kaler, J., 2016. Broken biosecurity? Veterinarians' framing of biosecurity on dairy farms in England. Preventive Veterinary Medicine 132, 20–31. https://doi.org/10.1016/j.prevetmed.2016.06.001

Shortall, O., Sutherland, L.-A., Ruston, A., Kaler, J., 2018. True Cowmen and Commercial Farmers: Exploring Vets' and Dairy Farmers' Contrasting Views of 'Good Farming' in Relation to Biosecurity. Sociologia Ruralis n/a-n/a. https://doi.org/10.1111/soru.12205

Silva, E. de, Ferguson, N.M., Fraser, C., 2012. Inferring pandemic growth rates from sequence data. Journal of The Royal Society Interface 9, 1797–1808. https://doi.org/10.1098/rsif.2011.0850

Simmons, M.P., 2014. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. Molecular Phylogenetics and Evolution 80, 267–280. https://doi.org/10.1016/j.ympev.2014.08.021

Sniehotta, F.F., Presseau, J., Araújo-Soares, V., 2015. On the development, evaluation and evolution of health behaviour theory. Health Psychology Review 9, 176–189. https://doi.org/10.1080/17437199.2015.1022902

Sniehotta, F.F., Presseau, J., Araújo-Soares, V., 2014. Time to retire the theory of planned behaviour. Health Psychology Review 8, 1–7. https://doi.org/10.1080/17437199.2013.869710

Sok, J., Hogeveen, H., Elbers, A.R.W., Oude Lansink, A.G.J.M., 2016. Using farmers' attitude and social pressures to design voluntary Bluetongue vaccination strategies. Preventive Veterinary Medicine 133, 114–119. https://doi.org/10.1016/j.prevetmed.2016.09.016

Stack, J.C., Welch, J.D., Ferrari, M.J., Shapiro, B.U., Grenfell, B.T., 2010. Protocols for sampling viral sequences to study epidemic dynamics. Journal of The Royal Society Interface rsif20090530. https://doi.org/10.1098/rsif.2009.0530

Stadler, T., 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. Journal of Theoretical Biology 261, 58–66. https://doi.org/10.1016/j.jtbi.2009.07.018

Stadler, T., Kouyos, R., Wyl, V. von, Yerly, S., Böni, J., Bürgisser, P., Klimkait, T., Joos, B., Rieder, P., Xie, D., Günthard, H.F., Drummond, A.J., Bonhoeffer, S., Study, the S.H.C., 2012. Estimating the Basic Reproductive Number from Viral Sequence Data. Mol Biol Evol 29, 347–357. https://doi.org/10.1093/molbev/msr217

Stadler, T., Kuhnert, D., Bonhoeffer, S., Drummond, A.J., 2013. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences 110, 228–233. https://doi.org/10.1073/pnas.1207965110

Stan Development Team, 2018. "RStan: the R interface to Stan." R package version 2.17.3.

Steel, M., 2005. Should phylogenetic models be trying to 'fit an elephant'? Trends in Genetics 21, 307–309. https://doi.org/10.1016/j.tig.2005.04.001

Sun, Z., Lorscheid, I., Millington, J.D., Lauf, S., Magliocca, N.R., Groeneveld, J., Balbi, S., Nolzen, H., Müller, B., Schulze, J., Buchmann, C.M., 2016. Simple or complicated agent-based models? A complicated issue. Environmental Modelling & Software 86, 56–67. https://doi.org/10.1016/j.envsoft.2016.09.006

Sutherland, L.-A., Burton, R.J.F., Ingram, J., Blackstock, K., Slee, B., Gotts, N., 2012. Triggering change: Towards a conceptualisation of major change processes in farm decision-making. Journal of Environmental Management 104, 142–151. https://doi.org/10.1016/j.jenvman.2012.03.013

Swinkels, J.M., Hilkens, A., Zoche-Golob, V., Krömker, V., Buddiger, M., Jansen, J., Lam, T.J.G.M., 2015. Social influences on the duration of antibiotic treatment of clinical mastitis in dairy cows. Journal of Dairy Science 98, 2369–2380. https://doi.org/10.3168/jds.2014-8488

Tago, D., Hammitt, J.K., Thomas, A., Raboisson, D., 2016. The Impact of Farmers' Strategic Behavior on the Spread of Animal Infectious Diseases. PLOS ONE 11, e0157450. https://doi.org/10.1371/journal.pone.0157450

Tiffany, A., Dalziel, B.D., Njenge, H.K., Johnson, G., Ballah, R.N., James, D., Wone, A., Bedford, J., McClelland, A., 2017. Estimating the number of secondary Ebola cases resulting from an unsafe burial and risk factors for transmission during the West Africa Ebola epidemic. PLOS Neglected Tropical Diseases 11, e0005491. https://doi.org/10.1371/journal.pntd.0005491

Toma, L., Low, J.C., Vosough Ahmadi, B., Matthews, L., Stott, A.W., 2015. An analysis of cattle farmers' perceptions of drivers and barriers to on-farm control of Escherichia coli O157. Epidemiology & Infection 143, 2355–2366. https://doi.org/10.1017/S0950268814003045

Uddin, S., Hossain, L., 2011. Time Scale Degree Centrality: A Time-Variant Approach to Degree Centrality Measures, in: 2011 International Conference on Advances in Social Networks Analysis and Mining. Presented at the 2011 International Conference on Advances in Social Networks Analysis and Mining, pp. 520–524. https://doi.org/10.1109/ASONAM.2011.57

USDA, 2014. Brucellosis regionalization risk assessment model: An Epidemiologic Model to Evaluate the Risk of B. abortus Infected Undetected Breeding Cattle Moving out of the Designated Surveillance Areas in Idaho, Montana, and Wyoming.

Valdano, E., Poletto, C., Giovannini, A., Palma, D., Savini, L., Colizza, V., 2015. Predicting Epidemic Risk from Past Temporal Contact Data. PLoS Comput Biol 11, e1004152. https://doi.org/10.1371/journal.pcbi.1004152

Valdes-Donoso, P., VanderWaal, K., Jarvis, L.S., Wayne, S.R., Perez, A.M., 2017. Using Machine Learning to Predict Swine Movements within a Regional Program to Improve Control of Infectious Diseases in the US. Front. Vet. Sci. 4. https://doi.org/10.3389/fvets.2017.00002

Valeeva, N.I., van Asseldonk, M.A.P.M., Backus, G.B.C., 2011. Perceived risk and strategy efficacy as motivators of risk management strategy adoption to prevent animal diseases in pig farming. Preventive Veterinary Medicine 102, 284–295. https://doi.org/10.1016/j.prevetmed.2011.08.005

Van Andel, M., Hollings, T., Bradhurst, R., Robinson, A., Burgman, M., Gates, M.C., Bingham, P., Carpenter, T., 2018. Does Size Matter to Models? Exploring the Effect of Herd Size on Outputs of a Herd-Level Disease Spread Simulator. Front. Vet. Sci. 5. https://doi.org/10.3389/fvets.2018.00078

VanderWaal, K., Morrison, R.B., Neuhauser, C., Vilalta, C., Perez, A.M., 2017. Translating Big Data into Smart Data for Veterinary Epidemiology. Front. Vet. Sci. 4. https://doi.org/10.3389/fvets.2017.00110

Vaughan, T.G., Kühnert, D., Popinga, A., Welch, D., Drummond, A.J., 2014. Efficient Bayesian inference under the structured coalescent. Bioinformatics 30, 2272–2279. https://doi.org/10.1093/bioinformatics/btu201

Verelst, F., Willem, L., Beutels, P., 2016. Behavioural change models for infectious disease transmission: a systematic review (2010–2015). Journal of The Royal Society Interface 13, 20160820. https://doi.org/10.1098/rsif.2016.0820

Vernon, M.C., 2011. Demographics of cattle movements in the United Kingdom. BMC Veterinary Research 7, 31. https://doi.org/10.1186/1746-6148-7-31

Vernon, M.C., Keeling, M.J., 2012. Impact of regulatory perturbations to disease spread through cattle movements in Great Britain. Preventive Veterinary Medicine 105, 110–117. https://doi.org/10.1016/j.prevetmed.2011.12.016

Victory, K., Coronado, F., Ifono, S., Soropogui, T., Dahl, B., 2015. Ebola Transmission Linked to a Single Traditional Funeral Ceremony — Kissidougou, Guinea, December, 2014–January 2015. Morb Mortal Wkly 64, 386–388.

Visschers, V.H.M., Meertens, R.M., Passchier, W.F., DeVries, N.K., 2007. How Does the General Public Evaluate Risk Information? The Impact of Associations with Other Risks. Risk Analysis 27, 715–727. https://doi.org/10.1111/j.1539-6924.2007.00915.x

Volz, E.M., 2012. Complex population dynamics and the coalescent under neutrality. Genetics 190, 187–201. https://doi.org/10.1534/genetics.111.134627

Volz, E.M., Romero-Severson, E., Leitner, T., 2017. Phylodynamic Inference across Epidemic Scales. Mol Biol Evol 34, 1276–1288. https://doi.org/10.1093/molbev/msx077

Webb, T.L., Sheeran, P., 2006. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin 132, 249–268. https://doi.org/10.1037/0033-2909.132.2.249

Wengraf, T., 2001. Qualitative Research Interviewing. SAGE Publications, Ltd, 1 Oliver's Yard, 55 City Road, London England EC1Y 1SP United Kingdom.

Wesolowski, A., Buckee, C.O., Engø-Monsen, K., Metcalf, C.J.E., 2016. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. J Infect Dis 214, S414–S420. https://doi.org/10.1093/infdis/jiw273

Whay, H., 2007. The journey to animal welfare improvement. Animal Welfare 16, 117–122.

Whelan, S., Liò, P., Goldman, N., 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends in Genetics 17, 262–272. https://doi.org/10.1016/S0168-9525(01)02272-7

Widgren, S., Engblom, S., Bauer, P., Frössling, J., Emanuelson, U., Lindberg, A., 2016. Data-driven network modelling of disease transmission using complete population movement data: spread of VTEC O157 in Swedish cattle. Veterinary Research 47, 81. https://doi.org/10.1186/s13567-016-0366-5

Wiethoelter, A.K., Sawford, K., Schembri, N., Taylor, M.R., Dhand, N.K., Moloney, B., Wright, T., Kung, N., Field, H.E., Toribio, J.-A.L.M.L., 2017. "We've learned to live with it"—A qualitative study of Australian horse owners' attitudes, perceptions and practices in response to Hendra virus. Preventive Veterinary Medicine 140, 67–77. https://doi.org/10.1016/j.prevetmed.2017.03.003

Wilkie, R., 2005. Sentient commodities and productive paradoxes: the ambiguous nature of human–livestock relations in Northeast Scotland. Journal of Rural Studies 21, 213–230. https://doi.org/10.1016/j.jrurstud.2004.10.002

Wilson, L., Rhodes, A.P., Dodunski, G., 2015. Parasite management extension – challenging traditional practice through adoption of a systems approach. New Zealand Veterinary Journal 63, 292–300. https://doi.org/10.1080/00480169.2015.1056853

Wood, B.A., Blair, H.T., Gray, D.I., Kemp, P.D., Kenyon, P.R., Morris, S.T., Sewell, A.M., 2014. Agricultural Science in the Wild: A Social Network Analysis of Farmer Knowledge Exchange. PLOS ONE 9, e105203. https://doi.org/10.1371/journal.pone.0105203

Worby, C.J., Lipsitch, M., Hanage, W.P., 2014. Within-Host Bacterial Diversity Hinders Accurate Reconstruction of Transmission Networks from Genomic Distance Data. PLoS Computational Biology 10. https://doi.org/10.1371/journal.pcbi.1003549

World Organisation for Animal Health OIE, 2015. Zoning and compartmentalisation. In: Terrestrial Animal Health Code 2015 [WWW Document]. URL http://www.oie.int/index.php?id=169&L=0&htmfile=chapitre_zoning_compartment.htm (accessed 5.12.16).

