

Collection of empirical equine contact network data to quantify the effect of non-homogenous mixing patterns on disease dynamics in Ontario

by

Rachael Michal Milwid

A Thesis

presented to

The University of Guelph

In partial fulfilment of requirements
for the degree of

Doctor of Philosophy

in

Population Medicine

Guelph, Ontario, Canada

© Rachael Michal Milwid, August 2018

ABSTRACT

COLLECTION OF EMPIRICAL EQUINE CONTACT NETWORK DATA TO QUANTIFY THE EFFECT OF NON-HOMOGENOUS MIXING PATTERNS ON DISEASE DYNAMICS IN ONTARIO

Rachael Michal Milwid
University of Guelph, 2018

Advisor(s):
Dr. Amy Greer
Dr. Terri O'Sullivan

This thesis assessed the suitability of using homogeneous mixing to describe the contact structure of different equine populations in Canada, and the associated effects on disease dynamics within the respective populations. This was accomplished by the development of an appropriate contact data collection method for animal agricultural settings, the collection of equine contact data, and the characterization of the equine contact networks and the corresponding disease spread dynamics. Traditional radio-frequency identification (RFID) tags were modified to enable data storage on the tags' flash memory. The modified tags were deployed to collect contact pattern data from 4 equine facilities in Ontario. The collected data were used to generate contact networks that were analyzed with both traditional and non-traditional network analysis techniques. The contact networks were used to inform the contact rate of a network epidemic model that was used to quantify the effect of different network structures on the epidemiological outcomes. The model had a typical SEIR structure and incorporated both vaccination and isolation. Equine influenza was used as a case study. The thesis resulted in several important outcomes. First, the modified RFID technology provided a feasible method for contact data collection, specifically in animal agricultural settings.

Second, the contact networks exhibited similar traits across facilities, such as patterns in the relative degree centrality and a failure to satisfy the assumption of homogenous mixing. Third, the empirical contact networks resulted in epidemic curves with similar epidemic durations, peak times, and peak heights when used to inform the contact rate of the network epidemic model. Furthermore, differences in the incidence curves were observed when comparing the empirical networks to theoretical networks such as a homogenous mixing network. Overall, the results indicated that while it is important to use empirical contact data for the characterization of disease dynamics within populations, it is possible to generalize the disease dynamics and associated biosecurity strategies for equine facilities with similar characteristics. The improved understanding gained from this research regarding contact data collection and analysis and the importance of empirical networks in the simulation of disease dynamics enables the improved ability to inform equine biosecurity strategies.

DEDICATION

This thesis is dedicated to the memory of my grandma, Maureen Sonia Milwid, and my friend, Jeff White, whose constant encouragement and enthusiasm for this thesis will always be remembered.

ACKNOWLEDGEMENTS

I would like to thank my advisors Drs. Amy Greer and Terri O'Sullivan for their continued support and guidance during my PhD. Your patience, experience, and insight have helped me grow both personally and professionally. Along with my advisory committee, Drs. Zvonimir Poljak and Marek Laskowki, you have formed a cohesive group that has facilitated interesting discussions providing the forum for knowledge expansion. Thank you for all your help and support throughout my PhD.

I am extremely grateful to the study participants for their willingness and enthusiasm to participate in the study. Your patience, interest and eagerness to help in the study has been exemplary. This thesis would not have been possible without you.

To our summer students, Enise Decluwe-Tulk, Beatrice Hai, and Kamel Omer, thank you for everything. Your company and help with the data collection aspect of the study was greatly appreciated.

To my colleagues and friends in Guelph. Thank you for your insightful thoughts, suggestions, and company. You have truly helped to make Guelph feel like a home away from home. I enjoyed getting to know you all from both a personal and professional standpoint and look forward to collaborating on projects in the future.

Lastly, thank you to my family and friends for your help, support, and encouragement. Dad, your proof-reading skills have taught me lots, and Yon, your ongoing patience and help answering my programming questions have helped me in ways that words just

can't express. Mom, thanks for the late-night phone calls and listening to every practice talk I have ever given. Bobba and Zai, your interest in my work means the world to me.

STATEMENT OF WORK

Chapter 2: The initial research proposal was prepared by Drs. Amy Greer and Terri O'Sullivan. The specific study details were determined by Rachael Milwid, and Drs. Amy Greer and Terri O'Sullivan. The facility recruitment was managed by Dr. Terri O'Sullivan. The facility survey was prepared by Rachael Milwid with input from Drs. Amy Greer and Terri O'Sullivan. The radio-frequency identification firmware was modified by Alexander Olpin and Joel Cummings. The tag deployment was carried out by Rachael Milwid and Terri O'Sullivan, as well as members of the math.epi.lab. The data was analyzed by Rachael Milwid with input from Drs. Amy Greer, Terri O'Sullivan, Zvonimir Poljak, and Marek Laskowski. The manuscript was written by Rachael Milwid and reviewed and edited by Drs. Amy Greer, Terri O'Sullivan, Zvonimir Poljak, and Marek Laskowski.

Chapters 3 and 4: The RFID tag and the facility survey from chapter 2 were used in these chapters. The tag deployment and collection were carried out by Rachael Milwid, and Terri O'Sullivan, as well as members from the math.epi.lab. The data was analyzed by Rachael Milwid with input from Drs. Amy Greer, Terri O'Sullivan, Zvonimir Poljak, and Marek Laskowski. The manuscript was written by Rachael Milwid and reviewed and edited by Drs. Amy Greer, Terri O'Sullivan, Zvonimir Poljak, and Marek Laskowski.

Chapter 5: The model was built by Rachael Milwid with input from Drs. Amy Greer and Terri O'Sullivan. The model outcomes were analyzed by Rachael Milwid with input from Dr. Amy Greer. The manuscript was written by Rachael Milwid and reviewed and edited by Drs. Amy Greer, Terri O'Sullivan, Zvonimir Poljak, and Marek Laskowski.

TABLE OF CONTENTS

Abstract.....	ii
Dedication	iv
Acknowledgements	v
Statement of Work.....	vii
Table of Contents.....	viii
List of Tables.....	xiv
List of Figures.....	xvi
List of Abbreviations	xxi
List of Appendices	xxiii

CHAPTER 1

Introduction, literature review, and thesis objectives.....	1
1.1 Introduction	1
1.2 Disease transmission models.....	3
1.2.1 An overview of mathematical models	3
1.2.2 History and use.....	4
1.2.3 Model formulation	5
1.2.4 Assumption of homogenous mixing.....	6
1.3 Network analysis	7
1.3.1 An introduction to graph theory.....	7
1.3.2 Types of networks	8
1.3.3 Data collection	10
1.3.4 Network formation	11
1.3.5 Network analysis	12

1.3.6	Contact networks and disease transmission.....	13
1.4	Radio-frequency identification tags	15
1.4.1	Radio-frequency identification system	15
1.4.2	Applications	16
1.4.3	Advantages and disadvantages.....	16
1.5	Equine influenza.....	17
1.5.1	A brief background	17
1.5.2	Epidemiology	18
1.5.3	Signs and symptoms	18
1.5.4	Prevention and control strategies	19
1.5.5	Disease transmission models and EI.....	20
1.6	Thesis overview and objectives.....	22
1.7	References.....	25
1.8	Figures	31

CHAPTER 2

Validation of modified radio-frequency identification tag firmware, using an equine population case study	34
2.1 Abstract.....	34
2.2 Introduction	35
2.3 Methods	39
2.3.1 System validation procedure	39
2.3.2 Active RFID tags.....	41
2.3.3 Tag specifications.....	41
2.3.4 Setup	42

2.3.5	Data storage and analysis	45
2.4	Results	46
2.4.1	Technical results.....	46
2.4.2	Graphs.....	47
2.5	Discussion.....	49
2.6	Tables	54
2.7	Figures	55

CHAPTER 3

Quantification of contact patterns using radio-frequency identification proximity tags on an equine facility in Ontario, Canada: a pilot study		58
3.1	Abstract.....	58
3.2	Introduction	59
3.3	Methods	64
3.3.1	RFID tags	64
3.3.2	Pilot study	64
3.3.3	Data analysis	66
3.3.4	Study Assumptions.....	69
3.4	Results	71
3.4.1	Quantifying contact patterns within the facility	71
3.4.2	The assumption of homogenous mixing	72
3.4.3	Generalizability of the contact patterns.....	72
3.4.4	Comparison of the tag-acquired network to the survey-acquired network .	72
3.5	Discussion.....	73
3.5.1	Benefits and limitations.....	79

3.5.2	Future work.....	81
3.6	References.....	83
3.7	Tables	86
3.8	Figures	88

CHAPTER 4

Comparison of the dynamic networks of four equine boarding and training facilities.....	95	
4.1	Abstract.....	95
4.2	Introduction	96
4.3	Methods	99
4.3.1	Study design and data collection process.....	99
4.3.2	Network description	101
4.3.3	Statistical analysis	103
4.4	Results	104
4.4.1	Descriptive results	104
4.4.2	Assessment of the assumption of homogeneous mixing.....	104
4.4.3	Assessment of network generalizability	105
4.4.4	Assessment of network predictability.....	105
4.4.5	Computation and comparison of centrality measures	106
4.4.6	Regression analysis	106
4.5	Discussion.....	107
4.5.1	Limitations, benefits, and opportunities for future work.....	110
4.6	References.....	112
4.7	Tables	115

4.8	Figures	119
-----	---------------	-----

CHAPTER 5

Comparing the effects of non-homogenous mixing patterns on epidemiological outcomes in equine populations: A mathematical modelling study		127
5.1	Abstract.....	127
5.2	Introduction	128
5.3	Methods	131
5.3.1	Empirical networks	131
5.3.2	Contact network structure.....	132
5.3.3	Equine influenza disease transmission model	134
5.4	Results	137
5.4.1	No-intervention model informed with a weighted network.....	137
5.4.2	Vaccination and isolation models informed with unweighted networks....	138
5.5	Discussion.....	141
5.5.1	Study limitations	146
5.6	References.....	148
5.7	Tables	151
5.8	Figures	154

CHAPTER 6

Discussion		159
6.1	Summary of key findings	160
6.1.1	Feasibility of using the modified RFID tags in agricultural settings	160
6.1.2	Network analysis and the resulting network characteristics.....	162
6.1.3	Necessity of using the RFID tags to collect and record contact data	165

6.1.4	The effect and feasibility of different disease control intervention strategies	167
6.2	Limitations	168
6.3	Future work	173
6.4	Concluding remarks	176
6.5	References.....	177

APPENDICES

Appendix 1:	Supplementary materials for Chapter 3	179
Appendix 2:	Supplementary materials for Chapter 4	184
Appendix 3:	Supplementary materials for Chapter 5	188

LIST OF TABLES

Table 2.1 Summary weather statistics during the study week as recorded by a weather station in proximity to the facility (19).	54
Table 3.1: Summary of strengths and weaknesses of selected contact pattern collection methods.	86
Table 3.2: Descriptive measures for the contact network of horses residing on a single equine facility in southwestern Ontario, Canada. Both the density and clustering coefficient were always larger than 0.5.	87
Table 4.1: Descriptive statistics for each of the 4 participating equestrian facilities. The facilities were located in southwestern Ontario, Canada, and boarded 20-28 horses that ranged in age, sex, breed, and use. Each facility had a 100% enrolment rate.....	116
Table 4.2: Description of the networks for each day of the study. Contact data was collected over a 7-day period at 4 equestrian facilities in southwestern Ontario. The data was aggregated and used in the development of 7 networks describing the contacts that occurred on each study day. All of the facilities had multiple components, however, the number of components varied by day. Each network had between 20-28 nodes. The network densities were generally less than 0.5. Densities greater than 0.5 are bolded.	117
Table 4.3: ERGM ^o analysis for each facility. The analysis was carried out using the R ERGM (20) package in the Statnet suite version 2016.9 (9) and studied the effect of the distance between the horses' stalls or pasture location on the resulting network. The distance variable was not always statistically significant. Statistically significant values ($p < 0.05$) are indicated with bold text.	118
Table 5.1: Demographic information for the participating equestrian facilities. The facilities ranged in size, and discipline. Vaccination rates for equine influenza ranged from 46-100%.	151
Table 5.2: List of parameters and variables used in the network epidemic model. Parameter values were obtained from the peer-reviewed literature. The parameter symbol represents parameters found in the deterministic analogue (Figure 5.2).	152
Table 5.3: Key epidemiological results for the disease transmission model of equine influenza. The base model had an SEIR structure and did not include any disease prevention or control interventions. The epidemic model was simulated on different weighted networks including 4 empirical networks ("Empirical Networks 1-4"), a network with the average characteristics (average degree, and network size) of the empirical networks ("Random Mixing Network"), a Complete Network in which all the nodes were connected to each other, and a network with an average degree of 4 ("Minimal	

Network”). The empirical networks were the most similar to the Minimal Network. The Complete Network had the largest epidemic peak and the shortest epidemic duration.