Young, I., Rajić, A., Hendrick, S., Parker, S., Sanchez, J., McClure, J.T., McEwen, S.A., 2010. Attitudes towards the Canadian quality milk program and use of good production practices among Canadian dairy producers. Preventive Veterinary Medicine 94, 43–53. https://doi.org/10.1016/j.prevetmed.2009.11.018

Yupiana, Y., de Vlas, S.J., Adnan, N.M., Richardus, J.H., 2010. Risk factors of poultry outbreaks and human cases of H5N1 avian influenza virus infection in West Java Province, Indonesia. International Journal of Infectious Diseases 14, e800–e805. https://doi.org/10.1016/j.ijid.2010.03.014

Zimmerman, F.J., 2013. Habit, custom, and power: A multi-level theory of population health. Social Science & Medicine 80, 47–56. https://doi.org/10.1016/j.socscimed.2012.12.029

Zuckerkandl, E., Pauling, L., 1962. Molecular disease, evolution and genetic heterogeneity, in: Horizons in Biochemistry. Academic Press, pp. 189–225.

Appendix 1

1. Aggregation of repeated movements between the same pair of farms over a short time period

As described in the Section 2.1.1 of the manuscript, there were 2,643 instances where multiple batches of the same animal type were moved between the same source and destination farms over a short time period (≤ 31 days). We aggregated these movements into a single batch as follows. Firstly, for each distinct combination of source farm, destination farm, and age class, a batch that had the largest size was identified. If there were any batches that occurred within 15 days prior to or after the date of this batch, they were all integrated. The movement date for the newly created batch was set to be equivalent to the original date of the largest batch, and the size to be equivalent to the sum of the size of batches that were combined. When more than one batch had the largest size, the batch that had the earliest movement date was chosen. This integration process was repeated until there were no more batches that could be combined.

2. Descriptions of hurdle models

The distances that farmers move cattle are known to be associated with various demographic and herd management factors. In this study, movement distances towards destination farms were predicted by a regression model using characteristics of both the batch and the destination farm as predictors. It was reasonable to assume that short distance and longer distance movements have fundamentally different mechanisms; the latter is likely to be a commercial trade, whereas the former might be a more personal trade such as that between a landowner and a sharemilker, which is a unique feature of pastoral management systems in New Zealand (Blunden et al., 1997). To account for this, a hurdle regression model was employed where short (<10km) and longer (≥10km) distance movements were considered two distinct processes and modelled separately using a logistic regression and a zero-truncated count model (Dohoo et al., 2009). The 10km threshold for short and long distance movements was defined arbitrarily, after considering that the movement distance of more than half of livestock movements between farms in New Zealand was estimated to be less than 10 km (Sanson, 2005).

Table S1 shows the distribution of <10km movements and mean movement distances towards farms in each of 16 regions in New Zealand (Figure S1); both were significantly heterogeneous over regions (chi-squared test for the distribution and Kruskal-Wallis test for mean distances, both p <0.001). This might be partially explained by the substantial heterogeneity in the farm density over regions as well as distance to livestock trading markets.

The Euclidian distance for each observed batch movement was calculated based on geographic coordinates of the source and destination farms. The calculated distances were further categorized so that <10 km movements were coded as 0 and any distances above 10km were then categorised with a 20 km interval (i.e. coded 1 for movements ≥ 10km and < 30km, coded 2 for movements ≥ 30km and < 50km, and so on. The logistic regression part first modelled the probability of the movement being < 10km or ≥ 10km. Conditional to movements being ≥ 10km, the count part modelled the distribution of categorised movement distances.

The herd-level factors included in the regression model were (1) geographical region (16 categories), (2) the farm-level in-degree and out-degree if farms existed before the study period, (i.e. number of farms from which a given farm purchased cattle, and number of farms to which a given farm sold cattle, respectively), (3) number of calves born in the 2009 season, and (4) number of cows culled in the 2009 season. The batch-level factors included (1) month of movement, (2) batch type (calves, heifers, and adults), and (3) batch size.

Given farms that just started their business ('new farms') were assumed to have different trading patterns from farms that had been in existence for a while ('existing farms'), two separate hurdle models were developed for two farm types.

3. Removal of repeated short distance movements from a network rewiring process

As described above, our regression models were developed to predict short and long movement distances. Given that limited explanatory variables are available, these models are not designed to explain the presence of farms that repeatedly received short distance (<10 km) movements. Among 24,530 batches in the dataset, 595 farms received more than one short distance movement. We therefore only included one randomly selected <10km batch per each batch type for each destination farm. This excluded 1,087 batches and remaining 23,443 batches were used in the network rewiring model.

4. Construction of hurdle models

Manual forward selection was performed to identify significant categorical covariates, initially for the logistic part and then for the count part. Those covariates significant at 0.15 level, using a log-likelihood ratio (LR) test, were retained in the model. Continuous variables were subsequently added to the model as fractional polynomial functions using the multivariable fractional polynomial algorithm available in R package mfp, which identifies the most suitable fractional polynomial transformation with a backward elimination procedure at a significance level of 0.05 (Sauerbrei et al., 2006). The transformation was performed by combining numbers of following power terms -3, -2, -1, -0.5, 0, 0.5 1, 2, 3 as well as a log transformation of the original continuous covariate. If the addition of a covariate caused unrealistic inflation in the standard error of at least one of coefficients, this was considered a sign of multicollinearity and the covariate was not included in the model. The count part was initially modelled as a zero-truncated Poisson regression and then replaced by a zero-truncated negative binomial model because LR test showed a significantly better fit of the latter ($p < 0.001$). Interactions terms were assessed only between the batch type and month using LR test if they both remained in the final model; only this interaction, if present, was deemed reasonably interpretable. Evaluations of the constructed models were performed by plotting Pearson residuals against the fitted values and comparing the distributions of model predicted and observed movement frequencies (Dohoo et al., 2009). Constructed final models were used to predict the probabilities of each movement falling into each distance interval. A multinomial trial based on these probabilities was then performed to select a distance for each batch that was in turn used to identify a best outward stub in the network rewiring model as described in the manuscript. Confidence intervals for the predicted conditional distances from the hurdle model were obtained through 1,000 times bootstrap procedures.

5. Hurdle regression results

The scatterplot of Pearson residuals over fitted values shown in Figure S2 suggests that while predictions for long distance movements were generally good, those for short distance movements were underestimated. Nevertheless, the overall predictive abilities of the final hurdle models for both new and existing farms were fairly good as shown in Figure S3.

The adjusted coefficients for explanatory variables that remained in the final multivariable hurdle regression model for new farms were summarised in Table S2. For the count part of the hurdle model, the geographical region of destination farms and the batch type remained in the final model. The month of movement was not identified to be a significant covariate thus interaction terms were not assessed. For the logistic part, the variable batch size, geographical region of destination farms, and batch type remained in the final model. The linear form of size was identified to best fit the data.

Tables S3 and S4 showed the regression estimates for explanatory variables that remained in the count part and the logistic part of the final multivariable hurdle model for existing farms. The number of animals culled in the 2009 season did not remain in the logistic part of the final model. Interaction terms between the batch type and month were statistically significant (p value < 0.001) in both the count and logistic part of the hurdle model for existing farms. To aid the interpretation of these interactions, predicted results were displayed graphically (Figure S4) rather than regression coefficients being shown. The functional forms of each continuous covariate that best fitted the data were also shown in Table S3 and S4. Given that a meaningful interpretation of coefficients of fractional polynomials is not straightforward (Dohoo et al., 2009), the predicted outcomes from the multivariable regression models were also graphically displayed in Figure S5 and S6.

6. Interpretation of hurdle regression results

The hurdle regression models were used solely to better predict the livestock movement distance and the regression results should be interpreted only as preliminary information given the limited number of available explanatory variables. However, we note some important findings from the regression models below.

6.1. Movements towards new farms

The odds of movement distance being ≥10km were 0.24 (95% CI 0.09 – 0.66) times less for Marlborough and 0.44 (95% CI 0.27 – 0.71) times less for West Coast that for Taranaki. This means that the probabilities of new farms in these regions receiving movements from farms <10km away were relatively high, compared with those in Taranaki. However, the count ratios were 3.29 (95% CI 0.98 – 11.08) for Marlborough and 2.92 (95% CI 1.79 – 4.76) for West Coast (Table S2). This suggests that new farms in these two regions tended to purchase cattle from more distant source farms than the farms in Taranaki. Heifers and adults were more likely to be moved distances ≥ 10km than calves (Table S2). If the movements were ≥ 10km, however, the distance for batches with heifers or adults was significantly shorter (count ratio 0.69 for heifer and 0.77 for adult) than that of calves (Table S2). While larger batches were more likely to be moved distances < 10km, the batch size was not associated with distances for movements moved ≥ 10km.

6.2. Movement towards existing farms

During the period between August and November, calves were more likely to be moved < 10km than heifers and adults (Figure S4A). Predicted conditional distances for the same period were also shorter for calves compared to heifers (Figure S4B). Movement distance for adults and calves was similar between July and January then calves were moved for longer distances between February and June. In contrast to new farms, the probabilities of movement towards farms being ≥ 10km were not significantly different for the Marlborough and West Coast regions compared to Taranaki (Table S4). Movement distance to farms in these two regions was significantly longer than that to farms in Taranaki (Table S3). Our results suggest that while new farms in these two regions were more likely to purchase livestock from neighbour farms, existing farms tended to purchase livestock from more distant farms. Given that the Marlborough and West Coast regions are designated as Movement Control Areas (Area 4 in this manuscript), where the risk of bTB transmission from wildlife is high, it is interesting to examine if these observed heterogeneous trading patterns over existing and new farms were associated with the deemed high bTB risk.

Figures S5 and S6 show the predicted probabilities of movements being <10km and the predicted conditional distance for each of continuous covariates after adjusting by other covariates. In general, destination farms with a larger number of calves born, a higher in-degree, and a lower out-degree in the 2009 season had the lower probability of receiving batches from farms located within 10km. Consistent to this trend, movements over 10km to farms with these characteristics had longer distances (Figure S6). Batches towards farms with a higher number of animals culled in the 2009 season also moved longer distances. Potential explanations for these findings are as follows. Farms that culled a greater number of animals and/or a higher in-degree in the 2009 season may have needed to purchase cattle from more distant farms to meet their relatively high number of required replacements. Farms with higher number of calves born were more likely to be large herds, which also may have needed to purchase greater numbers of cattle and hence some of their source farms could be farther away. Another possibility is that these farms might have more economic resources, which allowed them to have options to purchase livestock from more distant farms. A previous study also found that larger farms tended to move cattle over greater distances (Sanson, 2005). It is of importance to understand the underlying socioeconomic factors and farmers' perceptions that are associated with these observations. Movements towards farms with higher out-degree tended to be shorter; we suspect these farms may play a role as traders. Given these farms could be potentially disease super-spreaders, their behaviour should be investigated in depth.

Table S1. Descriptive statistics of cattle batch movements stratified by each categorical covariate representing regions and DCA status of destination farms, an indicator whether or not the farm existed before the study period, and the batch types. Mean distance represents the average distance of batches sent to farms with a given category.

Variable	Level	No. farms[1]	No. batches	No. <10km (%)	Mean distance (km)
Region[2]	Auc	202	662	108 (16.3)	88.5
	Bop	376	1208	253 (20.9)	63.0
	Can	720	3197	415 (13.0)	233.6
	Eac	11	148	0 (0)	194.7
	Hba	67	829	31 (3.7)	261.1
	Man	410	1386	234 (16.9)	90.6
	Mar	29	97	21 (21.6)	181.5
	Nel	3	73	1 (1.4)	297.5
	Nor	395	1301	236 (18.1)	108.9
	Ota	188	608	103 (16.9)	231.9
	Sou	481	1688	226 (13.4)	315.1
	Tar	782	2832	656 (23.2)	47.4
	Gba	66	253	48 (19.0)	172.7
	Wai	1864	9515	1266 (13.3)	87.3
	Wel	75	224	32 (14.3)	111.2
	Wco	161	480	141 (29.4)	111.5
DCA[3]	4	339	1456	213 (14.6)	180.0
	3	954	4215	533 (12.6)	163.9
	2	691	3023	455 (15.1)	167.3
	1b	1275	6227	884 (14.2)	72.2
	1a	2547	9444	1669 (17.7)	134.3
Farm status	New	938	2937	665 (22.6)	124.5
	Existing	4901	21593	3106 (14.4)	131.0

Batch type	Calf	1501	4504	777 (17.3)	159.9
	Heifer	2814	5676	794 (14.0)	158.3
	Adult	4201	14350	2200 (15.3)	109.9

[1] Number of destination farms fell into each covariate category. Each value in Batch type represented the number of destination farms that received at least one given batch type.

[2] 9 farms were unknown for their regions

Auc: Auckland, Bop: Bay of Plenty, Can: Canterbury, Eac: East Coast, Hba: Hawks Bay, Man: Manawatu, Mar: Martinborough, Nel: Nelson, Nor: Northland, Ota: Otago, Sou: Southland, Tar: Taranaki, Gba: Golden Bay, Wai: Waikato, Wel: Wellington, Wco: West Coast

[3] 33 farms and 165 movements sent to these farms were unknown for their DCA status. Area 4: Movement Control Area, Area 3: Special Testing Area Annual, Area 2: Special Testing Area Biannual, Area 1a: Special Testing Area Dairy, Area 1b: Surveillance Area

Table S2. Estimates of count part and logistic part of the hurdle model for new farms. See Table S1 for abbreviations.

			95% CI		
Covariates	Level	Ratio	Lower	Upper	p[1]
Count part[2]					
Region	Tar	Baseline			
	Auc	1.93	1.28	2.92	0.003
	Bop	1.68	1.23	2.30	0.002
	Can	6.82	5.39	8.63	<0.001
	Eac	2.12	0.76	5.87	0.15
	Hba	2.75	1.75	4.31	<0.001
	Man	1.97	1.44	2.70	<0.001
	Mar	3.29	0.98	11.08	0.05
	Nel	3.13	0.50	19.74	0.23
	Nor	2.77	2.03	3.79	<0.001
	Ota	10.07	6.81	14.91	<0.001
	Sou	9.68	7.36	12.74	<0.001
	Gba	4.71	1.40	15.88	0.01
	Wai	1.02	0.82	1.27	0.88
	Wel	2.27	0.82	6.29	0.11
	Wco	2.92	1.79	4.76	<0.001
Batch type	Calf	Baseline			
	Heifer	0.69	0.55	0.87	0.002
	Adult	0.77	0.62	0.96	0.01
Logistic part[3]					

Region	Tar	Baseline			
	Auc	1.09	0.63	1.89	<0.001
	Bop	1.06	0.70	1.60	0.76
	Can	1.90	1.36	2.65	0.79
	Eac	NA[5]	NA[5]	NA[5]	0.98
	Hba	4.66	1.82	11.95	0.001
	Man	1.25	0.81	1.92	0.32
	Mar	0.24	0.09	0.66	0.005
	Nel	NA[5]	NA[5]	NA[5]	0.99
	Nor	1.08	0.72	1.63	0.72
	Ota	2.36	1.31	4.25	0.004
	Sou	2.27	1.50	3.43	<0.001
	Gba	1.43	0.29	7.15	0.66
	Wai	0.97	0.74	1.28	0.84
	Wel	0.64	0.21	1.91	0.43
	Wco	0.44	0.27	0.71	<0.001
Size[4]		0.996	0.995	0.997	<0.001
Batch type	Calf	Baseline			
	Heifer	1.46	1.09	1.96	0.01
	Adult	1.60	1.24	2.06	<0.001

[1] p values derived from Wald test

[2] Ratio indicates the count ratio of the transformed movement distance given the distance is ≥10km

[3] Ratio indicates the odds ratio of the movement being ≥10km vs <10km

[4] Added as a linear term

[5] Estimates unavailable due to the small sample size in the category

Table S3. Estimates of count part of the hurdle model for existing farms. See Table S1 for abbreviations.

Covariates	Level	Ratio	95% CI Lower	95% CI Upper	p[1]
Count part[2]					
Region	Tar	Baseline			
	Auc	2.59	2.22	3.01	<0.001
	Bop	1.61	1.42	1.83	<0.001
	Can	5.92	5.36	6.55	<0.001
	Eac	6.01	4.61	7.84	<0.001
	Hba	5.56	4.82	6.42	<0.001
	Man	2.35	2.09	2.64	<0.001
	Mar	7.01	4.89	10.06	<0.001
	Nel	11.43	7.91	16.50	<0.001
	Nor	3.17	2.81	3.58	<0.001
	Ota	5.71	4.85	6.74	<0.001
	Sou	6.94	6.18	7.79	<0.001
	Gba	4.22	3.35	5.31	<0.001
	Wai	1.55	1.43	1.68	<0.001
	Wel	2.70	2.15	3.40	<0.001
	Wco	3.93	3.26	4.73	<0.001
Size[3]	(size/10)^-1				<0.001
	(size/10)^-0.5				<0.001
Born 2009[3]	log((born_2009+1)/100)				<0.001
	((born_2009+1)/100)^0.5				<0.001
	((born_2009+1)/100)^1.5				0.001
	((born_2009+1)/100)^-0.5				<0.001
Cull 2009[3]	log((cull_2009+1)/100)				0.004
	((cull_2009+1)/100)^-1				<0.001

	((cull_2009+1)/100)^3 * log((cull_2009+1)/100)	<0.001
In-degree 2009[3]	((Indegree_2009 + 1)/10)^-1	<0.001
	((Indegree_2009 + 1)/10)^-0.5	<0.001
Out-degree 2009[3]	((Outdegree_2009 + 1)/10)^1	<0.001
	((Outdegree_2009 + 1)/10)^1 * log((Outdegree_2009 + 1)/10)	<0.001

[1] p values derived from Wald test

[2] Ratio indicates the count ratio of the movement distance given the distance is ≥ 10km

[3] Coefficients were impossible to interpret for fractional polynomials, therefore, only the functional forms of each continuous variable were shown. Each variable represented followings in the 2009 season: born 2009; number of calves born, cull 2009; number of culling, in-degree 2009; number of source farms from which each farm purchased cattle, out-degree 2009; number of destinations farms to which each farm sold cattle.

Table S4. Results from the logistic part of the hurdle regression model for existing farms. See Table S1 and S3 for abbreviations.

			95% CI		
Covariates	Level	Ratio	Lower	Upper	p[1]
Logistic part[2]					
Region	Tar	Baseline			
	Auc	1.73	1.34	2.23	<0.001
	Bop	1.25	1.04	1.50	0.02
	Can	1.81	1.54	2.13	<0.001
	Eac	NA[3]	NA[3]	NA[3]	0.94
	Hba	3.40	2.24	5.19	<0.001
	Man	1.59	1.32	1.92	<0.001
	Mar	1.55	0.82	2.91	0.18
	Nel	6.74	0.92	49.50	0.06
	Nor	1.37	1.13	1.65	0.001
	Ota	1.47	1.14	1.90	0.003
	Sou	1.81	1.49	2.19	<0.001
	Gba	0.92	0.64	1.31	0.63
	Wai	1.63	1.44	1.83	<0.001
	Wel	2.28	1.49	3.48	<0.001
	Wco	0.92	0.71	1.19	0.52
Born 2009[4]	((born_2009 + 1)/100)^0.5				<0.001
	((born_2009 + 1)/100)^0.5 * log((born_2009 + 1)/100)				<0.001
In-degree 2009[4]	((Indegree_2009 + 1)/10)^-0.5				<0.001
	((Indegree_2009 + 1)/10)^-0.5 * log((Indegree_2009 + 1)/10)				<0.001
Out-degree 2009[4]	(Outdegree_2009 + 1)/10				0.006
	((Outdegree_2009 + 1)/10)^2				<0.001

---

[1] p values derived from Wald test

[2] Ratio indicates the odds of movement being ≥ 10km vs < 10km

[3] Estimates unavailable due to the small sample size in the category

[4] Coefficients were impossible to interpret for fractional polynomials, therefore, only the functional forms of each continuous variable were shown.

Table S5. The median movement frequencies within and between DCAs obtained over 1000 simulations of the network rewiring model under the risky scenario. Numbers in the bracket represent the 2.5th and 97.5th percentiles of the simulated distributions.

	Destination DCA				
Source DCA	Area 4	Area 3	Area 2	Area 1b	Area 1a
Area 4	64 (53, 74)	52 (43, 62)	27 (19, 36)	17 (11, 23)	821 (805, 836)
Area 3	135 (119, 152)	415 (388, 442)	189 (168, 210)	317 (291, 342)	2507 (2471, 2545)
Area 2	174 (156, 192)	429 (402, 457)	364 (339, 390)	332 (306, 357)	1409 (1374, 1447)
Area 1b	291 (273, 312)	1276 (1240, 1312)	806 (774, 837)	3613 (3558, 3667)	862 (819, 904)
Area 1a	706 (680, 733)	1875 (1832, 1919)	1476 (1439, 1513)	1631 (1580, 1676)	3366 (3316, 3415)

Table S6. The median movement frequencies within and between DCAs obtained over 1000 simulations of the network rewiring model under the safe scenario. Numbers in the bracket represent the 2.5th and 97.5th percentiles of the simulated distributions.

	Destination DCA				
Source DCA	Area 4	Area 3	Area 2	Area 1b	Area 1a
Area 4	339 (321, 357)	256 (236, 275)	213 (194, 234)	129 (115, 145)	16 (8, 26)
Area 3	349 (328, 373)	1139 (1097, 1177)	701 (667, 735)	1153 (1116, 1192)	191 (161, 221)
Area 2	239 (219, 262)	718 (686, 753)	688 (655, 719)	687 (650, 717)	362 (336, 391)
Area 1b	202 (183, 220)	953 (916, 991)	568 (539, 600)	2912 (2860, 2964)	2217 (2172, 2269)
Area 1a	244 (221, 268)	981 (940, 1020)	691 (656, 725)	1024 (978, 1068)	6176 (6122, 6230)

Figure S1. Map showing 16 regions in New Zealand. See Table S1 for the abbreviations for region name.

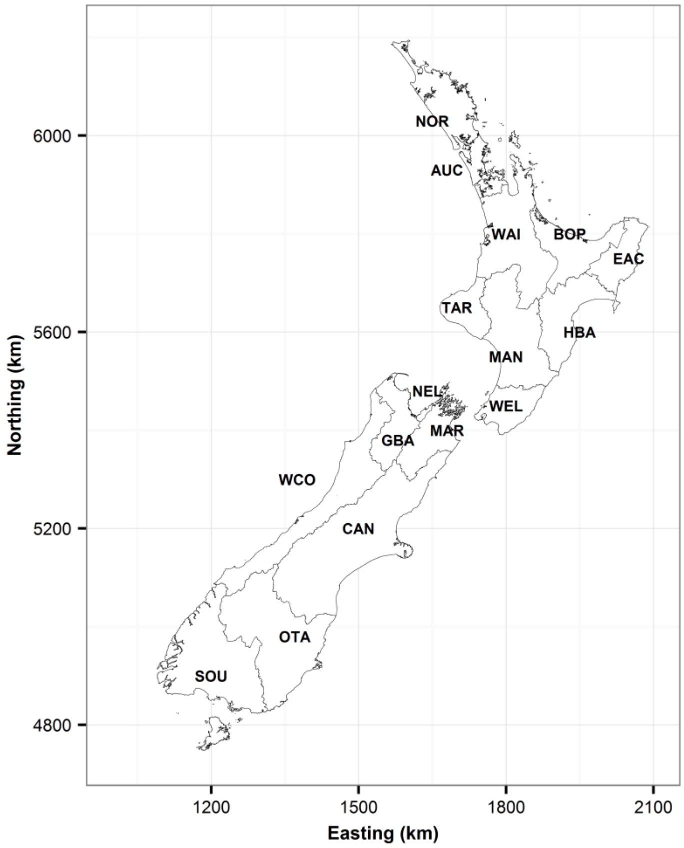
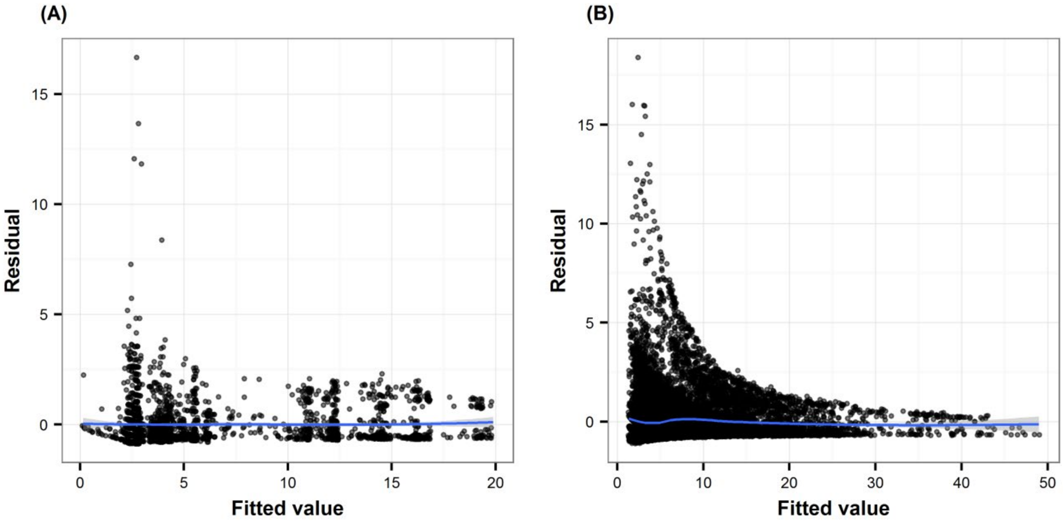