..... 153

LIST OF FIGURES

Figure 1.1: Descriptive statistics of Canada’s horse population. Panel a) shows the proportion of horses in each province in Canada in 2010 (88). Ontario had the second largest equine population, preceded by Alberta. Panel b) describes Canada and Ontario’s equine population by use. With the exclusion of young horses, the majority of horses were used in sport competitions (88). 31

Figure 1.2: Cyclical process of mathematical modelling. The process starts with conceptualizing a real-world problem. Simplifying assumptions should be made based on the system aspects that are important to the problem. The real-world problem can then be translated and analyzed with mathematical tools and empirical data. Lastly, the model results should be interpreted with respect to the original problem. The feasibility of the results should be assessed, and the model should be redefined to accommodate the updated expectations. 32

Figure 1.3: Sample disease transmission model. Panel a): Simple compartmental SIR (susceptible-infectious-recovered) model. This model does not contain demographic processes such as birth and death. Individuals in the susceptible compartment (S) move to the infectious compartment (I) at a given rate β . Infectious individuals can recover at a rate γ , after which they transfer to the recovered compartment. Panel b): analogous network epidemic model. The population has an explicit contact structure. Susceptible individuals in the population transfer to the infectious stage (indicated with a grey node), after which they recover (indicated with a black node). 33

Figure 2.1: The RFID tag system and components. Panel a) shows the components used to flash the tags. The components required to flash a tag are: a computer, a tag (labelled A), a programmer-device (labelled B), and a Nordic board (labelled C). Panel b) shows the traditional RFID setup. In addition to a computer and tag, the communicating tags must be within 10 m of the reader (labelled D) in order to transmit the data to the computer. Panel c) depicts the modified system, in which the tags operate independently of the computer (top line). The data is downloaded to the computer with the use of the programmer-device. The button on the tag is indicated with a black circle in the top panel. 55

Figure 2.2: Selected horse and pasture contact networks. The horses are represented by circular nodes and the pastures are represented by square nodes. Panel a) represents the contacts that occurred on day 1 of the study, panel b) represents the contacts that occurred on day 4 of the study, and panel c) represents contacts that occurred on day 5 (Saturday) of the study. The thickness of the edges between each pair of nodes represents the contact duration for the respective day. Therefore, a larger edge represents a longer contact duration. 56

Figure 2.3: Sample heat map of aggregated tag contacts. The heat maps show the amount of contact that tags (based on individual tag assignment) had with each other. Darker colours indicate a higher duration of contact (hours) throughout the day. 57

Figure 3.1: Graphical representations of contact network characteristics. Panel a) shows a simple, non-directed, non-weighted network. Panel b) shows a directed network. In this network, the contact originates at one node (e.g. node 4) and goes towards another node (e.g. node 3). Panel c) contains a weighted network in which edges with larger weights (larger line widths) imply more contact between the respective nodes. The network in panel d) shows both a dynamic network over a three-day period, as well as a static representation of the same network, marked as “final network”. 88

Figure 3.2: OpenBeacon RFID proximity tag with onboard storage (a), horse wearing a tag on his halter (b), and halters with a tag on the nose piece hung on a pasture fence (c)..... 89

Figure 3.3: Map of the barn (a) and facility (b). Horses are represented by the numbers in the stall. The shape surrounding the numbers signifies the pasture in which the horse was turned out. The triangle represents pasture 1, the circle represents pasture 2 and the square represents pasture 3. The barn was separated into aisles, which were used to construct heat maps to quantify the amount of contact between the horses in the different aisles (Figure 3.4). The horses were turned out in pastures 1-3, except for on the weekend when at least one horse was moved to pasture 4. Each pasture shape in the barn contains a number which refers to the horse IDs in the heat maps in Figure 3.4. 90

Figure 3.4: Summary study results. Networks in which horses are laid out in relative locations of the stalls in the barn (Figure 3) are found in (a). Horses’ nodal shapes are coded by pasture number. The edge width represents the contact duration ($\log(\text{seconds})^2$) between respective horses. (b) shows heat maps of the duration of contact (hours) throughout the day. Horses are numbered in conjunction with the numbering system in Figure 3. Lastly, (c) shows heat maps for the contact between horses in each wing. Only three days of results are shown (day 1 (top row), day 5 (middle row), and day 7 (bottom row)), as the results were similar for each day of the study. The full set of figures can be found in the supplementary information in Figures S3.1-S3.3. 91

Figure 3.5: Selected centrality measures for each horse on each study day. Centrality measures include eigenvector centrality (a), degree (b), and strength (c). Each day, a different horse was the most central horse (day 1=light blue, day 2= orange, day 3=grey, day 4= yellow, day 5=dark blue, day 6= green, day 7= black). The eigenvector centrality ranged between 0.0 and 0.7, while the degree and strength ranged between 1 and 8, and 0 and 40 respectively. No trends were evident in the centrality plots. 92

Figure 3.6: Heat map of the 7 combined dynamic networks. The heat map is not suggestive of homogenous mixing. The colored panels on the left and top axes represent the different clustering of horses by pasture..... 93

Figure 3.7: Heat map of the Jaccard similarity index. The heat map compares the similarity in contacts between participants on each day of the study. All days share at least ~54% of identical contacts. 94

Figure 4.1: Heat maps of the contact that occurred on each facility on day 1 of the study. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. Panels a) and b) show the heat maps for facilities 1 and 2 respectively while panels c) and d) represent facilities 3 and 4 respectively. A strong contact duration (hours per day) exists on the diagonal of the heat map between horses who shared a pasture. There is no evidence of homogeneous mixing at this level of aggregation. 119

Figure 4.2: Aggregated, week-long heat maps for each facility. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. Panels a) and b) contain the heat maps for facilities 1 and 2 respectively while panels c) and d) represent facilities 3 and 4 respectively. Heat maps are visually similar to the daily heat maps in Figure 4.1. There is a strong trend on the diagonal between horses who shared a pasture. There is no evidence of homogeneous mixing at this level of aggregation. 120

Figure 4.3: Boxplot for Jaccard similarity index to quantify the similarity in contacts between participating horses on each day of the study for each facility. The similarity in contacts can be used to generalize the contacts that occurred over the course of the study period. Roughly 11-62% of contacts were repeated over all the networks. 121

Figure 4.4: Jaccard similarity indices for the comparison of the survey-based networks and the networks formed with the logger acquired data. Survey-based networks were formed by assuming that all horses that shared a pasture or barn came in contact. The survey-based networks were able to correctly predict 8.7%-78.7% of contacts that occurred on the four equestrian facilities..... 122

Figure 4.5: Normalized degree centrality for each horse during the study week. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. Different coloured lines represent the different study days (blue=day 1, orange=day 2, grey=day 3, yellow=day 4, black=day 5, green=day 6, and navy=day 7). Overall, the same trends occurred on each of the study days. In other words, with the exception of isolates, horses that had a larger degree compared to other horses had a larger degree on each day of the study compared to the other horses.. 123

Figure 4.6: Normalized strength for each horse during the study week. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4

respectively. Different coloured lines represent the different study days (blue=day 1, orange=day 2, grey=day 3, yellow=day 4, black=day 5, green=day 6, and navy=day 7). Similar trends occurred as those in Figure 4.5. 124

Figure 4.7: Comparison of the normalized degree and normalized strength over the study period. The line on each graph represents the regression line. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. A statistically significant, weak to moderate positive correlation, computed with the spearman correlation coefficient (r) was observed for each facility. 125

Figure 4.8: Normalized betweenness for the horses in each component. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. Gaps in the graphs represent isolates (for which it is impossible to calculate the betweenness centrality). No trends were observed for the betweenness between the facilities. 126

Figure 5.1: Sample theoretical networks used in the network epidemic model. Panel a) represents the Minimal Network which has an average of 48 edges. Panel b) represents the Random Mixing Network which has an average of 87.6 edges. Panel c) represents a Complete Network in which each node is connected to every other node. 154

Figure 5.2: Deterministic analogue of the study model. In this model, horses start out susceptible to infection (S). A proportion (ρ) of the susceptible horses are vaccinated (S_v). Both vaccinated and unvaccinated horses can come in contact. Furthermore, both vaccinated and unvaccinated horses can become infected (E), however the rate of infection differs between the vaccinated horses (transmission probability = p_v) and unvaccinated horses (transmission probability = p). The transmission rate of the pathogen between infectious horses (I) and susceptible horses is determined through the contact rate (c) and the transmission probability. At the end of the infectious period (σ and σ_v for vaccinated and unvaccinated horses respectively), horses can recover (R), where it is assumed that they have full immunity to re-infection. A proportion (η) of the unvaccinated, infectious horses (I) are isolated (Q). These horses remain in isolation for a pre-determined length of time (θ), before they move to the recovered compartment. 155

Figure 5.3: Epidemiological results for the model in the absence of interventions. Panel a) contains plots of the prevalence over time, and panel b) contains plots of the incidence over time. Similar shaped curves were produced for each empirical network. The peak incidence and prevalence for the Complete and Random Networks were larger than the empirical networks, and the epidemic duration was shorter than the empirical networks. The peak incidence and prevalence for 3 of the empirical networks were the smallest, and the epidemic duration for these networks was the longest. 156

Figure 5.4: Heat maps of the cumulative incidence (%) when isolation and vaccination were implemented in the population. The heat maps represent Empirical Networks 1-4

(panels a-d), the Random Mixing Network (panel e), the Complete Network (panel f), and the Minimal Network (panel g). Increasing the proportion of the population isolated and/or vaccinated generally decreased the cumulative incidence. However, a non-linear trend in the cumulative incidence was observed for high isolation rates used in addition to a vaccination program. Other than a 100% vaccination and isolation rate, the implementation of any intervention strategy had minimal effect on the cumulative incidence for the Complete Network..... 157

Figure 5.5: Prevalence plots for Empirical Network 1. Empirical Networks 2-4, the RMN, and the Complete and Minimal Networks resulted in similar trends. The plots show the effect of varying the proportion of the infectious population that is isolated as the proportion of the population that is vaccinated is kept constant. The proportion isolated is portrayed with different line colours: red solid line=0%, pink dotted line=25%, blue dotted line= 50%, green dotted line=75%, and black solid line=100%. Panels a)-e) represent a population in which 0, 25,50, 75, and 100% of horses are vaccinated. For vaccination rates of 0-50%, increasing the isolation rate decreased the disease prevalence. However, the difference in prevalence due to the increasing isolation rates decreased as the proportion vaccinated increased. By a 75% vaccination rate, there was no difference in the prevalence with increasing isolation rates. 158

LIST OF ABBREVIATIONS

CFIA	Canadian Food Inspection Agency
EI	Equine influenza
ERGM or p^*	Exponential random graph model
gHZ	Gigahertz
HA	Haemagglutinin
ID	Identification number
JSI	Jaccard similarity index
kHz	Kilohertz
m	Meter
MHz	Megahertz
NPV	Negative predictive value
PPV	Positive predictive value
R_0	Basic reproductive number
RFID	Radio-frequency identification
RMN	Random mixing network

SEIR	Susceptible-exposed-infectious-recovered
SI	Susceptible-infectious
SIR	Susceptible-infectious-recovered
SLIR	Susceptible-latent-infectious-recovered
sn	Sensitivity
sp	Specificity

LIST OF APPENDICES

Appendix 1: Supplementary materials for Chapter 3	179
Appendix 2: Supplementary materials for Chapter 4	184
Appendix 3: Supplementary materials for Chapter 5	188

CHAPTER 1

Introduction, literature review, and thesis objectives

1.1 Introduction

Infectious diseases pose a major challenge to populations worldwide, with impacts at the societal, economical, and political levels (1). The 2009 Animal Health Act (2) as defined by the Government of Ontario describes 4 goals with respect to animal health in Ontario, Canada:

1. To protect animal health;
2. To establish biosecurity measures, i.e. measures to aid in the prevention, detection, response, control, and recovery from threats that can affect both animal and human health;
3. To regulate animal related activities that can affect both animal and human health, and;
4. To ensure the safety of animal by-products for human use (2).

The threats to animal health as discussed in objective 2 of the Animal Health Act include threats arising from infectious diseases. Infectious diseases are unique due to their continual emergence and their capacity to change over time (3). These attributes make the implementation of effective prevention and control strategies challenging. In the absence of intervention strategies, the introduction and spread of an infectious disease can have deleterious effects on the population, including death (3). Therefore, the World Organisation for Animal Health (OIE Office International des Epizooties) identified 3 strategies in the Sixth Strategic Plan (4) for minimizing the impact of

infectious diseases: 1) control risks at the human-animal-environment interface, 2) build trust between stakeholders through transparency and good communication, and 3) strengthen the sustainability and capacity of National Veterinary Services. The Strategic Plan enumerated 3 methods of accomplishing these strategies: 1) scientific excellence, 2) diversity, participation, commitment and transparency of the OIE, and 3) the OIE governance. The described goals are multi-disciplinary and are not mutually exclusive. For example, research tools such as mathematical models can be used to inform policy regarding the development of prevention and control strategies (5). In the context of infectious diseases, mathematical models describe the trends in disease spread given the implementation of different prevention and control strategies (6,7). The models provide an efficient, ethical, and inexpensive method of evaluating the efficacy of the interventions in the given population (5).

The use of a population's movement and contact patterns in order to characterize mechanisms of disease spread has recently gained momentum for animal populations (8). This has largely been demonstrated with respect to the 2001 foot and mouth disease outbreak in the United Kingdom, where extensive movement data has been recorded over time (9). Unfortunately, this data availability does not extend to all international regions for all animal species. Furthermore, the areas that do collect movement data seldom collect within-farm contact data (10). This paucity of data extends to the Canadian equine population.

The Canadian agricultural census defines a farm as any agricultural establishment that produces and sells agricultural products (11). Using this definition, the 2016 census of

agriculture reported that 291,561 horses resided within 39,164 farms in Canada (12). Furthermore, Equine Canada estimated that in 2010, the second largest proportion of horses were located in Ontario (22.1%), surpassed only by Alberta (32.5%) (Figure 1.1). The contribution of horses to Canadian agriculture and culture has implications for the agricultural and entertainment industries. With respect to the entertainment sector, the majority of the Ontario equine population was used for sport competitions (not including racing) (20%) followed by pleasure riding (17%) (Figure 1.1). Other reported uses for horses were lessons (4%), breeding (14%), and racing (13%). These uses have direct implications for the Canadian economy, for which the equine industry contributes more than 19 billion dollars per year (13). Therefore, a disruption to the equine population, for instance, due to an infectious disease epidemic, could have a large impact on individual and groups of horses, individual horse owners, and the Canadian economy.

1.2 Disease transmission models

1.2.1 An overview of mathematical models

Mathematical models are simplified representations of real-world systems (14). The process of mathematical modelling is cyclical and begins with the conceptualization of a problem (Figure 1.2) (15). Given the complexities of real-world systems, simplifying assumptions are developed based on the aspects of the system that are the most important to represent in the model (14,16). The conceptual model can then be translated into a mathematical model that can be simulated and analyzed using observed data. The model outcomes should be validated with the observed data. There are different reasons for model outcomes that do not reflect the expected disease

dynamics including: 1) an erroneous model, in which case the model should be adjusted given the constraints of the original system (14), and 2) the model has identified a biological aspect of the disease transmission process that is otherwise unknown.

1.2.2 History and use

The use of mathematical models to study the epidemiology of infectious diseases will hereafter be referred to as mathematical epidemiology. The history of mathematical epidemiology dates back to the 18th century when Daniel Bernoulli investigated the effect of variolation (inoculation of an individual with the virus) on life expectancy (17). However, it was only in the late 19th and early 20th centuries that the foundations of mathematical epidemiology were formalized. Many individuals contributed to the early development of mathematical modelling including P.D. En'ko, Sir R.A. Ross, A.G. McKendrick, and W.O. Kermack (17).

Sir R.A. Ross conducted studies on the transmission of malaria. Ross concluded that a possible mechanism for reducing malaria transmission was to minimize the mosquito population. Through these studies, Ross conceived the epidemiological definition of the basic reproduction number (R_0) (17). The basic reproduction number is defined as the average number of secondary infections caused by a single infectious individual in a totally susceptible population (18). Kermack and McKendrick, influenced by Ross, developed a model to describe the spread of disease in a population (19). In the model, individuals were grouped into compartments based on their disease state, including susceptible, infectious, and recovered. The model which is often referred to as a compartmental model, used differential equations to describe the rate of change

between individuals in each compartment (20). These models still serve as the basis for contemporary mathematical models.

1.2.3 Model formulation

The unit of interest in disease transmission models is generally individuals, or groups of individuals in a population (21) and should be selected based on population level characteristics such as the population size and the available data regarding individual-level traits. Models that focus on group-level dynamics are referred to as compartmental models, while models that include individual-level heterogeneity and focus on the dynamics of individuals in the population are referred to as individual- or agent-based models (22). Each group of individuals (or each individual in an individual-based model) is assigned a disease status, for example, susceptible (S), infectious (I), and recovered (R) (Figure 1.3). A model including the S, I, and R disease states is referred to as an SIR model, and along with the model extensions such as the SEIR model (where E represents an exposed group of individuals that is infected, but not yet infectious) is an appropriate framework with which to model population level disease dynamics (21). Disease transmission models simulate the rate at which individuals transition between the different disease states over time (20). For example, in the SIR model, when individuals in the susceptible compartment interact with individuals in the infectious compartment, the susceptible individuals can become infected at a rate β (Figure 1.3). Therefore, the number of individuals leaving the susceptible compartment and entering the infectious compartment can be expressed as βSI . After a defined infectious period (γ), infectious individuals can move to the recovered compartment at a rate $\frac{1}{\gamma}$. The rate

of change for each compartment (S, I, and R) can be described using the system of equations:

$$\begin{cases} \dot{S} = -\beta SI \\ \dot{I} = \beta SI - \gamma I \\ \dot{R} = \gamma I \end{cases}$$

Disease transmission models can be characterized as deterministic or stochastic (14). Deterministic models are governed by a set of parameters that determine the model outcome. Therefore, for a specified input, every simulation of the model will provide identical output, representing the average behavior of the system (22). Deterministic models are inappropriate for small populations in which homogenous mixing is unlikely to occur and stochasticity is likely to have a larger impact on population level outcomes. In these situations, it is more suitable to model the biological processes using stochastic models (20). In contrast to deterministic models, stochastic models are governed by chance. In stochastic models the future behavior of the population is based only on the current time point. Given that stochastic models consider a distribution of possible behaviors, each simulation results in different outcomes (20,22).

1.2.4 Assumption of homogenous mixing

Host heterogeneity can result from a variety of factors including age, susceptibility to infection, and mode of disease transmission (23). This variability can impact the contact structure of a population, and consequently affect the rate of disease transmission within the population (23). Since contact pattern data can be challenging to collect, models often assume homogenous mixing in which individuals mix randomly and have

an equal chance of coming in contact (24). The assumption of homogenous mixing is implemented in disease transmission models through the transmission rate, β , which consists of two parameters: 1) the contact rate and, 2) the probability of transmission given an effective contact (25).

Although existing models that have assumed homogenous mixing have a history of providing reliable predictions, the assumption of homogenous mixing does not accurately represent the mixing patterns for all populations (23). Heterogeneity in contact patterns can influence the spread and transmission of infectious diseases (25) especially in small populations that are subject to small, chance variations in initial conditions (26). Therefore, the inappropriate use of the assumption of homogenous mixing can result in incorrect model outcomes. An alternative method to the assumption of homogenous mixing is to use heterogenous mixing rates. Heterogeneous mixing rates can be incorporated in the model in different ways (see (23) for an in-depth discussion of the different methodology), ranging from using a time-dependent transmission rate ($\beta(t)$) in compartmental models, to specifying a matrix of contact rates for different groups or individuals within the population (23,27,28).

1.3 Network analysis

1.3.1 An introduction to graph theory

Graph theory has a history dating back to 1735, when Euler solved the Konigsberg bridge problem (29). The Konigsberg bridge problem consisted of finding a path that traversed 7 bridges without crossing a single bridge twice. By transforming the map into a series of points connected with lines, Euler used the properties of the figure to prove

that there was no solution to the Königsberg bridge problem (30). Since then, other problems such as Hamilton's game and the four colour theorem have been solved in a similar fashion (30). In essence, the solution to each of these problems has involved using a representative figure of points connected by lines, otherwise known as a graph (30). The field of graph theory is constantly evolving and being applied to different scientific disciplines. Newman and Watts (2011) emphasize that the current applications of graph theory differ from the traditional applications of graph theory in 3 ways: 1) current applications focus on both empirical and theoretical questions, 2) networks do not only describe static systems, but can vary over time, and 3) networks are viewed as structures upon which dynamical processes can occur (31).

Graph theory has been applied to disciplines such as chemistry, computer science, and social networks, each with its own set of terminology (30,32). In 1954, John Barnes coined the term "social network" to describe the interactions between social entities (33). Social networks, hereafter referred to as contact networks, consist of nodes, representing the individual entities in the network, and edges, representing the contact between the nodes (8,33). The networks can be represented with graphs or matrices, and analysed using network analysis methods (33).

1.3.2 Types of networks

Both the research question and the population of interest should be considered when deciding on an appropriate structural and temporal representation of the contact network. Structural characteristics include attributes such as edge weight and directionality. Edge weight can be used to describe the strength of an interaction, for

example the duration or frequency of a contact event between two individuals (34,35). Edge weights can be represented in graphs with characteristics such as the edge width, or in matrices by populating the matrix entry between two nodes with the numerical representation of the weight (36). In contrast, in unweighted graphs, the presence of an edge between two nodes indicates contact between the nodes. This contact, or lack thereof is indicated with a 1 or 0 respectively in the contact matrix. Directed graphs describe the direction of a relationship, for example, movement from one location to another location. Directed graphs are visualized with an arrow connecting the two nodes, pointing in the direction of the contact. Alternatively, for interactions in which an event occurs simultaneously between the two entities, such as contact between two individuals, the network can be undirected (37).

Networks can also include a temporal component (i.e. static or dynamic processes). Due to the challenges of collecting and analyzing dynamic data, most studies to date have focused on static networks (38,39). Static networks are appropriate representations for populations in which the contacts exhibit minimal change over time (40). Therefore, static networks are represented with graphs in which the structure of the nodes and edges is fixed (40). In contrast, for populations in which the contact structure does change over time, it is often more appropriate to use dynamic networks to describe the contact structure of the population over time (41). In these instances, the network representation can change according to population level changes such as demographic (e.g. birth and death) and behavioural changes (e.g. avoidance behaviours) (40). Dynamic networks can be represented in different ways including

multiple static networks or videos that display the network edges as they evolve over time (39).

1.3.3 Data collection

The formation of representative contact networks is largely determined by the quality and quantity of the collected data (42). Traditional data collection methods such as observation, surveys, and diaries have made the collection of large sets of reliable data challenging (31,43). Each data collection method has different strengths, weaknesses, and inherent biases. Observation has historically taken a hands-on approach where researchers observe and record the contacts that occur between individuals in the population. Technological advances have also enabled automated recording of the contacts using video recording technologies (44). Regardless of the observation methods used, this method of data collection can be expensive and difficult to implement for large populations. Although observational methods can be invasive and subject to technological constraints such as software facial recognition issues, they are capable of providing information on the types of social contacts that occur within the population (44).

Ego-centric data collection methods, focus on the collection of data associated with a specific individual (45). Ego-centric methods include contact diaries and surveys. Contact diaries and surveys involve the participant recording the type of contact that occurred within a specified time frame (42,44). These methods enable the specification of the type of contact, as well as other details such as the contact frequency and duration (46). While contact diaries and surveys are relatively easy to administer, they

are subject to recall and reporting biases, as well as a possible low response rate (44,46). Furthermore, these data collection methods focus on the network of each reporting individual as opposed to the population-level network (44).

Recently, proximity sensing technology has been used to collect and record contact pattern data. Proximity technology includes both Bluetooth® and radio-frequency identification (RFID) technology (44). Data collection through proximity sensing technology requires that all participants (members of the population being studied) wear a signal transmitting tag, enabling the aggregation of the signals to form a network (44). While this data collection method allows for large amounts of data to be collected over long periods of time, it is limited to the collection of data from the individuals wearing the tags. Furthermore, proximity technology can be subject to radio errors in which contacts are recorded that never occurred, or reciprocal contacts are not recorded (44).

Technological advancements have enabled the collection of large of amounts of data with increasing accuracy. Given the abundance of available data, the next step in the analytical process is to determine the biological implications of different data implementation schemes, and consequently, how to integrate the data into models in a meaningful way (47).

1.3.4 Network formation

Network formation can be accomplished in two different ways. If a sufficient amount of data has been collected to accurately represent the population-level contact patterns, then the collected data can be used to construct the network (33). This is often achieved

by defining contact matrices in which the column and row names represent the nodes, and each cell contains the amount (or the presence or absence) of contact between the nodes (48). However, for some populations (e.g. marginalized populations) the population characteristics (e.g. demographics, risk factors, etc.) can make it difficult to collect sufficient data to characterise the contact network (48). In these situations, statistical methods can be used to infer the contacts that occurred (49). These methods include exponential random graph models otherwise known as p^* models, and use the characteristics of the collected data to generate contact networks for a population of a specified size (50).