Figure S2. Scatter plots of Pearson residuals and fitted values obtained from the final hurdle regression model for (A) new farms, and (B) existing farms. Locally weighted scatter plot smoothing lines shown in blue indicated roughly equal variance of residuals throughout fitted values. Note that the unit of x axis was the categorised distance.



0

Figure S3. Distributions of the observed and predicted distance obtained from the final hurdle model for (A) new farms, and (B) existing farms. Red and blue lines represent the observed and predicted distance distributions, respectively. Note that the unit of x axis is the categorised distance.

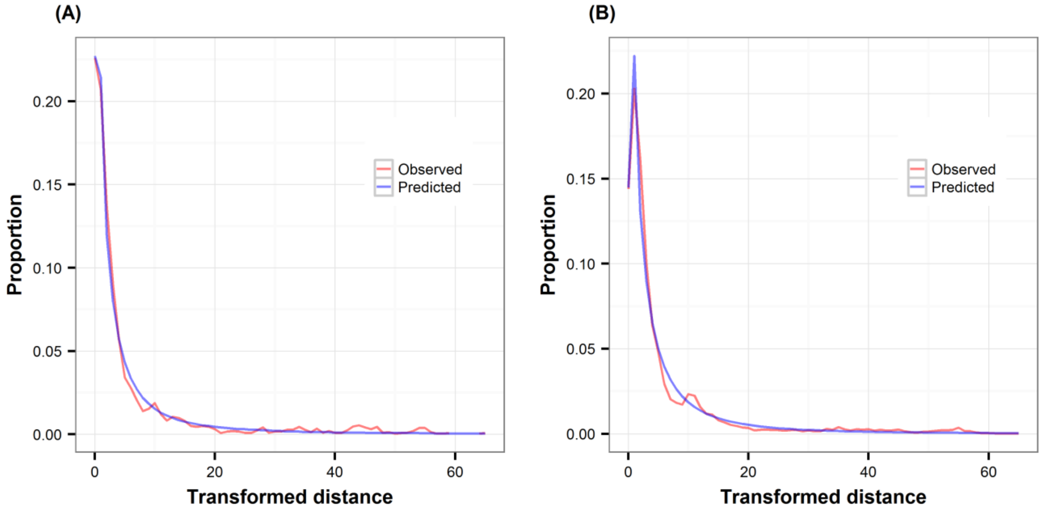


Figure S4. Predictions from the final multivariable hurdle model for existing farms over different combinations of the batch type and month. As an example, predictions were made for batches with median size (= 6) towards existing farms in Taranaki that had median numbers for each of in-degree 2009 (= 2), out-degree 2009 (= 2), cull 2009 (= 39), and born 2009 (= 59). (A) Predicted probabilities and their 95% confidence intervals of movements being <10km for each combination of the batch type and month. (B) Predicted conditional distances and their bootstrap 95% confidence intervals for movements for each combination of the batch type and month. Conditional distance is the predicted mean from the zero-truncated negative binomial model given the distance was ≥ 10km (i.e. predictions were independent from the logistic part). Lines indicate calf (red), heifer (green), and adult (red) batch. X-axis indicates the month of a given movement.

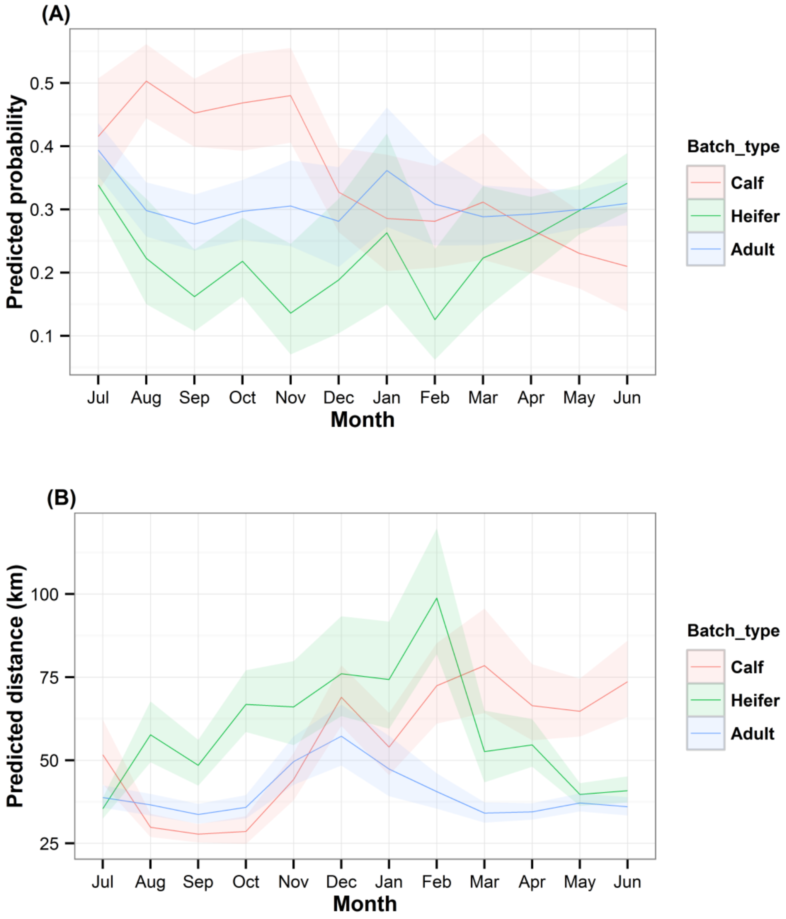


Figure S5. Predicted probabilities and their 95% confidence intervals of movements towards existing farms being <10km as a function of fractional polynomials of each of three continuous covariates that remained in the logistic part of the final multivariable model. Identified sets of fractional polynomials can be found in Table S4. As an example, probabilities were calculated for batches carrying adult that occurred in April towards existing farms in Taranaki that had median values for continuous covariates (i.e., born 2009 = 59, in-degree 2009 = 2, out-degree 2009 = 2) other than the covariate that was used for x-axis in each plot. In each plot y-axis represents the predicted probability as a function of x-axis that represents (A) born 2009, (B) in-degree 2009, and (C) out-degree 2009.

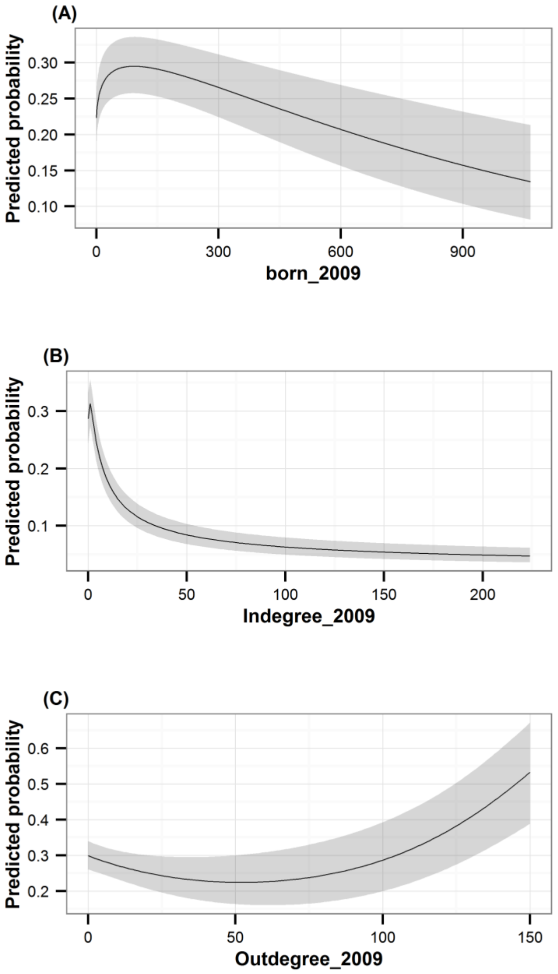


Figure S6. Predicted conditional distances and their bootstrap 95% confidence intervals for movements as a function of fractional polynomials of each of five continuous covariates that remained in the count part of the final multivariable model. Identified sets of fractional polynomials can be found in Table S3. As an example, predicted distances were calculated for batches carrying adult that occurred in April towards existing farms in Taranaki that had median values for continuous covariates (i.e., size=6, born 2009 = 59, cull 2009 = 39, in-degree 2009 = 2, out-degree 2009 = 2) other than the covariate that was used for x-axis in each plot. In each plot y-axis represents the predicted unconditional distance as a function of x-axis that represents (A) size, (B) born 2009, (C) cull 2009, (D) in-degree 2009, and (E) out-degree 2009. Conditional distance is the predicted mean from the zero-truncated negative binomial model given the distance was ≥ 10km (i.e. predictions were independent from the logistic part).

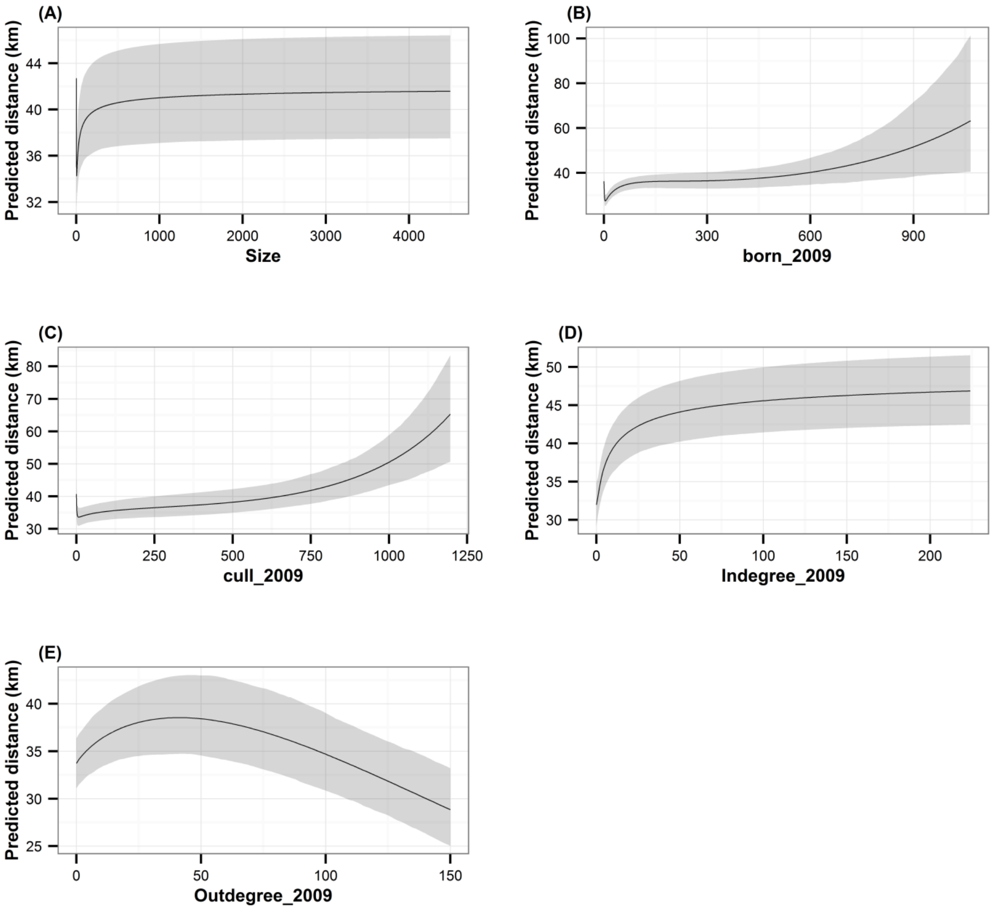


Figure S7. Maps showing monthly dairy cattle movement patterns between July 2010 and June 2011. Batch movements were aggregated over 16 regions, whose boundaries are shown by white lines and locations are indicated by circles. Lines represented movements from North towards South in direction (red) and the opposite (blue). Line widths represented the frequency of batch between a given pair of regions. Circle size represented the total number of batches received in each region in a given month. Circle colour represented the proportion of within-region movements to the total movements received in a given month.