1.3.5 Network analysis

The analysis of contact networks has applications for describing the disease spread potential within the population. Traditional network analysis includes aspects such as determining a node's importance in the network (centrality), as well as community detection algorithms. These analyses are largely applied to static networks as the network analysis methods for dynamic networks are not well developed (39). Moreover, analytic methods become increasingly intractable with the increased complexity of the system.

Centrality measures quantify the importance of a node in the network and often include degree, closeness, betweenness, and eigenvector centrality (8,51,52). Briefly, degree refers to the number of edges that are incident to a node (53). The degree centrality has a direct implication in the disease spread potential since nodes with a higher degree have a greater chance of transmitting a disease to other nodes or becoming infected by

connected nodes (52). Closeness provides an estimate of how close a node is to all of the other nodes in the network. Individuals with a high closeness play an important role in the transmission of the disease to other individuals in the network. Betweenness represents the probability that the shortest path between two nodes passes through a node of interest. Betweenness is important to disease transmission as nodes with a high betweenness can serve as cut-points. Cut-points are nodes whose removal will segment the graph, inhibiting the spread of disease between the two components (8,54). Lastly, eigenvector centrality is a measure of a node's secondary contacts. Eigenvector centrality is important to disease transmission as nodes with a higher eigenvector centrality have an increased potential of transmitting and contracting the disease (53).

Community structures are defined as densely connected groups of nodes that are weakly connected to the rest of the network (55). It has been reported that communities have a high clustering coefficient, which is a measure describing the extent to which nodes form clusters (54,56). In their study of the effect of community structure on epidemic spread, Wu and Liu (2008) reported that the epidemic potential will decrease with an increase in the clustering coefficient. The community structure of a network can be detected with algorithms, each of which has different strengths and weaknesses. Harenberg et al., (2014) describe community detection algorithms in more detail (55).

1.3.6 Contact networks and disease transmission

The simulation of a disease transmission model over a contact network is referred to as a network epidemic model (57). Contact networks can influence the trends in disease

spread within a population (58,59), and enable the incorporation of individual-level heterogeneities in disease transmission models (60). When analysing contact networks with respect to disease dynamics, nodes represent hosts and edges represent interactions enabling the transmission of pathogens (40,61). Numerous studies have identified the impact of networks on disease dynamics such as the epidemic duration and severity, however, many of these studies were conducted on static networks (59). While static networks may be appropriate to model rapidly spreading pathogens which cause acute infections, they do not necessarily reflect the epidemic behaviour of diseases that continue to circulate for longer periods of time, such as endemic diseases, during which individuals can modify their behaviour (40,62). In situations where the underlying network is likely to change throughout the epidemic, static networks do not capture the epidemic behaviour as well as dynamic networks (38).

The use of networks to represent the heterogeneities in the host population is gaining momentum in the literature. However, there are still many challenges associated with the use of networks in epidemiological modelling. For instance, it is not always intuitive how the structural properties of the underlying network such as edge weights relate to the transmission dynamics (61). Furthermore, calculating epidemiological measures such as the basic reproductive number become complicated when adding in the extra dynamics contributed by the contact network (61). Regardless of these challenges, networks have the capacity to further the understanding gained from disease transmission models.

1.4 Radio-frequency identification tags

1.4.1 Radio-frequency identification system

Radio-frequency identification (RFID) technology uses radio-waves to send and receive signals (63). The RFID system generally consists of 3 components: 1) an antenna, 2) a transceiver, and 3) a transponder hereafter referred to as a tag (64). The antenna enables communication between the transceiver and the tag. The transceiver is responsible for data acquisition, and the tag reads and records data. When the antenna and transceiver are combined, they are referred to as a reader (64).

The RFID system can be characterized in two ways: 1) by the data storage capabilities, and 2) by the power source (64). In the first classification scheme, the tags are said to have either read-only or read/write memory. Read-only tags do not have data storage capabilities, while read/write tags can both read and store data (64). In the second classification scheme, RFID tags can be classified as active or passive: active tags use a battery to power the system, while passive tags are powered by radio signals transmitted from the RFID reader (65). Active tags generally have read-write memory and a long communication range, while passive tags are generally read-only with shorter communication ranges (63,64).

The RFID frequency ranges vary and are regulated by governmental agencies (63), however, there are 3 frequencies which are generally used for RFID applications. Low frequency bands range from 100-500 kilohertz (kHz). Low frequency RFID tags are generally used in applications such as inventory control and have a low reading rate and short to medium read range. Mid-range frequencies range from 10-15 megahertz

(MHz) and are generally used in applications such as access control. These tags have a short to medium read ability and a medium reading speed. Lastly, high frequencies range from 850-950 MHz and 2.4-5.8 gigahertz (GHz). These tags are often used in toll collection systems and have a long read range with a high reading speed (63).

1.4.2 Applications

Radio-frequency identification technology has been implemented in applications including supply chain tracking, toll collection, and animal identification (63). In terms of contact network data collection, RFID technology has been used to collect contact data between individuals wearing the tags in both human and animal populations. For example, human studies have used data collected with proximity tags to study disease outbreaks amongst school children, and health care workers and patients (66,67). Animal studies, have been conducted on animals such as deer and possums to study the animals' mating patterns (68,69).

1.4.3 Advantages and disadvantages

Although RFID technology has many advantages, it also has inherent disadvantages (65). For example, while the RFID tags do not require a direct line of sight with the reader, objects containing metal may interrupt the normal transmission process and result in unpredictable performance. Furthermore, RFID technology has a high communication frequency however the transmission may result in incorrect data entries due to the distortion, deflection, interference, and absorption of radio waves (65,70).

Despite its limitations, the decreasing size and cost of RFID technology has resulted in an increase in use for agricultural animal research purposes (70). The ability of RFID technology to capture large amounts of data make it a good candidate for applications such as recording contact pattern data, body temperature, and traceability information (70–73). All of these applications enable the enhanced ability to predict animal disease dynamics using methods such as early warning signals and network analysis.

1.5 Equine influenza

1.5.1 A brief background

Equine influenza (EI) is considered to be the most important equine disease due to its transmissibility and the associated economic impact (74). Equine influenza is a respiratory disease caused by the equine influenza virus (type A). Although both the H3N8 and H7N7 subtypes have been identified as the cause of past outbreaks, the H7N7 subtype has not been active since 1979 (75).

The first reported EI outbreak occurred in Eastern Europe circa 1956 (76). The outbreak was caused by the H7N7 subtype, which was subsequently identified from outbreaks worldwide (76). In 1963, the H3N8 subtype was identified as the cause of a Miami, Florida epidemic (76), and has since become the only circulating subtype.

Prior to 2007, EI had been the cause of equine-related outbreaks worldwide with the exception of geographically isolated areas such as Australia, New Zealand, Iceland, and Japan (77). However, in 2007 Australia experienced an EI epidemic that was attributed to imported horses (78). The Australian outbreak has reinforced the importance of

having a prevention and control strategy in place prior to an outbreak in order to minimize the epidemic impact on the population (79).

1.5.2 Epidemiology

Equine influenza can be transmitted through aerosol, wind, nose-to-nose contact, fomites, and human contact (76,80). The virus is heat and cold sensitive and can be killed by disinfectants (76). Equine influenza is highly contagious, with a morbidity rate of up to 100% and an R_0 of 2-10 (80–84). Although EI infection can occur in horses of all ages, young naïve horses, and unvaccinated horses have an increased susceptibility to the disease (76). After infection, the virus has an incubation period of 1-5 days, however, shedding can begin as early as 24 hours after infection and can last for 7-10 days (76,80). Once recovered, horses exhibit natural immunity to re-infection for at least 1 year (80,85).

Equine influenza outbreaks are more likely to occur during cooler seasons when groups of horses spend a significant amount of time in a shared air space, i.e. indoor stabling (76). Similarly, viral shedding is increased during race and competition seasons when large groups of horses congregate together (76).

1.5.3 Signs and symptoms

Equine influenza can be recognized by clinical signs that include fever, lethargy, coughing, anorexia, and/or nasal discharge (76,80). Most of these clinical signs can last for 1-5 days, however, coughing can last up to 3 weeks. The majority of EI infections resolve without incident, however severe infections or secondary bacterial infections

can result in severe outcomes or sequela. Pregnant horses may abort and susceptible foals may develop pneumonia which can be fatal (76,80). Reduced severity of clinical signs can be observed in horses with a partial immunity from vaccination or previous infections (76,80).

1.5.4 Prevention and control strategies

Vaccination and immunity

Equine influenza vaccines were first developed circa 1960 (85). Although both modified-live vaccines and killed (inactivated) vaccines are available in North America, the majority of vaccines contain an inactivated whole virus (76). Since the vaccine efficacy is highly dependent on the amount of circulating antibodies against the influenza haemagglutinin (HA) glycoprotein, it will be ineffective in horses with low antibody levels (85). Given the imperfect nature of the vaccine, vaccinated horses that become infected can be asymptomatic and are still capable of shedding the virus (75,76,80). In this case, the vaccine acts to reduce the severity of clinical signs experienced by vaccinated horses. It is recommended that horses be re-vaccinated every 6 months (75,76).

Other control measures

Both quarantine, i.e. the segregation of horses that were potentially exposed to a pathogen, and isolation, i.e. the segregation of ill horses, can be used to prevent the spread of EI. Due to the ease with which equine influenza can be transmitted, horses that are thought to be infected with EI should be quarantined. Furthermore, it is

advisable to quarantine all horses who had contact with infectious horses. Horses should be quarantined on arrival to a new facility (76,80). Lastly, infectious horses should be isolated for the duration of the infectious period.

Biosecurity for the Canadian equine population

In 2016, the Canadian Food Inspection Agency (CFIA) in collaboration with Agriculture and Agri-Food Canada, Equine Canada, and other stakeholders developed a biosecurity standard containing biosecurity guidelines and recommendations. The guide, which defines biosecurity as a set of principles and practices that can be used to minimize the risks attributed to pathogens, discusses the importance of the implementation of biosecurity strategies, principles of prevention and control, and provides suggestions about how to minimize the transmission and spread of equine infectious diseases (86). The document, however, was meant as a guide and states that the suggestions provided are voluntary. Interestingly, there are currently no mandatory biosecurity regulations for equine health in Canada.

1.5.5 Disease transmission models and EI

The application of disease transmission models to animal infectious diseases is increasing in the literature (77), however, few models exist describing the transmission process of equine influenza within populations. Previous EI models tended to assume homogenous mixing between the horses, and generally focused on assessing the effects of vaccination as an EI intervention. The following section summarizes a selection of the previously published models in this area.

Glass et al. (2002) used a stochastic SEIR (susceptible-exposed-infectious-recovered) model to simulate an EI epidemic. The model assumed homogenous mixing and included the effects of a partially effective vaccine. The study population was a single yard of horses and the model was simulated for the length of a single epidemic. Given the short simulation time, the model did not incorporate waning immunity. An epidemic curve from a New York race track was used to validate the model. The authors concluded that vaccination decreased the size of the epidemic (77).

Garner et al. (2011) developed a stochastic SLIR (susceptible-latent-infectious-recovered) model to demonstrate the effects of early vaccination use on the 2007 Australian EI outbreak. The model was geographically explicit and used a single premises as the unit of study. Data from the Australian EI outbreak was used to validate the model. The authors concluded that the early deployment of a vaccination program was successful in reducing the number of infected premises within the population of interest (83).

Baguelin et al. (2010) developed a stochastic, metapopulation, SEIR model. The model assumed that each unit in the model (e.g. a yard of horses) was homogenous in its response to infection, and well mixed with respect to the potential for coming in contact. The authors included vaccination in the model, and assumed that at the start of the model, the study population had heterogenous levels of vaccine coverage across the individual horse yards. The authors concluded that the size of the epidemic was largely influenced by the size of the initial infectious facility. Furthermore, vaccination was

shown to be beneficial at the individual level, but was impacted by the vaccine coverage in other yards at a population level (87).

Finally, Park et al. (2004) used a stochastic SEIR model to examine the effect of strain heterology (i.e. strains which are similar but not identical) in a training-yard. The model studied the effect of vaccination on the EI prevalence while taking the horses' pre-exposure antibody level and seroconversion into account. The authors concluded that an outbreak was less likely to occur in a population vaccinated with a vaccine containing the circulating viral strain than with a vaccine containing a viral strain from a different lineage than the circulating strain (81).

1.6 Thesis overview and objectives

The ability to predict disease dynamics and effective prevention and control strategies in advance of an outbreak has the potential to minimize the epidemic outcomes. This in turn can reduce the costs associated with epidemics to both governmental agencies and individual horse owners. However, in order to use mathematical methods to make relevant and realistic predictions, it is necessary to parameterize the model with accurate and appropriate data. There is currently a scarcity of data with respect to equine contact patterns, making it challenging to use mathematical tools to understand the disease dynamics within equestrian facilities. To this end, this thesis presents a starting point to fill this gap in knowledge, by using empirical data to aid in the biosecurity decision making process of equine health.

The rationale behind this thesis was to substantiate the suitability of representing equine contact patterns with homogenous mixing. This was accomplished through different studies, each building on the previous study. Specifically, the traditional RFID system was modified to enable the collection of contact data in large, animal agricultural areas. The modified RFID system was used to collect contact pattern data from different equine facilities in order to characterize and compare the contact structure of equine facilities in Ontario. Lastly, the effect of different contact structures on disease dynamics was analyzed through the use of network models. Equine influenza was used as a case study. The process and results of the described objectives are described in this thesis in 4 chapters. The individual aspects of the thesis are documented in chapters 2-5 and are as follows:

- Chapter 2 describes and validates the methodology used to collect and record contact pattern data within an equestrian facility.
- Chapter 3 uses traditional and non-traditional network analysis methods for the characterization of the collected data. This chapter serves as a pilot study to document the analytic techniques possible with the data collected in chapter 2.
- Chapter 4 uses network analysis methods to characterize and compare the contact networks for 4 single-barn equestrian facilities located in southwestern Ontario, Canada.

- Chapter 5 uses network epidemic models to compare the epidemic outcomes when different contact networks were used to inform the model contact rate.

The last chapter of this thesis summarizes the main results and concludes with a discussion of the limitations of the described studies, as well as directions for future work.

1.7 References

1. Fonkwo PN. Pricing infectious disease. The economic and health implications of infectious diseases. *EMBO Rep.* 2008;9(SUPPL. 1):3–7.
2. Government of Ontario. Animal Health Act, 2009. 2018.
3. Cohen ML. Changing patterns of infectious disease. *Nature.* 2000;406(6797):762–7.
4. OIE World Organisation for Animal Health. The Sixth Strategic Plan (2016-2020) [Internet]. 2016. Available from: <http://www.oie.int/>
5. Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N. Mathematical models in the evaluation of health programmes. *Lancet.* 2011;378(9790):515–25.
6. Huppert A, Katriel G. Mathematical modelling and prediction in infectious disease epidemiology. *Clin Microbiol Infect.* 2013;19(11):999–1005.
7. Lofgren E, Halloran ME, Rivers CM, Drake JM, Porco TC, Lewis B, et al. Opinion: Mathematical models: A key tool for outbreak response. *Proc Natl Acad Sci.* 2014;111(51):18095--18096.
8. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis.* 2009;56(4):109–20.
9. Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *J R Soc Interface.* 2006;3(10):669–77.
10. Chen S, White BJ, Sanderson MW, Amrine DE, Ilany A, Lanzas C. Highly dynamic animal contact network and implications on disease transmission. *Sci Rep.* 2014;4:4472–9.
11. Statistics Canada. Census of Agriculture [Internet]. 2017. Available from: <http://www23.statcan.gc.ca>
12. Statistics Canada. Table 004-0224 - Census of Agriculture, other livestock on census day, every 5 years [Internet]. CANSIM (database). [cited 2017 Oct 27]. Available from: <http://www5.statcan.gc.ca/cansim/>
13. Equine Canada. the Economic Contribution From Horses. 2010;65–76.
14. Torres N V., Santos G. The (mathematical) modeling process in biosciences. *Front Genet.* 2015;6:354.
15. Motta S, Pappalardo F. Mathematical modeling of biological systems. *Brief Bioinform.* 2012;14(4):411–22.

16. Levins R. The strategy of model building in population biology. *Am Sci.* 1966;54(4):421–31.
17. Brauer F. Mathematical epidemiology is not an oxymoron. *BMC Public Health.* 2009;9(1):S2.
18. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface.* 2005;2(4):281–93.
19. Diekmann O, Heesterbeek H, Metz H. The legacy of Kermack and McKendrick. *Publ Newt Inst.* 1995;5:95–115.
20. Allen LJ, Brauer F, Van den Driessche P, Wu J. *Mathematical epidemiology.* Cachan JMM, Groningen FT, Paris BT, editors. Springer; 2008.
21. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. *Phys Life Rev.* 2016;18:66–97.
22. Mishra S, Fisman DN, Boily M-C. The ABC of terms used in mathematical models of infectious diseases. *J Epidemiol Community Heal.* 2011;65(1):87–94.
23. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface.* 2007;4(16):879–91.
24. Volz E, Meyers LA. Susceptible-infected-recovered epidemics in dynamic contact networks. *Proc R Soc B Biol Sci.* 2007;274(1628):2925–34.
25. Hu H, Nigmatulina K, Eckhoff P. The scaling of contact rates with population density for the infectious disease models. *Math Biosci.* 2013;244(2):125–34.
26. Vynnycky E, White R. *An introduction to infectious disease modelling.* Oxford University Press; 2010.
27. Kong L, Wang J, Han W, Cao Z. Modeling heterogeneity in direct infectious disease transmission in a compartmental model. *Int J Environ Res Public Health.* 2016;13(3):253.
28. Brooks CP, Antonovics J, Keitt TH. Spatial and Temporal Heterogeneity Explain Disease Dynamics in a Spatially Explicit Network Model. *Am Nat.* 2008;172(2):149–59.
29. Alexanderson GL. About the Cover: Euler and Königsberg’s Bridges: a Historical View. *Bull Am Math Soc.* 2006;43(406):567–73.
30. Foulds LR. *Graph theory applications.* Springer Science & Business Media; 2012.
31. Newman M, Watts A-LBDJ. *The structure and dynamics of networks.* Princeton University Press; 2011.
32. Trudeau RJ. *Introduction to graph theory.* Courier Corporation; 2013.

33. Wasserman S, Faust K. *Social network analysis: Methods and applications*. Vol. 8. Cambridge university press; 1994.
34. Yang Z, Zhou T. Epidemic spreading in weighted networks: An edge-based mean-field solution. *Phys Rev E*. 2012;85(5):56106.
35. Salathe M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci*. 2010;107(51):22020–5.
36. Kolaczyk ED, Csárdi G. *Statistical Analysis of Network Data with R*. 2014;65. Available from: <http://link.springer.com/10.1007/978-1-4939-0983-4>
37. Masuda N. Directionality of contact networks suppresses selection pressure in evolutionary dynamics. *J Theor Biol*. 2009;258(2):323–34.
38. Vernon MC, Keeling MJ. Representing the UK's cattle herd as static and dynamic networks. *Proc R Soc B Biol Sci*. 2009;276(1656):469–76.
39. Kolaczyk ECG. *Statistical Analysis of Network Data with R [Internet]*. Vol. 65, Springer. Springer Science & Business Media; 2014. 179-195.
40. Bansal S, Read J, Pourbohloul B, Meyers LA. The dynamic nature of contact networks in infectious disease epidemiology. *J Biol Dyn*. 2010;4(5):478–89.
41. Butts CT. Revisiting the Foundations of Network Analysis. *Science (80-)*. 2009;325(5939):414–6.
42. Marsden P V. *Network Data and Measurement*. *Annu Rev Sociol*. 1990;16(1):435–63.
43. Mikolajczyk RT, Kretzschmar M. Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Soc Networks*. 2008;30(2):127–35.
44. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings DAT. Close encounters of the infectious kind: Methods to measure social mixing behaviour. *Epidemiol Infect*. 2012;140(12):2117–30.
45. Marsden P V. Egocentric and sociocentric measures of network centrality. *Soc Networks*. 2002;24(4):407–22.
46. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin Microbiol Infect*. 2014;20(1):10–6.
47. Machens A, Gesualdo F, Rizzo C, Tozzi AE, Barrat A, Cattuto C. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect Dis*. 2013;13(1):185–200.

48. Handcock MS, Gile KJ. Modeling Social Networks From Sampled Data. *Ann Appl Stat.* 2010;4(1):5–25.
49. Krivitsky PN, Morris M. Inference for social network models from egocentrically sampled data , with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann Appl Stat.* 2017;11(1):427–55.
50. Welch D, Bansal S, Hunter DR. Statistical Inference to advance network models in epidemiology. *Epidemics.* 2011;3(1):38–45.
51. Borgatti SP. Centrality and network flow. *Soc Networks.* 2005;27(1):55–71.
52. Dubé C, Ribble C, Kelton D, McNab B. A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound Emerg Dis.* 2009;56(3):73–85.
53. Silk MJ, Croft DP, Delahay RJ, Hodgson DJ, Boots M, Weber N, et al. Using social network measures in wildlife disease ecology, epidemiology, and management. *Bioscience.* 2017;67(3):245–57.
54. Makagon MM, McCowan B, Mench JA. How can social network analysis contribute to social behavior research in applied ethology? *Appl Anim Behav Sci.* 2012;138(3–4):152–61.
55. Harenberg S, Bello G, Gjeltema L, Ranshous S, Harlalka J, Seay R, et al. Community detection in large-scale networks : a survey and empirical evaluation. *WIREs Comput Stat.* 2014;6:426–39.
56. Wu X, Liu Z. How community structure influences epidemic spread in social networks. *Phys A Stat Mech its Appl.* 2008;387(2–3):623–30.
57. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci.* 2015;370(1669):20140107–20140107.
58. Krause J, Croft DP, James R. Social network theory in the behavioural sciences: Potential applications. *Behav Ecol Sociobiol.* 2007;62(1):15–27.
59. Fefferman NH, Ng KL. How disease models in static networks can fail to approximate disease in dynamic networks. *Phys Rev E - Stat Nonlinear, Soft Matter Phys.* 2007;76(3):031919–03120.
60. Craft ME, Caillaud D. Network models: An underutilized tool in wildlife epidemiology? *Interdiscip Perspect Infect Dis.* 2011;2011:1–12.
61. Pellis L, Ball F, Bansal S, Eames K, House T, Isham V, et al. Eight challenges for network epidemic models. *Epidemics.* 2015;10:58–62.
62. Gross T, D’Lima CJD, Blasius B. Epidemic dynamics on an adaptive network. *Phys Rev Lett.* 2006;96(20):208701–5.

63. Roberts CM. Radio frequency identification (RFID). *Comput Secur.* 2006;25(1):18–26.
64. Domdouzis K, Kumar B, Anumba C. Radio-frequency identification (RFID) applications: A brief introduction. *Adv Eng Informatics.* 2007;21(4):350–5.
65. McCombie W, Welt BA. Radio Frequency Identification Technology. In: Kirk-Othmer Encyclopedia of Chemical Technology. Wiley Online Library; 2011.
66. Gemmetto V, Barrat A, Cattuto C. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis.* 2014;14(1):695–705.
67. Mastrandrea R, Soto-Aladro A, Brouqui P, Barrat A. Enhancing the evaluation of pathogen transmission risk in a hospital by merging hand-hygiene compliance and contact data: a proof-of-concept study. *BMC Res Notes.* 2015;8(1):426.
68. Walrath R, Van Deelen TR, VerCauteren KC. Efficacy of proximity loggers for detection of contacts between maternal pairs of white-tailed deer. *Wildl Soc Bull.* 2011;35(4):452–60.
69. Ramsey D, Spencer N, Caley P, Efford M, Hansen K, Lam M, et al. The effects of reducing population density on contact rates between brushtail possums: Implications for transmission of bovine tuberculosis. *J Appl Ecol.* 2002;39(5):806–18.
70. Ruiz-Garcia L, Lundadei L. The role of RFID in agriculture-applications, limitations and challenges. *Comput Electron Agric.* 2011;79(1):42–50.
71. Feng J, Fu Z, Wang Z, Xu M, Zhang X. Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. *Food Control.* 2013;31(2):314–25.
72. Schaefer AL, Cook NJ, Bench C, Chabot JB, Colyn J, Liu T, et al. The non-invasive and automated detection of bovine respiratory disease onset in receiver calves using infrared thermography. *Res Vet Sci.* 2012;93(2):928–35.
73. Taylor PS, Hemsworth PH, Groves PJ, Gebhardt-Henrich SG, Rault JL. Ranging behaviour of commercial free-range broiler chickens 1: Factors related to flock variability. *Animals.* 2017;7(7):54–68.
74. Cullinane A, Elton D, Mumford J. Equine influenza - surveillance and control. *Influenza Other Respi Viruses.* 2010;4(6):339–44.
75. Lewis NS, Daly JM, Russell CA, Horton DL, Skepner E, Bryant NA, et al. Antigenic and genetic evolution of Equine Influenza A (H3N8) Virus from 1968 to 2007. *J Virol.* 2011;85(23):12742–9.
76. Myers C, Wilson WD. Equine influenza virus. *Clin Tech Equine Pract.* 2006;5(3):187–96.