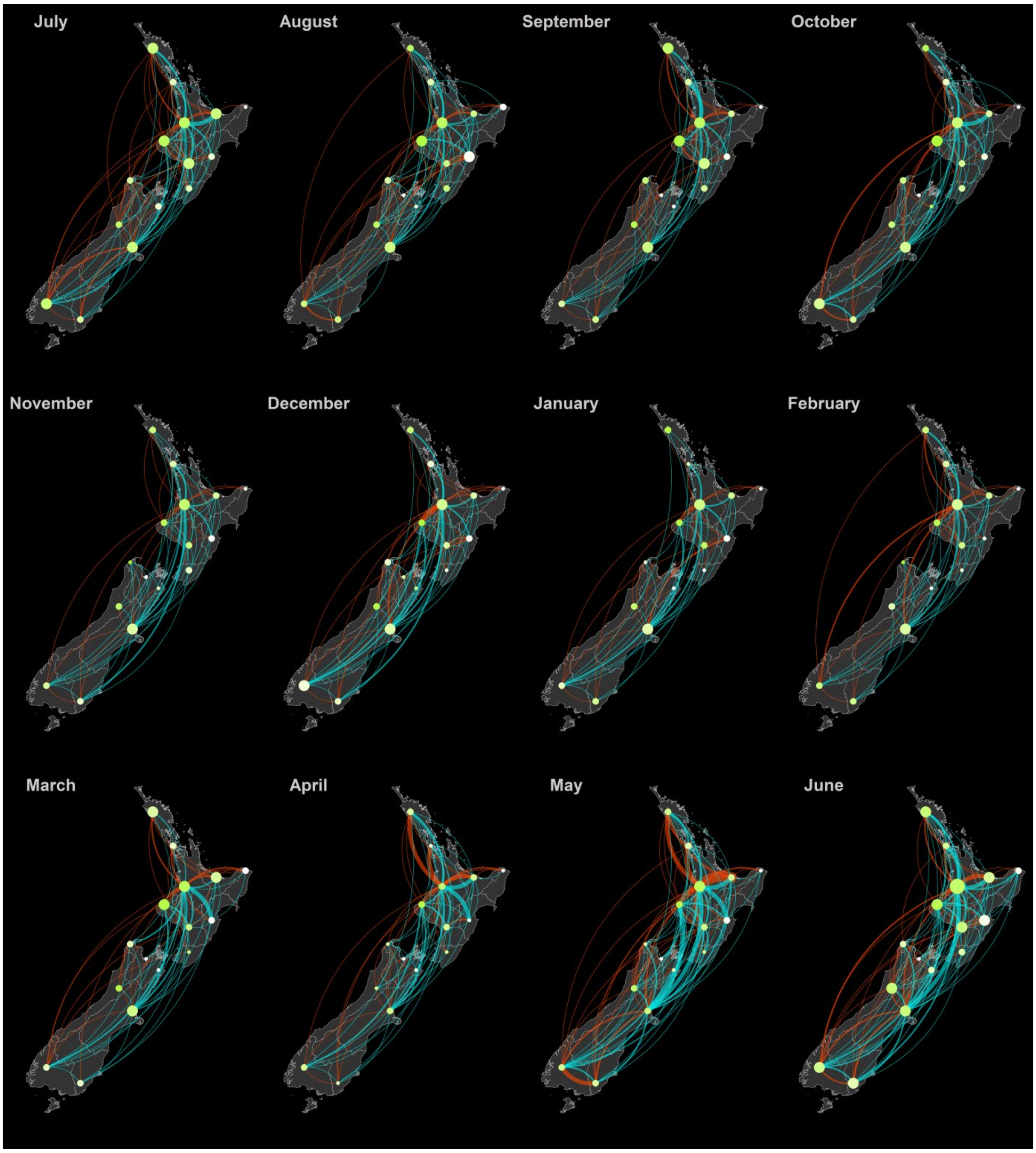


Figure S8. Simulated and observed frequencies of movements between each of 25 combinations of source and destination DCA status. DCA status of source farms was (A) Area 4, (B) Area 3, (C) Area 2, (D) Area 1b, and (E) Area 1a. Numbers shown in the strip of each plot indicate the DCA status of destination farms. Vertical dashed lines indicate the observed movement frequencies for each DCA status combination. Boxplots summarised (on x-axis) the 2.5th, 25th, 50th, 75th, and 97.5th percentile of the movement frequencies obtained from 1000 simulations of the network rewiring model under the random selection scenario (annotated as "Ra"), risky scenario ("Ri"), and safe scenario ("S"). Plots are highlighted in blue when the observed frequencies were lower than the 2.5th percentile of the simulated frequencies under the random scenario and movement were from the higher bTB risk area to the lower risk area.

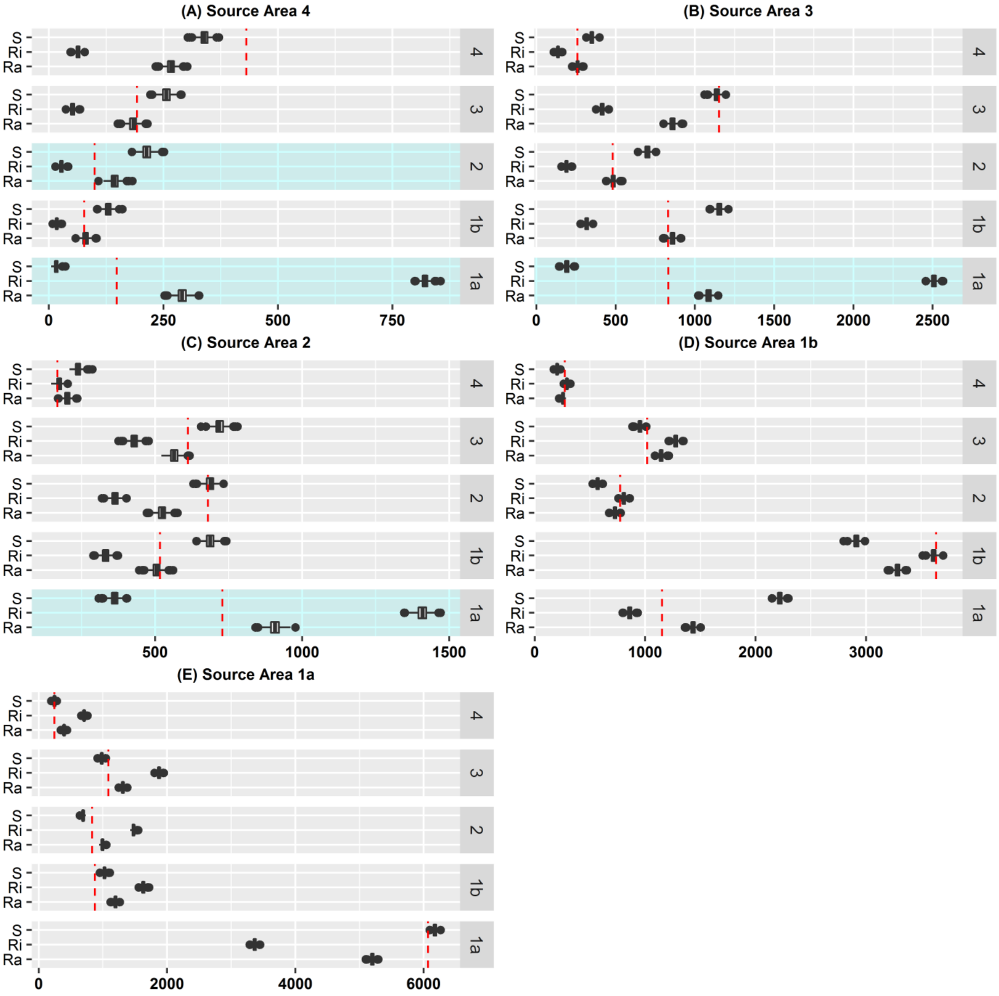


Figure S9. Distributions of movement distances simulated from the network rewiring model under the random selection scenario (black), safe scenario (blue), and risky scenario (red), and the observed distribution (dashed grey). Proportions indicate the median value calculated over 1000 simulations for each of three scenarios.

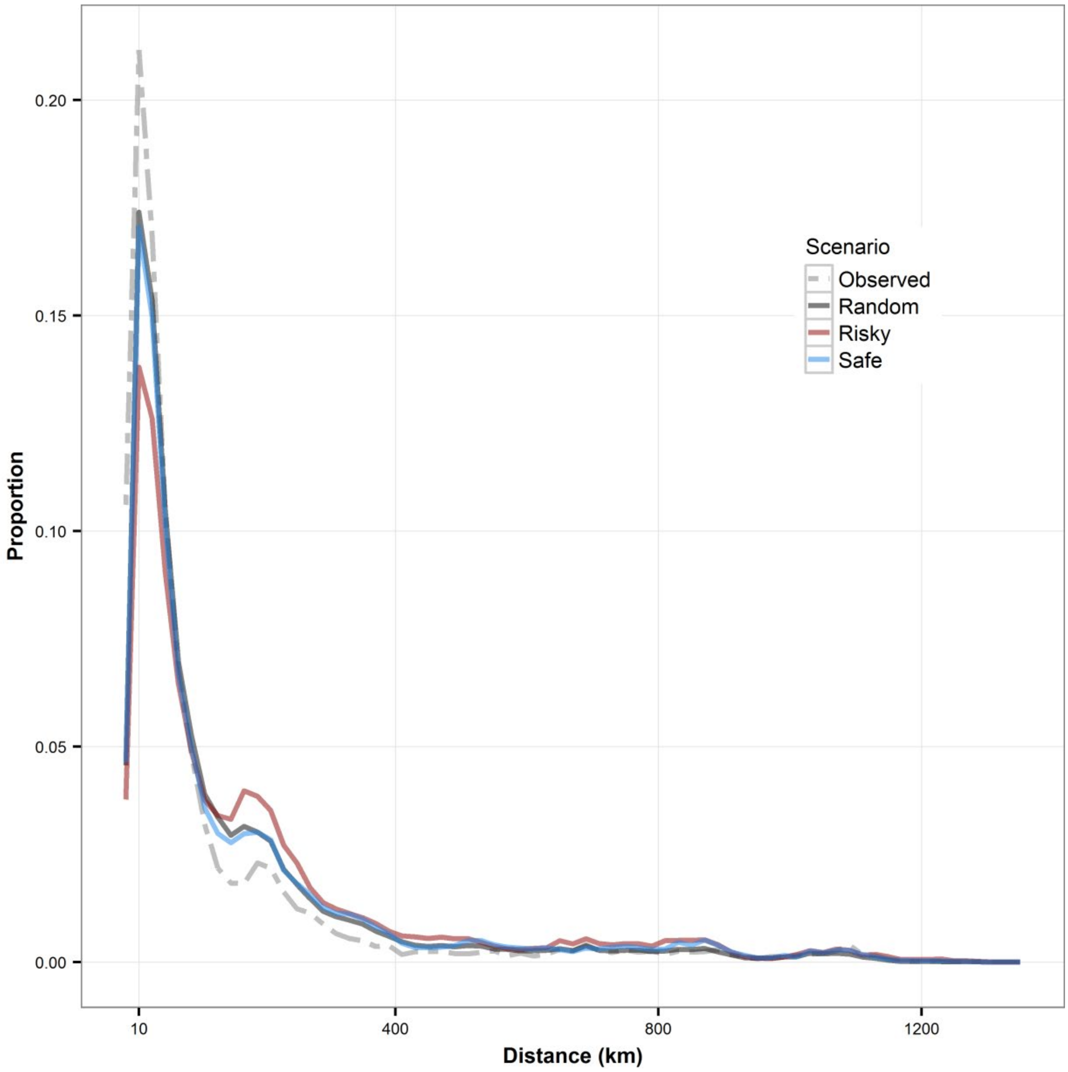


Figure S10. Boxplots showing the results of sensitivity analyses after removing ≤20km movements in the network rewiring model. Boxplots annotated as "Ra" represent results from the random selection scenario without removing ≤20km movements (i.e. same results annotated as "Ra" in Figure S8), whereas those annotated as "Se" represent results after removing ≤20km movements. Red and blue vertical dashed lines represent the observed movement frequency without and with removing ≤20km movements, respectively.