77. Glass K, Wood JLN, Mumford JA, Jesset D, Grenfell BT. Modelling equine influenza 1: a stochastic model of within-yard epidemics. *Epidemiol Infect.* 2002;128(3):491–502.
78. Cowled B, Ward MP, Hamilton S, Garner G. The equine influenza epidemic in Australia: Spatial and temporal descriptive analyses of a large propagating epidemic. *Prev Vet Med.* 2009;92(1–2):60–70.
79. Smyth GB, Dagley K, Tainsh J. Insights into the economic consequences of the 2007 equine influenza outbreak in Australia. *Aust Vet J.* 2011;89(SUPPL.1):151–8.
80. vanMaanen C, Cullinane A. Equine influenza virus infections: an update. *Vet Q.* 2002;24(2):79–94.
81. Park AW, Wood JLN, Daly JM, Newton JR, Glass K, Henley W, et al. The effects of strain heterology on the epidemiology of equine influenza in a vaccinated population. *Proc R Soc London B Biol Sci.* 2004;271(1548):1547–55.
82. Satou K, Nishiura H. Basic reproduction number for equine-2 influenza virus A (H3N8) epidemic in racehorse facilities in Japan, 1971. *J Equine Vet Sci.* 2006;26(7):310–6.
83. Garner MG, Cowled B, East IJ, Moloney BJ, Kung NY. Evaluating the effectiveness of early vaccination in the control and eradication of equine influenza--a modelling approach. *Prev Vet Med.* 2011;99(1):15–27.
84. Daly JM, Newton JR, Wood JLN, Park AW. What can mathematical models bring to the control of equine influenza? *Equine Vet J.* 2013;45(6):784–8.
85. Daly JM, Newton JR, Mumford JA. Current perspectives on control of equine influenza. *Vet Res.* 2004;35(4):411–23.
86. Canadian Food Inspection Agency. Canadian Food Inspection Agency [Internet]. Requirements for Livestock Producers: Livestock Identification and Traceability Program. 2016 [cited 2017 Apr 27]. Available from: <http://www.inspection.gc.ca/>
87. Baguelin M, Newton JR, Demiris N, Daly JM, Mumford JA, Wood JLN. Control of equine influenza: scenario testing using a realistic metapopulation model of spread. *J R Soc Interface.* 2010;7(42):67–79.
88. Equine Canada. Horses in Canada in 2010. 2011;(23):9–32.

1.8 Figures

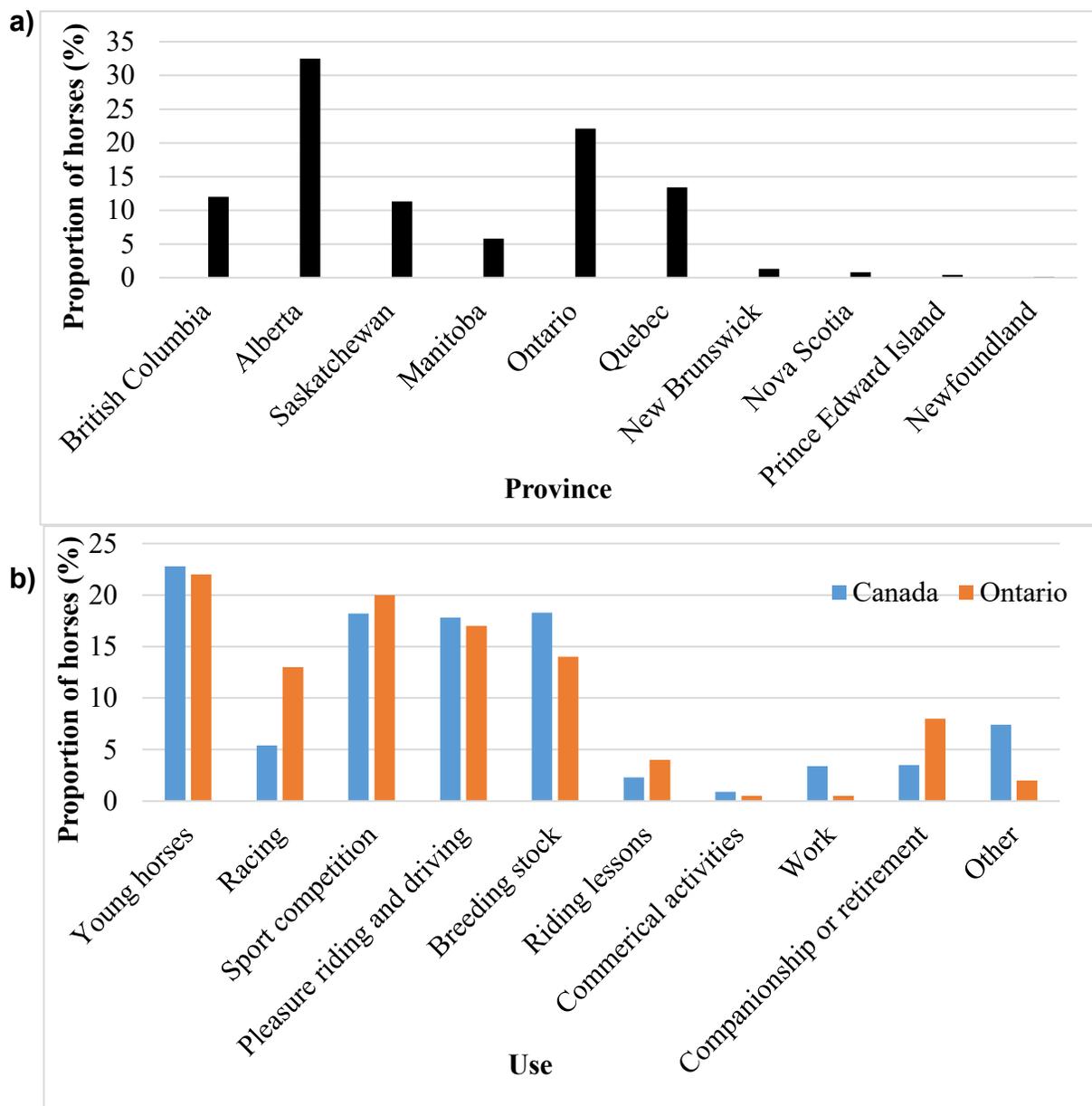


Figure 1.1: Descriptive statistics of Canada's horse population. Panel a) shows the proportion of horses in each province in Canada in 2010 (88). Ontario had the second largest equine population, preceded by Alberta. Panel b) describes Canada and Ontario's equine population by use. With the exclusion of young horses, the majority of horses were used in sport competitions (88).

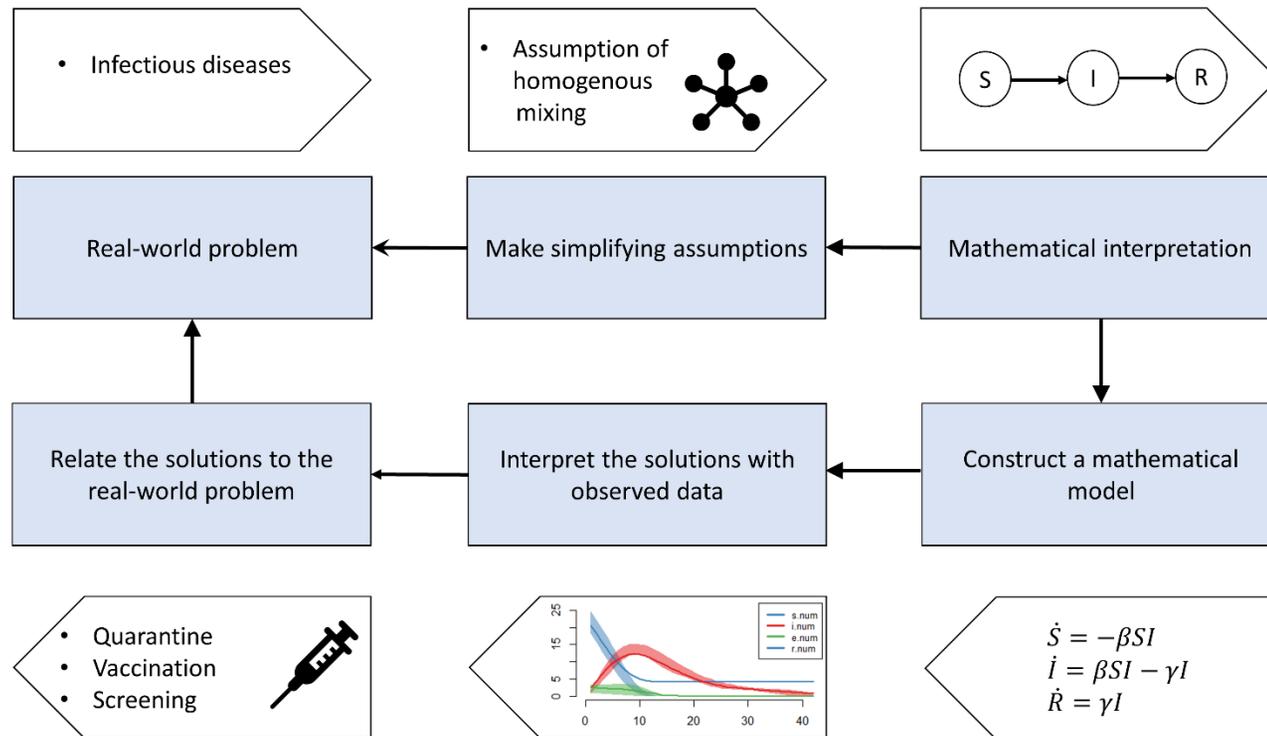


Figure 1.2: Cyclical process of mathematical modelling. The process starts with conceptualizing a real-world problem. Simplifying assumptions should be made based on the system aspects that are important to the problem. The real-world problem can then be translated and analyzed with mathematical tools and empirical data. Lastly, the model results should be interpreted with respect to the original problem. The feasibility of the results should be assessed, and the model should be redefined to accommodate the updated expectations.

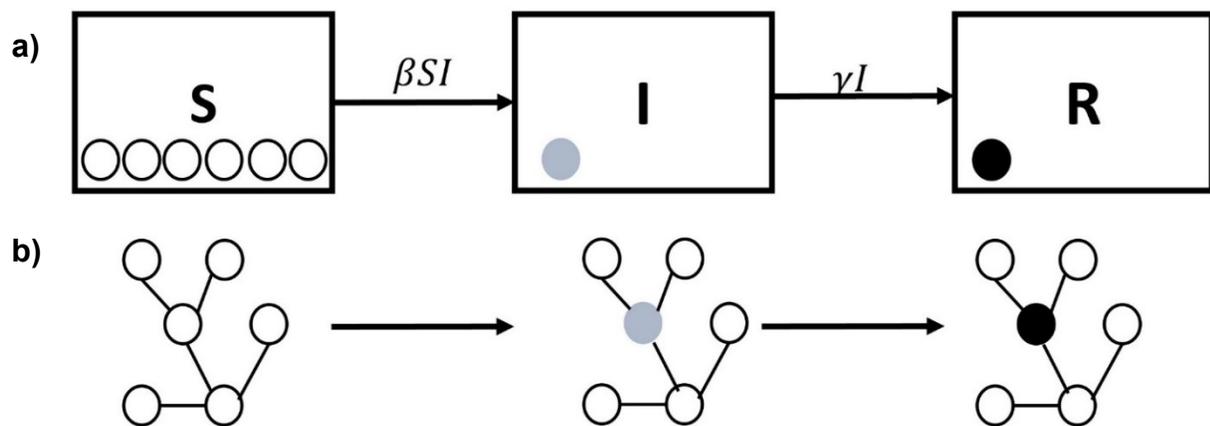


Figure 1.3: Sample disease transmission model. Panel a): Simple compartmental SIR (susceptible-infectious-recovered) model. This model does not contain demographic processes such as birth and death. Individuals in the susceptible compartment (S) move to the infectious compartment (I) at a given rate β . Infectious individuals can recover at a rate γ , after which they transfer to the recovered compartment. Panel b): analogous network epidemic model. The population has an explicit contact structure. Susceptible individuals in the population transfer to the infectious stage (indicated with a grey node), after which they recover (indicated with a black node).

CHAPTER 2

Validation of modified radio-frequency identification tag firmware, using an equine population case study

2.1 Abstract

Historically, contact pattern data has been collected using methods such as observation, surveys, and diaries. More recently, radio-frequency identification (RFID) tags have been used to collect contact pattern data. The RFID system generally consists of RFID tags and RFID readers. The readers, which are powered by an external power source, should be within 10 meters of the communicating tags. The RFID readers are challenging to implement in large agricultural settings due to the lack of a power source, the large area needed to be covered, and inclement weather.

Therefore, the purpose of this study was to validate a modified, active RFID system.

OpenBeacon firmware was modified to use the tag's onboard flash memory for data storage. The tags were deployed within an equine facility which was used as a case study. Tags were attached to the horses' halters, placed in strategic locations around the facility, and were used by facility staff and trainers during the week-long deployment. When the tags came within a pre-defined distance of each other (2 meters), they communicated in a peer-to-peer fashion. The tags recorded the contact event participant identifiers, as well as the start and end times of the contact events. At the end of the study period, the data was downloaded to a computer off-site, where it was analyzed using descriptive network analysis methods.

The resulting networks were biologically plausible and corresponded to the facility schedule. Although the facility had 6 pastures, only 3 were used during the week, with an additional pasture being used on the weekend. This change of protocol was reflected in the networks, confirming the validity of the technology. Most of the batteries held charge for the majority of the study period with only 12% of batteries losing charge before the 6th study day.

2.2 Introduction

Disease transmission models are often used to evaluate the expected impact of disease prevention and control strategies. The models often rely on the assumption of homogenous mixing with each individual having an equal probability of coming in contact with any other individual in the population. This mixing pattern is reflected in the disease transmission term, where disease transmission is dependent on the probability of pathogen transmission given an effective contact and the contact rate (2,3). While this assumption is convenient, it is not appropriate for all populations, such as small or highly structured populations where contacts are not random. In this context, the homogenous mixing assumption does not capture the complexity of the population's contact structure and can result in an over- or under-estimate of the projected impact of the disease control strategies examined in the model (4). One method of correcting this over-simplification is by using data on contact networks to explicitly describe the contact patterns (5).

Contact networks describe the relationships between individuals, or groups of individuals in a population (6). They are often visualized with graphs consisting of nodes

and edges. Nodes represent individuals such as people, places or locations, and edges represent the relationships between the nodes (6). In order to create a contact network, data need to be collected or obtained in such a way that describes a representative portion of the relationships between the individuals in the population of interest. This includes the type of contact, i.e. direct or indirect contact, as well as the duration and frequency of the contact events. Traditional methods for collecting contact network data include the observation and recording of the contacts that occurred during the observation period, journals or surveys in which research participants document the contact events, and more recently, radio frequency identification (RFID) tags (2,7). Each of these methods have different strengths and weaknesses related to the incorporation of biases (e.g. recall bias) and resource requirements (e.g. time and cost), and hence how representative the data is of the true population (7).

The previously described methods of data collection have been used in numerous animal and human studies. For example, Mastrandrea et al., (2015) (7) compared data collected in a French high school using both paper surveys and RFID tags (7). This study reported that contacts of shorter durations were not self-reported using the survey instrument, however, these short duration contacts were recorded by the RFID tags. Other studies which have used RFID technology to collect contact pattern data include hospital studies and animal studies (8–10). In the hospital studies, hospital staff and patients wore RFID tags to track their contacts (8), while in the animal studies, animals were tracked to study mating and contact patterns (9,10).

Although the use of RFID technology for network analysis is relatively new, the technology has a history dating back to World War II. It was used by the British “Identity Friend or Foe system” to distinguish enemy aircraft from their own (11). The technology has evolved over time, and has since been used in electronic article surveillance (late 1960s), tolling systems (1990s), and for identifying objects with the use of RFID labels such as library books or groceries (11).

An RFID system generally consists of two primary components: 1) a reader, and 2) a tag which transmits a signal to the reader. The RFID tags can be described based on the tag’s memory storage capabilities, and the tag’s power source.

In the first categorization scheme, the tags are identified as read-only or read-write tags. Read-only RFID tags have no data storage capabilities. Instead, they contain a unique identifier which is associated with a database containing information about the object. The read-write RFID tag has a memory source and is able to record data about the object (11).

The second type of classification system categorizes the tags based on their power source. Passive tags are activated by an electromagnetic field generated by the RFID reader. In other words, when the tag senses radio waves from the reader, it forms a magnetic field which supplies power to the tag. Active tags on the other hand, are powered by tag-integrated batteries and are able to emit a stronger signal than passive tags. Active tags also use higher radio frequencies as a result of the additional power source (11,12) and do not require the presence of fixed readers.

The use of RFID technology for research purposes has been applied to many disciplines including but not limited to: medicine, engineering, construction, and agriculture (11). This paper focuses on the data collection aspect of network quantification and analysis in an animal agricultural setting using an RFID system manufactured by OpenBeacon (OpenBeacon, Cambridge, United Kingdom (13)).

A typical OpenBeacon deployment protocol is represented in Figure 2.1. The setup includes active RFID tags that, when they come within a pre-defined distance of each other (i.e. 2 meters), exchange signals in a peer-to-peer fashion. If the communicating tags are within 10 meters (m) of a reader, the information exchanged is passed through the reader to a connected computer where the data can be stored.

The traditional deployment setup presents some challenges when implementing the infrastructure within agricultural settings. First, while the OpenBeacon tags contain built-in memory, the memory is not used to store data in the default configuration, creating a need for many RFID readers. Second, the reader, which must be within 10 meters of the tags in order to transmit the data to a computer, must be connected to a power source. While this is easily accomplished in indoor settings where power sources are readily available, it is more challenging in outdoor areas such as pastures or fields due to the large areas that would need to be covered, and the associated perceptions of barn fire risks and/or tripping hazards for animals or people. Furthermore, a large agricultural setting or facility would require the use of many strategically placed readers which could result in some contacts not being recorded, as they might occur out of range of the readers. Lastly, agricultural settings may present more extreme variations

in environmental conditions such as lower temperatures and precipitation, which can impact the operation and maintenance of the system.

In order to overcome these challenges of collecting data in an animal agricultural setting, the main objective of this work was to modify the traditional OpenBeacon tag firmware to utilize the onboard tag storage capability. A study conducted with the modified firmware is used to validate the updated system.

2.3 Methods

This study was reviewed and approved by the University of Guelph's Research Ethics Board (REB #16AP009) and Animal Care Committee (AUP #3518).

2.3.1 System validation procedure

OpenBeacon RFID tags were used to record the contacts that occurred within an equine facility in southwestern Ontario. A contact was defined as an interaction between tags that were within 2m of each other. A distance of 2m was chosen, because it was felt to represent contacts between nodes (horses) in the network that would be considered effective contacts for the possible transmission of an equine respiratory pathogen, e.g. Equine Influenza Virus (EIV), or *Streptococcus equi* also known as the causative agent of "strangles". While equine influenza (EI) can be transmitted over longer distances, for example, by aerosol, wind, and direct contact (14,15), strangles is transmitted through either fomites or respiratory droplets (16–18). To account for both pathogen profiles, we considered an effective contact to occur when individuals came within 2m of each other.

The case study took place over a 7-day period in November 2016. The average outdoor temperature as measured by the closest weather station during the RFID tag deployment ranged from -3 to 12.3 degrees Celsius (Table 2.1). During the study period, there was 0.0 to 13.6 mm of rain, and 0 to 1 cm of snow (Table 2.1) (19). Nine horses resided within the facility and all 9 horses participated in the study. Six people volunteered to participate in the study and included: barn workers, horse trainers, and other veterinary health care providers. All human participants wore the RFID tags on lanyards with the tag encased in a plastic conference badge (Avery Products Corporation, Whitby, Ontario, Canada) during shift hours. Outside of shift hours, the tags were stored in a central location outside the study area with an additional “stationary” tag. All contact events that occurred within the vicinity of the stationary tag were erased during the data-cleaning phase of the study.

Additional tags were placed around the facility at 17 static locations, including the barn doors, indoor riding arena, wash and grooming stalls, and cross ties. Tags were placed on the nosepieces of the horses’ halters which were kept on or near the horses at all times. When a horse was in its stall, the halter was placed on the stall door, and when the horse was in the pasture, the halter was placed on the pasture fence, near the proximity tag that was located at the main gate entrance to the field.

At the end of the study period, a questionnaire was completed by the facility manager. The questionnaire addressed issues such as: the study compliance difficulty, safety with respect to both the human and equine participants, and the invasiveness of the study on the day to day running of the facility.

2.3.2 Active RFID tags

To increase the utility of using the RFID tags in a farm (animal agricultural) setting, the tag firmware was modified to use the tag's 8 MB flash memory (Adesto Technologies, Sunnyvale, California, United States). This eliminated the need to transfer the data directly to a nearby RFID reader connected to a computer. Each tag contained both an internal identifier stored as a 32-bit unsigned integer, as well as an external identifier sticker. When tags came within a predefined distance (in this example, 2m, however distances of 0.5-6m are possible) of each other, the tags recorded the internal identifiers of the interacting tags. The process was repeated when the tags stopped communicating (when the distance between the tags increased to more than 2m), indicating the end of the contact event. A contact event was considered terminated when the tags were out of communication range for approximately 30 seconds. Every contact event that occurred used 16 bytes of memory on each of the tags involved in the interaction. Since the tags have 8MB of flash memory, then given an energy source with sufficient power to store many interactions over long periods of time, the tags can store up to 524,288 interactions, making them suitable for longer study periods with many participants.

The following sections provide details on the tag characteristics, methods of use, and the data storage and processing modifications.

2.3.3 Tag specifications

In the traditional setup, tags communicate with the RFID reader, which is connected to a computer on a network. The incorporation of an RFID reader provides the ability to

assign a date and time to each individual contact event. While this time-keeping mechanism is ideal in the traditional setup, the absence of a reader in agricultural settings makes time keeping more challenging. Individual tags are not manufactured with a real-time clock (RTC), making it impossible to record the actual time at which each contact event starts and ends. To overcome this challenge, the tags were programmed to include a clock substitute that started with the insertion of the coin battery (CR2032). Therefore, the time of each contact event is recorded relative to the insertion of the battery. The calendar date and time of the contact event can be back-calculated using the calendar date and time of the battery insertion.

In the current version of the tag firmware, one observed contact with another tag is recorded every second. For example, when tag A encounters tag B, the tags record this contact. If a third tag (tag C) happened to encounter tag B while tag B was recording the contact event with tag A, the recording action is put on hold momentarily while tag B senses the presence of tag C. The recording action between A and B will continue once tag B is finished sensing tag C, and will also record the contact with tag C. While this limitation affects the order of operations occurring within the tag (the recorded data will not necessarily be sequential), its effect on data quality is mitigated through sorting during the data processing stage.

2.3.4 Setup

The required software for the RFID tags can be run on linux and Apple MacOS operating systems. In addition to the firmware code, which is available upon request, a JLINK tool is required and can be downloaded from:

<https://www.segger.com/downloads/jlink/>. The JLINK tool enables communication between a Nordic board and the computer through a USB connection. In addition, the JLINK tool can be used for debugging purposes which is supported by using the JLINK tool in conjunction with the board.

Lastly, a computer is used for scripting and compiling the code. The code can then be loaded onto the tag chip with a compiler. The tag chip uses the ARM compiler which can be downloaded from: <https://launchpad.net/gcc-arm-embedded>.

The Nordic board and programmer-device (Figure 2.1) are necessary to flash (update the firmware) the developed firmware to the tags. Tags are flashed by attaching the programmer-device to the board and loading the firmware onto the tags with a series of terminal commands. The developed firmware is housed in a GitHub repository and is freely available to interested research groups upon request. Once each tag has been flashed, the tags can be used to collect the described contact data. The updated firmware contains three different functionalities which can be accessed via a button on the tags (see Figure 2.1 for the button location). The three functionalities are: transfer the data from the tags to the computer (one button push), check the tag battery status (two button pushes), and clear the tag memory (three button pushes). The tag's memory needs to be cleared before each subsequent deployment.

The clock starts keeping time as soon as the coin battery is inserted into the tag. Ideally, the batteries should be inserted into the tags immediately before the tag deployment in order to minimize the collection of extraneous data. The calendar day and time of the

battery insertion process should be recorded for use in the post-processing stage. The faster the initial battery insertion process is completed, the more reliable the time synchronicity between the tags will be since the time at which the tag receives power is the time at which the clock begins keeping time. The insertion of the battery also begins the peer-to-peer (a process in which both components have the same functionality, i.e. sending and receiving signals) signal transmission and receiving process. A light-emitting diode (LED) lights up when the battery is inserted correctly.

Once the tags have been deployed (either at specific pre-defined static locations or on individual research subjects e.g. horses), the tags will continue to send signals, receive signals, and record the data until the battery is either removed or is depleted of power. Once the battery is removed, the tags can be connected to a computer via the programmer-device and the data can be downloaded for processing.

After the data has been downloaded, the tag memory should be cleared in order to prepare the tags for the next deployment. The tag memory can be cleared either with or without the tag programmer-device; however, the process is power intensive and uses substantial amounts of battery power. The memory erasure process can take up to 200 seconds, with an average of 80 seconds. This process can be sped up in future firmware versions by resetting the clock to 0. Subsequent data collection would overwrite the previously collected data.

2.3.5 Data storage and analysis

The data is transferred from the tags to a computer where it is stored in a MySQL (Oracle Corporation, Redwood Shores, California) database. The following section describes the data cleaning and processing stages.