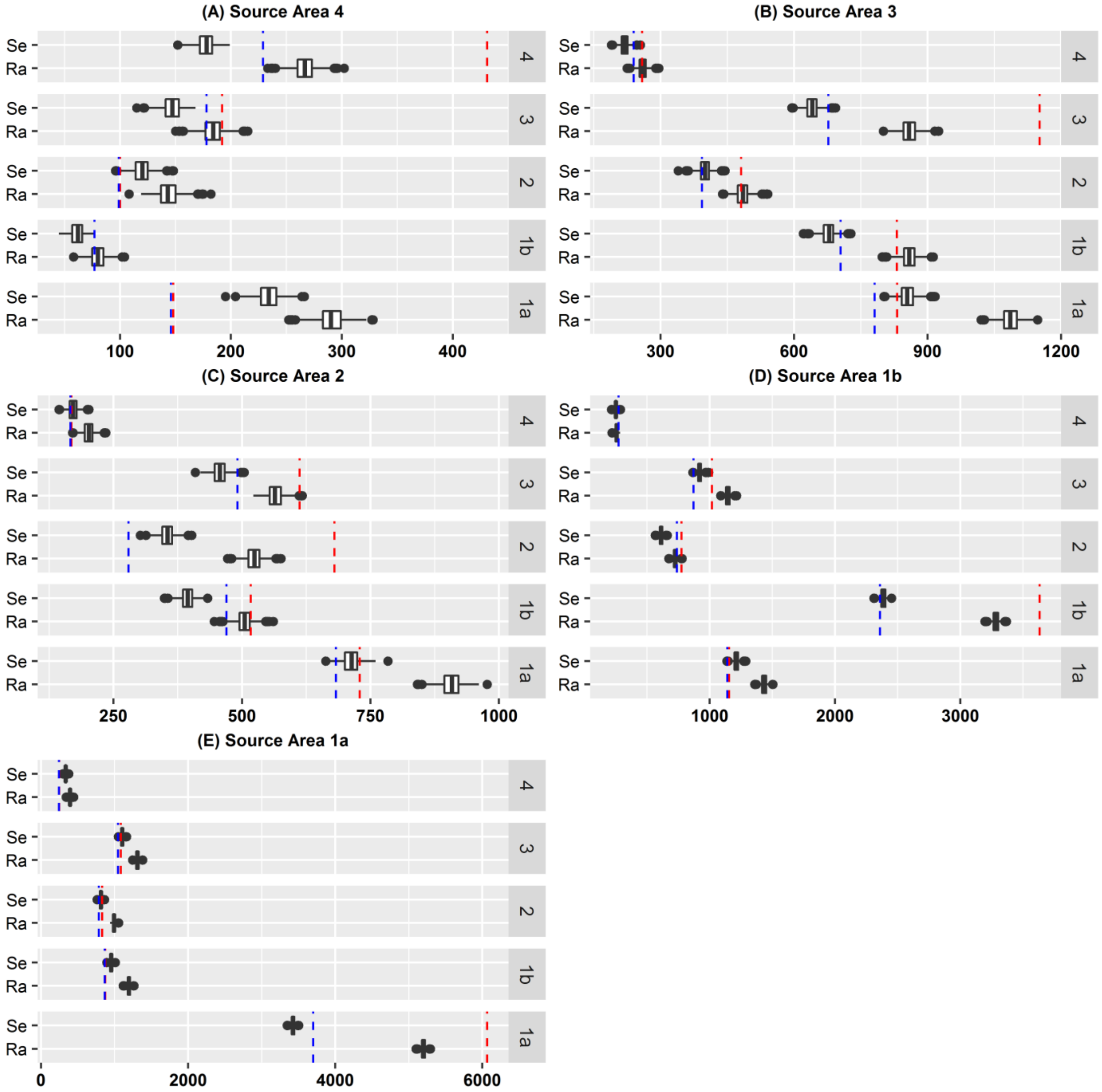
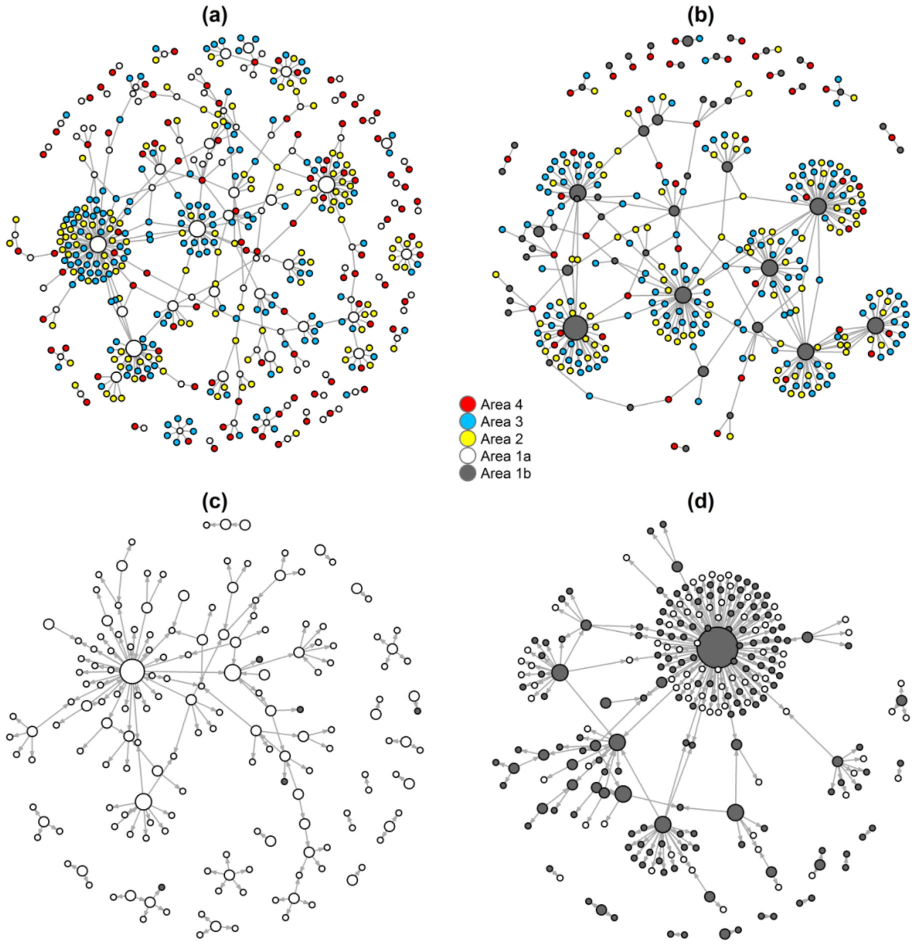


Figure S11. Network structures of batch movements in relation to DCA status. Presented structures are the subset of the whole network which only shows destination farms in Area 1a or 1b that purchased at least 1 batch from Area 4. **Top figures:** Networks showing on-farm movements to farms in Area (a) 1a and (b) 1b which purchased at least 1 batch from farms in Area 4. Destination farms in Area 1a and 1b are shown in white and grey, respectively. Circles in other colours represent source farms in Area 4, 3, or 2. Source farms in Area 1a or 1b are not shown. The circle size of destination farms indicates their total in-degree during the study period. **Bottom figures:** Networks showing off-farm movements from farms in Area (c) 1a and (d) 1b which purchased at least 1 batch from farms in Area 4. Arrows indicate the direction of movements. Circle colours represent farm's DCA status. Movements sold to farms in Area 1a or 1b are only shown. The circle size of source farms indicates their total out-degree during the study period.



Appendix 2

Table S1. Coefficients of the binary part of the final Bayesian mixed-effect zero-inflated negative binomial model used in Analysis 1. Note the coefficients represents the log-odds of the probability of not selling any cows.

			95% CI		
Variable	Form	Coefficient	Lower	Upper	ESS
Intercept		4.85	3.98	5.74	2334
Region[1]					
Bay of Plenty		0.19	-0.18	0.55	867
Canterbury		0.62	0.19	1.04	929
East Coast		NA[2]			
Hawks Bay		1.66	0.03	3.62	4000
Manawatu		0.27	-0.1	0.67	776
Marlborough		0.57	-0.26	1.4	2504
Nelson		NA[2]			
Northland		0.34	-0.04	0.73	1023
Otago		0.94	0.39	1.48	1646
Southland		0.88	0.4	1.35	1321
Taranaki		-0.01	-0.35	0.33	786
Gisborne		0.08	-0.5	0.66	1463
Waikato		-0.24	-0.56	0.09	755
Wellington		0.23	-0.33	0.78	1319
West Coast		0.61	0.12	1.09	1164
Average milk	(X/10)^-2	1.04	0.69	1.42	4000
Number of death	log(((X + 1)/10))^2	0.12	0.03	0.21	2527
	log(((X + 1)/10))	0.37	0.13	0.61	2539
Number of sold	(((X + 1)/10)^-1)	0.09	0.07	0.1	1307
Number of milking cows	((X/100)^-0.5)	2.3	1.79	2.85	2484
Number of culled	(((X + 1)/100)^-0.5)	0.14	0.03	0.26	2509
	log(((X + 1)/100))	0.24	0.04	0.44	2589
In-degree	(X+1)	-0.09	-0.13	-0.05	4000
Out-degree	(X+1)^-0.5	1.35	1.01	1.68	4000
Business in years	log((X/10))	-14	-17.1	-11.1	2896
	((X/10)^0.5)	36.4	29.6	43.6	2869
Milk price	((X/10)^3)	-0.78	-1.42	-0.15	3286

	((X/10)^3 * log((X/10)))	22.2	16.2	28.3	2566
Milk fat percentage	((X/10)^1)	-2.7	-4	-1.5	3494
SCC	((X/100)^1)	0.19	0.12	0.26	4000
SD(random effect)		0.75	0.6	0.91	561

[1] Baseline is Auckland.

[2] These regions were removed because the number of farms was very small.

Table S2. Coefficients of the count part of the final Bayesian mixed-effect zero-inflated negative binomial model used in Analysis 1.

			95% CI		
Variable	Form	Coefficient	Lower	Upper	ESS
Intercept		4.85	3.98	5.74	2334
Region[1]					
Bay of Plenty		-0.3	-0.51	-0.09	814
Canterbury		-0.23	-0.47	0.02	913
East Coast		NA[2]			
Hawks Bay		0.16	-1.3	1.85	4000
Manawatu		-0.31	-0.53	-0.09	878
Marlborough		-0.08	-0.66	0.5	1853
Nelson		-0.21	-1.52	1.33	4000
Northland		-0.26	-0.5	-0.04	850
Otago		-0.05	-0.38	0.28	1306
Southland		-0.32	-0.62	-0.04	1164
Taranaki		-0.57	-0.77	-0.38	778
Gisborne		-0.08	-0.42	0.24	1167
Waikato		-0.43	-0.62	-0.25	708
Wellington		-0.37	-0.68	-0.05	1461
West Coast		-0.2	-0.48	0.09	1176
Number of milking cows	log((X/100))	0.97	0.88	1.06	2215
Number of culled	log(((X+ 1)/100))	-0.03	-0.08	0.01	2664
In-degree	((X+1)^-0.5)	-1.85	-2.52	-1.18	2382
	((X + 1)^-0.5 * log((X + 1)))	-1.08	-1.57	-0.6	2469
Out-degree	(X+1)^-1	-0.14	-0.25	-0.03	4000
Business in years	((X/10)^1)	-0.82	-1.08	-0.56	4000

Milk price	((X/10)^-2)	-0.46	-0.84	-0.09	2239
	((X/10)^-2 * log((X/10)))	-0.31	-0.67	0.03	2221
Milk fat percentage	((X/10)^3)	0.8	0.03	1.54	3315
SCC	((X/100)^-0.5)	0.42	0.19	0.65	4000
SD(random effect)		0.41	0.36	0.46	853

[1] Baseline is Auckland.

[2] These regions were removed because the number of farms was very small.

Table S3. Coefficients of the final mixed-effect logistic regression for Analysis 3.

Variable	Form	Coefficient	SE	p-value
Intercept		45.3	3.48	<0.001
Frequency of being sold	((X+1)^-2)	-1.5	0.04	<0.001
Days animal spent in the herd	((X + 1)/1000)^1)	0.77	0.06	<0.001
	((X + 1)/1000)^0.5)	-1.35	0.15	<0.001
Age in years	((X/10)^-0.5)	-8.7	0.19	<0.001
	log((X/10))	-9.7	0.17	<0.001
SCC ranking	(X^3)	-1.1	0.02	<0.001
	(X^3 * log(X))	-0.75	0.15	<0.001
Milk volume ranking	(X^-1)	0.01	0.0004	<0.001
	(X^-0.5)	-0.3	0.009	<0.001
Milk fat ranking	(log X)	0.06	0.008	<0.001
	(X^3)	0.1	0.04	<0.001
Milk protein ranking	(X^3)	0.23	0.03	<0.001
Interval calving and herd test	((X+337)/1000^0.5)	-35.6	3.17	<0.001
	((X+337)/1000^0.5 * log(X+337)/1000)	35.2	2.65	<0.001
SD(random effect)		4.18		

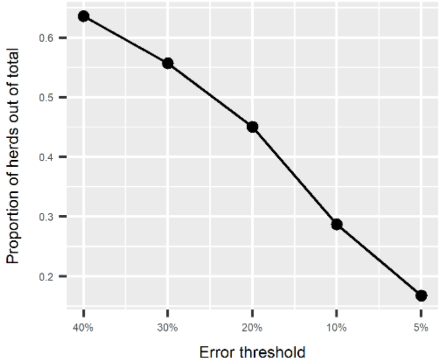


Figure S1. Proportion of eligible herds at each error threshold level from 40% to 5%.

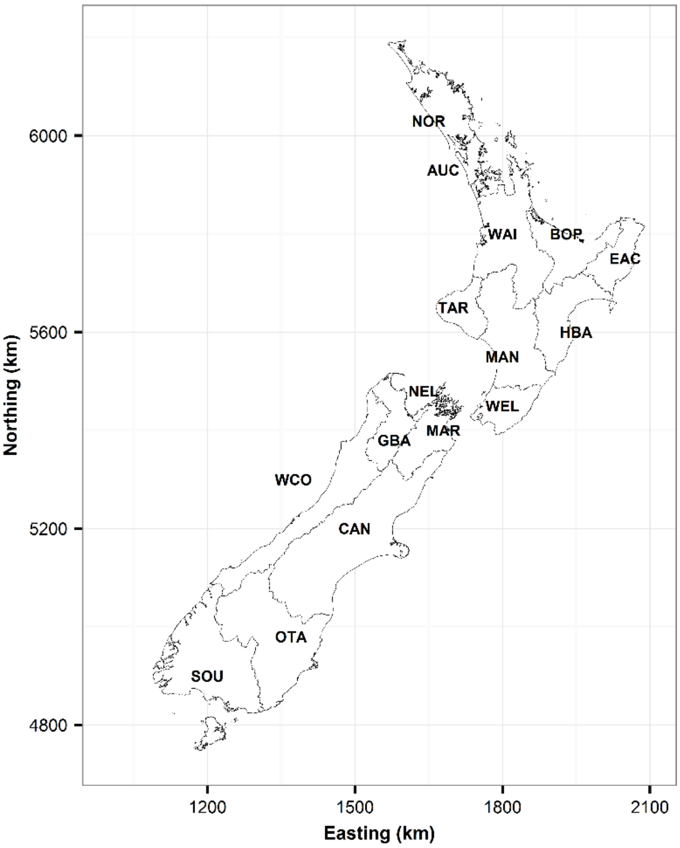


Figure S2. Locations and names of 16 geographical areas in New Zealand that were used in this analysis.

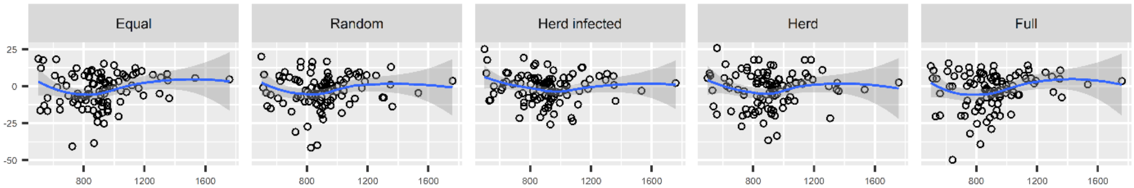
Appendix 3

Table S1. Results of final generalised additive models for each of three phylodynamic inferences in each sampling strategies.

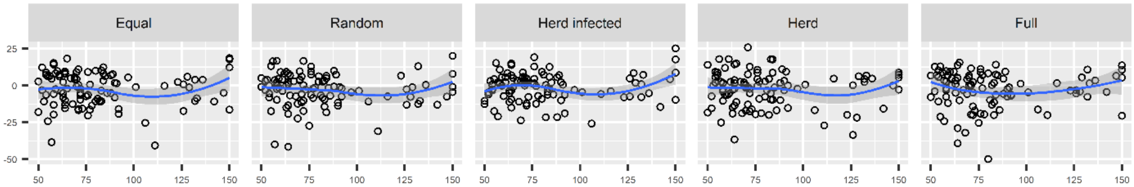
	Percent bias	Percent error	HPD size
Equal	P community (Fig S2A)	P infected -0.31 (0.008)	Tlast -0.0006 (<0.001)
Random	NA	NA	P region (Fig S2C) Tlast -0.0005 (<0.001)
Herd infected	P subsample (Fig S2B)	P subsample 0.25 (0.02)	Tlast -0.0004 (<0.001)
Herd	P community -0.58 (0.04)	Tlast -0.00007 (0.04)	Tlast -0.0005 (<0.001)
Full	P community -0.71 (0.01)	Tlast -0.0001 (0.02)	Tlast -0.0006 (<0.001)

Abbreviations: P community; proportion of maximum sample size collected from one community, P subsample; proportion of samples that were subsampled, P infected; proportion of infected farms that were sampled, T last; the last sampling date, P region; proportion of maximum sample size collected from one region, NA; there was no significant factors. Numbers after variable names indicate their coefficients and those in brackets p-value; these variables were identified to have a liner association with the outcome. Variables that do not show coefficients were those identified to have more complex associations, which were shown in corresponding figures.

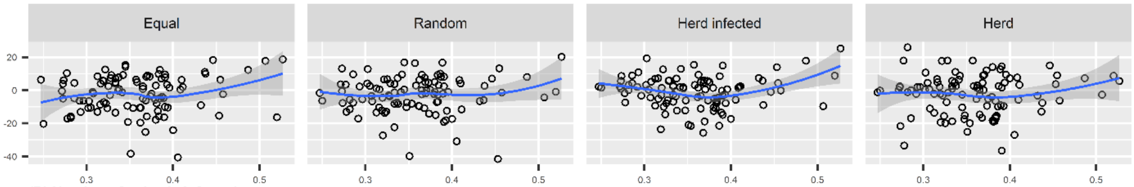
(A) Last sampling day



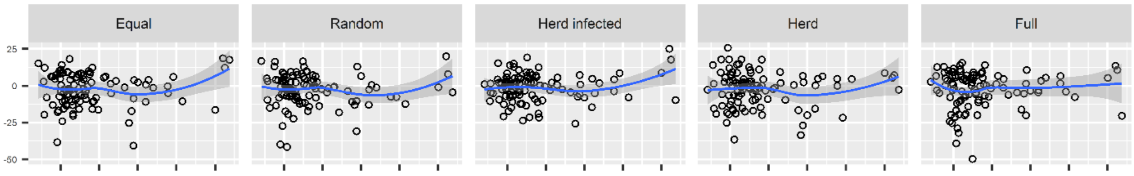
(B) Full sample size



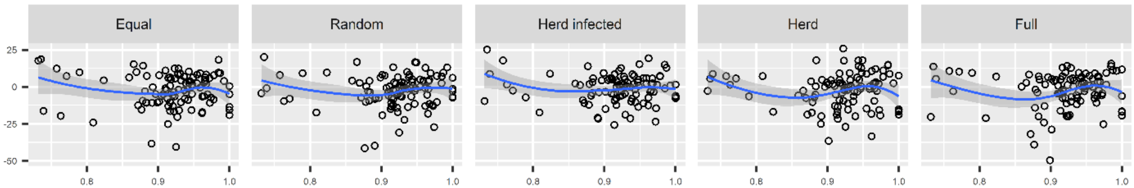
(C) Proportion of sample subsampled



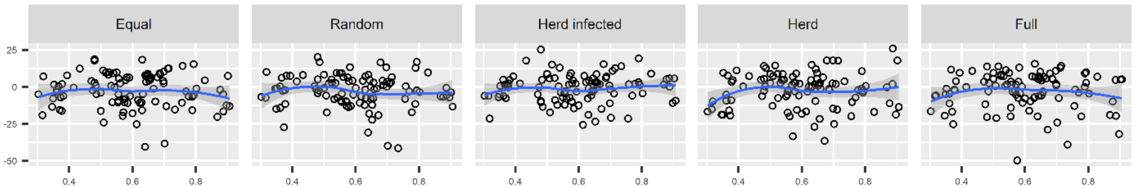
(D) Number of animals infected



(E) Proportion of infected farms sampled



(F) Proportion of max sample size per region



(G) Proportion of max sample size per community

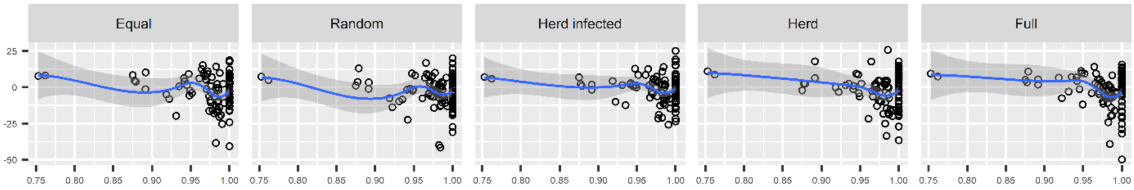


Figure S1. Scatter plots showing the associations between percent bias and each simulation characteristic. Y-axis shows percent bias. X-axis shows simulation characteristics including (A) the last day of sampling, (B) the total number of genetic sequences sampled, (C) the proportion of genetic sequences subsampled out of all samples, (D) the number of animals infected, (E) the proportion of infected farms sampled to the total number of infected farms, (F) the proportion of maximum sample size per geographical region to the total sample size, and (G) the proportion of maximum sample size per community to the total sample size. Blue line and grey areas represent the non-parametric loess smooth curves and their 95% confidence intervals.

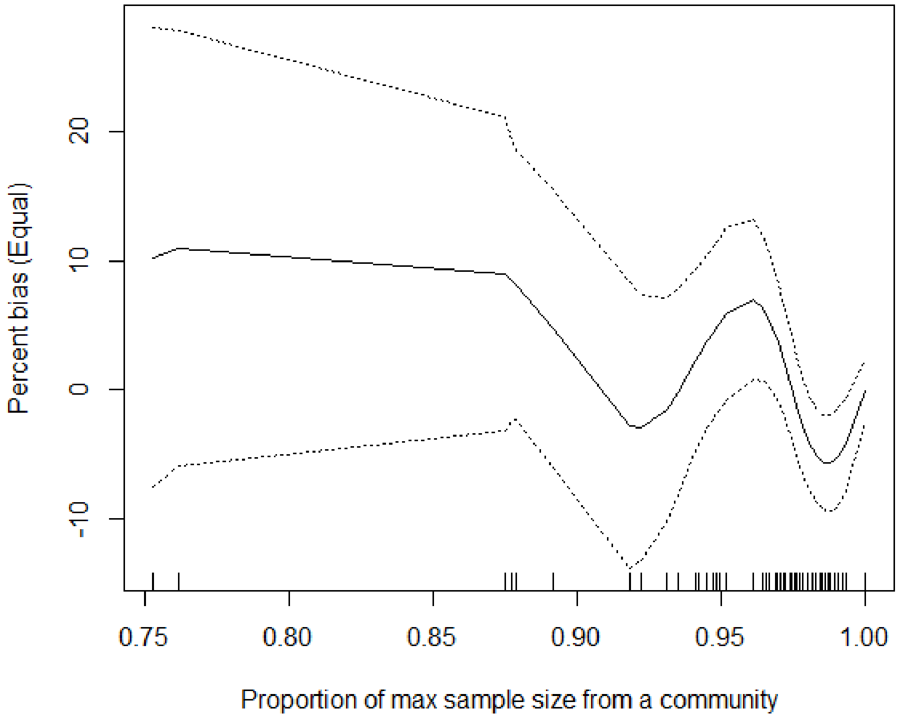


Figure S2A. Predicted association between percent bias and proportion of maximum sample size from one community to the total sample size in 'equal' scenario.

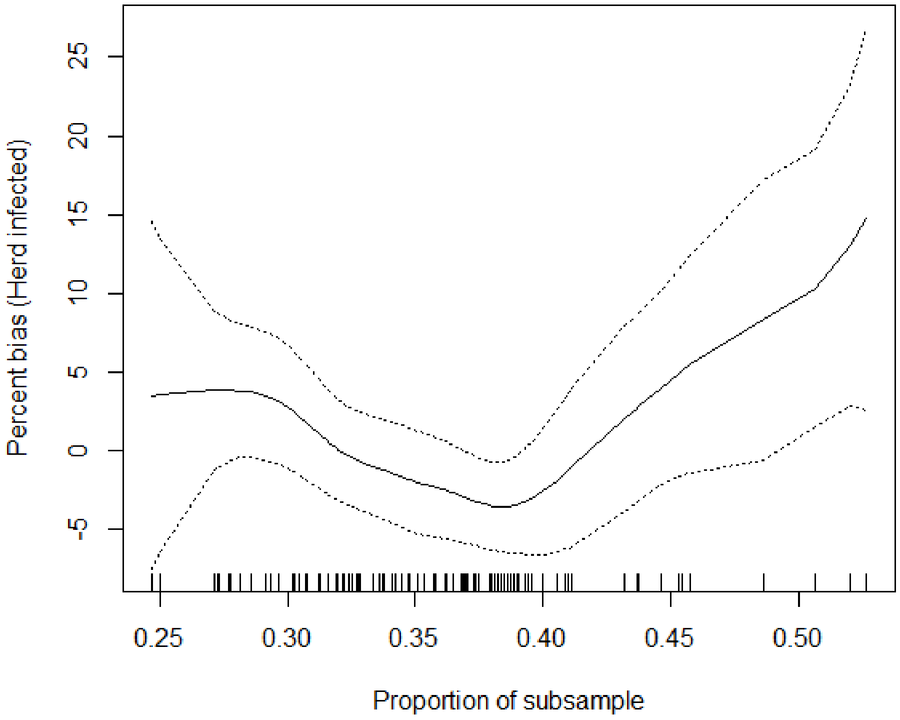


Figure S2B. Predicted association between percent bias and proportion of the number of genetic sequences subsampled to the total number of genetic sequences sampled in 'herd infected' scenario.

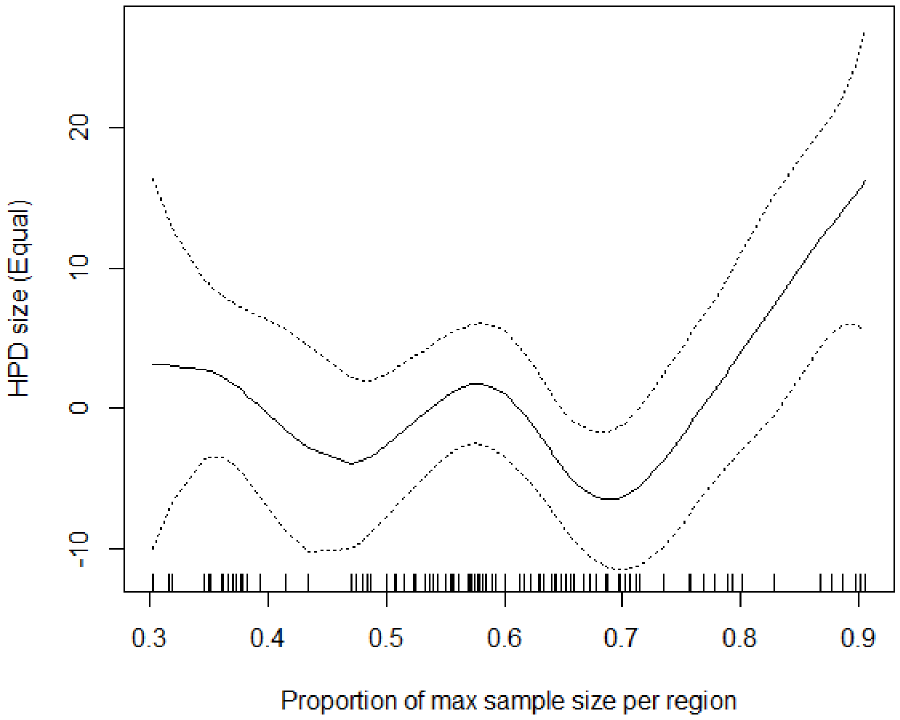


Figure S2C. Predicted association between HPD size and proportion of max sample size per region to the total number of genetic sequences sampled in ‘herd infected’ scenario.

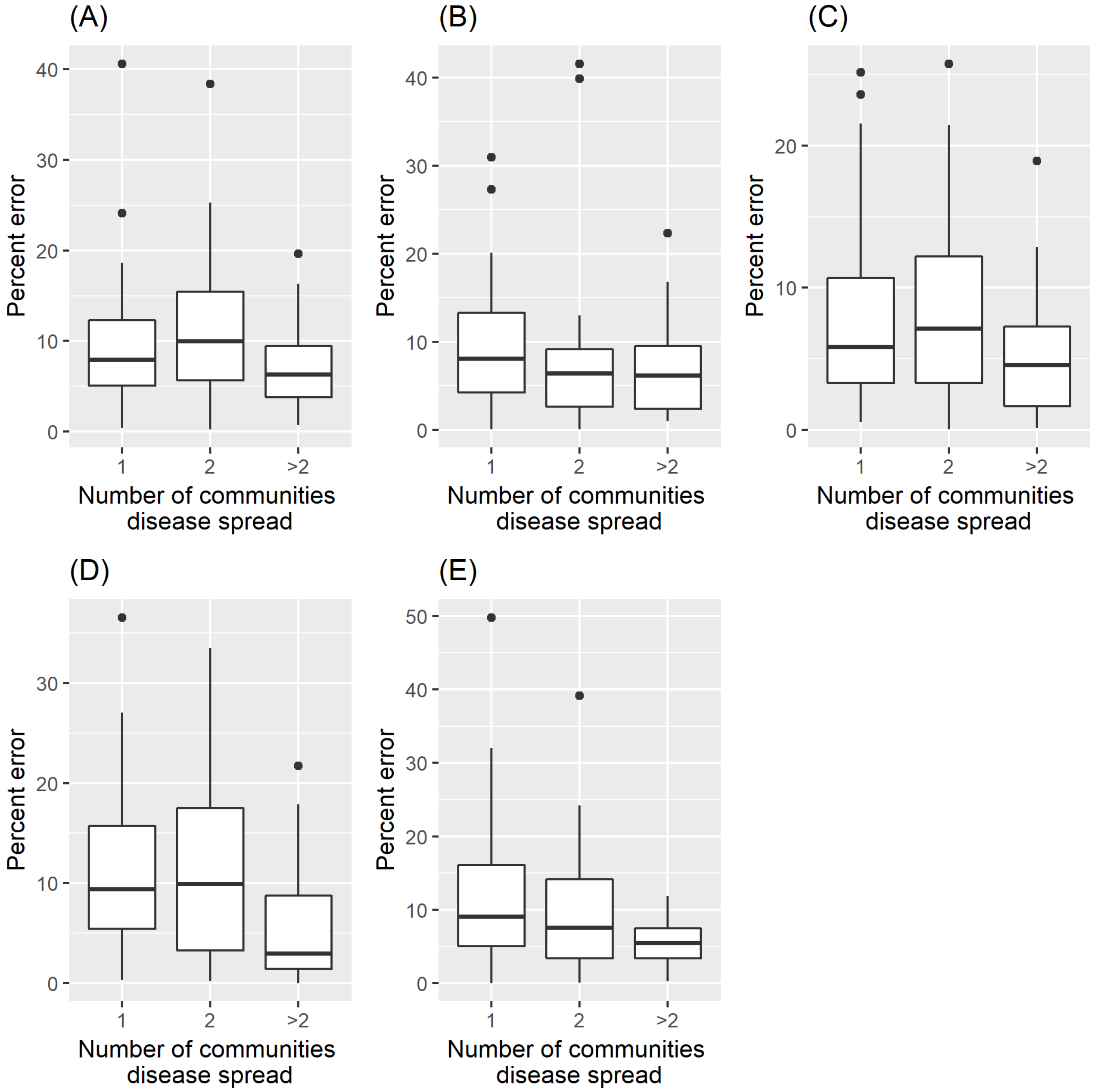


Figure S3. Boxplot showing the association between percent error and number of communities that disease spread. Each panel shows percent error in (A) equal, (B) random, (C) herd infected, (D) herd, and (E) full strategies.

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Arata Hidano

Name/Title of Principal Supervisor: Dr. Carolyn Gates

Name of Published Research Output and full reference:

Modelling dynamic human behavioural changes in animal disease models: challenges and opportunities for addressing bias

In which Chapter is the Published Work: Chapter 2

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 85%

  and / or

- Describe the contribution that the candidate has made to the Published Work:

Arata Hidano Digitally signed by Arata Hidano Date: 2018.07.12 14:54:33 +12'00'

Candidate's Signature

12/07/2018

Date

Principal Supervisor's signature

12/07/2018

Date

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Arata Hidano

Name/Title of Principal Supervisor: Dr. Carolyn Gates

Name of Published Research Output and full reference:

Evaluating the efficacy of regionalisation in limiting high-risk livestock trade movements

In which Chapter is the Published Work: Chapter 3

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate 85%

  and / or

- Describe the contribution that the candidate has made to the Published Work:

Arata Hidano Digitally signed by Arata Hidano Date: 2018.07.12 14:56:02 +12'00'

Candidate's Signature

12/07/2018

Date

Principal Supervisor's signature

12/07/2018

Date