The database is composed of multiple tables which contain information about: the trial number, each of the participating tag identifiers (i.e. a source tag (the tag from which the data is being accessed) and a partner tag (the tag that came in contact with the source tag)), the time of the contact event for both the source tag and the partner tag, and a number denoting the status of the contact event, where the status is one of: contact event begins, contact event is ongoing, or contact event ends. The time stamp from the source tag is used to calculate the duration of each contact event. The contact start time is subtracted from the end time in order to obtain the contact duration.

A list mapping the external identifier to the internal identifier of each of the tags was entered into a tag database prior to the tag deployment. The collected data is subsequently compared to this list in order to ensure that there are no aberrant data. Further data filtering is accomplished by entering the study deployment start and end dates and times. Any data that falls outside of this range (i.e. the study period) were discarded during the processing step. This processing step is important as it allows the researchers to remove the tag batteries at a convenient time and place, as opposed to at the original deployment location (e.g. in the case of inclement weather during the completion of the deployment).

Finally, the recorded contact events that occurred in the vicinity of the stationary tag (that was used to track tags that were not being used) were deleted. The data were aggregated by 24-hour periods (the first and last day of the study were less than 24 hours due to the deployment setup and clean up time). Data outputs include seven comma-separated value (CSV) files (one for each day of the deployment). The CSV files contain the internal identifier of the participating tags in each contact event on each respective day, as well as the “weight” of the contact event which represents the total contact duration for each pair of participants throughout the 24-hour period.

R version 3.3.0 (20) was used to conduct a descriptive network analysis, however, similar analyses can also be done using other programming languages. Further analyses can include plotting the contact networks, creating heatmaps to show the duration of contact between each pair of tags, and computing centrality measures (e.g. how important a node was in the network) such as the weighted and unweighted degree (e.g. the number and total duration of contacts a node had).

This paper will focus on validating the collected data during this test deployment with regard to horse-horse, horse-person, and horse-location contacts.

2.4 Results

2.4.1 Technical results

Most of the batteries used in this initial test deployment were operational for the full duration of the study period, however, some batteries lost charge before the end of the study (e.g. during day 6 of the study). In total, 4 out of 33 batteries recorded less than 6

days worth of data (two of the tags were attached to the entrance to the farm, so it is possible that no horses left the farm during this period). Out of the 33 batteries, 29 batteries lost charge within the 6th day of the study meaning that most of the tags were powered for the majority of the study period.

One of the tags assigned to be used by a veterinary health care professional during the tag deployment was not used during the study period and was stored with the stationary tag for the entire duration of the study. Therefore, after the data cleaning process, all data related to the specified tag should have been erased. However, after the data cleaning process, there were still multiple contacts between the unused tag and other tags. This may be due to the tags sporadically restarting, providing gaps of time in which the stationary tag was not communicating with other tags in the vicinity.

2.4.2 Graphs

Sample networks from selected study days can be found in Figure 2.2. The networks represent contacts between horses and pastures. The top panel shows the network of contacts that occurred on day 1 of the study, when the horses were turned out in one of three pastures. The middle and bottom panels show the contacts that occurred on days 4 and 5 of the study, when the horses came in contact with an additional pasture.

It was not stated by the barn manager which horses were moved to the fourth pasture for which days, rather, it was stated that a horse moved to a fourth pasture on the weekend. Figure 2.3 implies that this move occurred on days 4 or 5 of the study. It is unclear why the extra pasture tag was activated on days 4 and 5, but it is possible that

the horses either walked past the tag activating it (pasture 4 is located outside the barn door), or that this is when the horses were moved to the fourth pasture. Regardless, the tag recorded additional contacts on these days, as was expected by the knowledge gained from the facility schedule.

The contact data collected corresponds to the data that one would expect due to the facility structure and scheduling (these additional details were collected using a survey). The networks in Figure 2.2 indicate that horse tags had an increased amount of contact (indicated with a thicker line) with other horse tags that shared a pasture or barn aisle. This corresponds to the large amount of time spent in the pastures (8:20 am - 4:30 pm) and stalls. This trend is repeated daily, owing to the daily similarities in the facility schedule, and further validating the functionality of the tag's sensing and recording abilities.

Lastly, while most of the data appears to make sense from a biological point of view, there are some anomalies. For example, Figure 2.3 shows a heat map of the aggregated weights for each type of tag location (e.g. horse, barn, arena etc.). In other words, the heat map examines the total duration of contact between, for example, all horses and all pastures. From the figure, it is clear that at least one tag in the barn came in contact with the tag in the wash stall. Neither tag was within 2m of each other during the deployment. This is the only inconsistency of its kind, as the remainder of the data appears to be biologically plausible given the facility's daily schedule and the expected contacts. Interference, whether constructive or destructive is a common occurrence and is to be expected in radio transmission systems.

In general, over the course of the study period, horses had the most contact with each other, followed by horses and the barn (this includes the barn door and cross ties). People came in contact with most areas in the facility, however, they had minimal contact with the pastures. This is likely due to the fact that the contact between people and pastures was only recorded when they entered and exited the pasture.

2.5 Discussion

The evolution of RFID technology has provided opportunities to further research in all disciplines, including that of mathematical epidemiology. Although OpenBeacon RFID technology has been used in previous studies to collect contact data (21–24), this is the first study to make use of the onboard storage for data collection and to specifically describe the methods and tools developed for this type of research deployment. This adaptation of the traditional setup allows for the extension of the OpenBeacon RFID tags from primarily indoor settings, to outdoor settings, extending the range of potential interdisciplinary research in animal agricultural and veterinary settings.

Although the network size used in the pilot study was small, it provided the opportunity to test and validate the updated RFID firmware and data processing procedures.

Similar contact patterns were seen on each day, corresponding to similarities in the daily farm schedule. Although some differences exist in the contact network data collected, these differences are likely due to stochasticity in human behaviour. This stochasticity can impact aspects such as the horse's daily schedule (e.g. when they are exercised) and contacts (e.g. which other horses are in the vicinity of the cross ties being used). The biological implications of the study are discussed in Chapter 3.

With respect to the updated RFID technology, the tags functioned well with some imperfections. The batteries did not survive for the full seven days (some batteries fully discharged within the 6th day). Although this might be partly due to the cold temperatures, this problem has persisted in other subsequent studies that were conducted on larger facilities in warmer temperatures (Chapter 4). Furthermore, for unknown reasons, the tags do on some occasions appear to sporadically restart. While this error has the potential to cause a loss of data, the restart takes minimal time, thus minimizing the impact to the data. This discrepancy in the data is further exaggerated by the inability for multithreading and the lack of a real-time clock. However, this is in the order of milliseconds, resulting in minimal data loss. Lastly, due to the nature of radio-wave technology, participating tags might record slightly different contact durations. This can be attributed to different reasons including the tags ability to sense a limited number of tags at a time, and various forms of natural and artificial interference that can be anticipated in any type of radio transmission. A review of the recorded data from both the source and partner tags identified minimal differences between the tags' output. Therefore, given the data aggregation period of 24 hours, a heuristic was used to consider the original data as collected by a single tag, however, methods of compensating for data inconsistencies include averaging the recorded times from the participating tags, or using the maximum time recorded between the participating tags. Although limitations exist, it is important to consider them in the context of the study as a whole. First, the data was aggregated by 24-hour periods. Therefore, minor inconsistencies in time have minimal effect on the overall network description. Second,

alternate methods of data collection can be time consuming and costly and may not accurately reflect the entire network. For example, diaries and surveys tend to focus on the ego network (the network of the individuals in question), as opposed to focusing on the relationships within the population as a whole (25). Further biases include recall and non-response bias. Similarly, data collection by observation has the potential to miss observations and is subject to human error. In comparison to these alternate methods of data collection, the framework for collecting contact data using RFID tags described here has the potential to quantify a more reliable and comprehensive set of data. However, multiple data collection methods can be used to cross validate the data.

Third, RFID tags are becoming increasingly inexpensive and uncomplicated to deploy. The time component of the data collection involves preparing the tags, setting up the deployment, downloading the data, and clearing the tag memory at the completion of the deployment. The post-study questionnaire indicated that the RFID tags resulted in minimal intrusion into the daily activities of the farm. Horses appeared to be unaware of the tags, and from the point of view of the participants, the main difficulty was remembering to wear the tags and to move the horse halter from location to location (if it was not normally done).

Overall, we found the modified tags to provide reliable, biologically plausible data, which would be difficult to obtain using the traditional RFID system deployed in an agricultural setting or by using other methods. The tags provided an easy, safe, non-invasive, and non-intrusive means of collecting data that can be used to further network analysis research in animal agricultural settings.

References

1. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci.* 2015;370(1669):20140107–20140107.
2. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface.* 2007;4(16):879–91.
3. Najafi M, Laskowski M, de Boer PT, Williams E, Chit A, Moghadas SM. The effect of individual movements and interventions on the spread of influenza in long-term care facilities. *Med Decis Mak.* 2017;37(8):871–81.
4. Perkins SE, Cagnacci F, Stradiotto A, Arnoldi D, Hudson PJ. Comparison of social networks derived from ecological data: Implications for inferring infectious disease dynamics. *J Anim Ecol.* 2009;78(5):1015–22.
5. Craft ME, Caillaud D. Network models: An underutilized tool in wildlife epidemiology? *Interdiscip Perspect Infect Dis.* 2011;2011:1–12.
6. Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One.* 2015;10(9):e0136497.
7. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, van den Broeck W, et al. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS One.* 2011;6(2):e17144.
8. Ji W, White PCL, Clout MN. Contact rates between possums revealed by proximity data loggers. *J Appl Ecol.* 2005;42(3):595–604.
9. Walrath R, Van Deelen TR, VerCauteren KC. Efficacy of proximity loggers for detection of contacts between maternal pairs of white-tailed deer. *Wildl Soc Bull.* 2011;35(4):452–60.
10. Domdouzis K, Kumar B, Anumba C. Radio-frequency identification (RFID) applications: A brief introduction. *Adv Eng Informatics.* 2007;21(4):350–5.
11. Weinstein R. RFID : A technical overview and its application to enterprise. *IT professional.* 2005;7(3):27–33.
12. Bitmanufactory. Open hardware for open beacons [Internet]. [cited 2017 Apr 5]. Available from: <https://www.bitmanufactory.com/>
13. Myers C, Wilson WD. Equine influenza virus. *Clin Tech Equine Pract.* 2006;5(3):187–96.
14. Davis J, Garner MG, East IJ. Analysis of local spread of equine influenza in the park ridge region of Queensland. *Transbound Emerg Dis.* 2009;56(1–2):31–8.
15. Waller AS, Sellon DC, Sweeney CR, Timoney PJ, Newton JR, Hines MT.

- Streptococcal infections. In: *Equine Infectious Diseases*. Second Edi. Elsevier Inc.; 2014. p. 265–77.
16. Barnett C, Brumbaugh G, Holland R, Lunn P, Vaala W, Voss E, et al. Guidelines for vaccination of horses. American Association of Equine Practitioners. Lexington, Kentucky, USA.; 2001.
 17. Sweeney CR, Timoney JF, Newton JR, Hines MT. Streptococcus equi infections in horses: Guidelines for treatment, control, and prevention of strangles. *J Vet Intern Med*. 2005;19(1):123–34.
 18. Government of Canada. Daily data report for November 2016 [Internet]. [cited 2017 Nov 27]. Available from: <http://climate.weather.gc.ca>
 19. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.r-project.org/>
 20. Voirin N, Payet C, Barrat A, Cattuto C, Khanafer N, Régis C, et al. Combining high-resolution contact data with virological data to investigate influenza transmission in a tertiary care hospital. *Infect Control Hosp Epidemiol*. 2015;36(3):254–60.
 21. Machens A, Gesualdo F, Rizzo C, Tozzi AE, Barrat A, Cattuto C. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect Dis*. 2013;13(1):185–200.
 22. Barrat A, Cattuto C, Colizza V, Pinton JF, Van den Broeck W, Vespignani A. High resolution dynamical mapping of social interactions with active RFID. *Arxiv Prepr arXiv08114170*. 2008;1–16.
 23. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*. 2011;6(8):e23176.
 24. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin Microbiol Infect*. 2014;20(1):10–6.

2.6 Tables

Table 2.1 Summary weather statistics during the study week as recorded by a weather station in proximity to the facility (19).

Day	Max. temp. (°C)	Min. temp. (°C)	Ave. temp. (°C)	Total rain (mm)	Total snow (mm)
1	14.5	-1.0	6.8	0.6	0.0
2	11.0	3.0	7.0	0.4	0.0
3	13.0	-1.0	6.0	0.0	0.0
4	20.5	4.0	12.3	0.4	0.0
5	9.0	0.0	4.5	13.6	0.0
6	0.5	-3.5	-1.5	0.0	0.0
7	-0.5	-5.5	-3.0	0.0	1.0

2.7 Figures

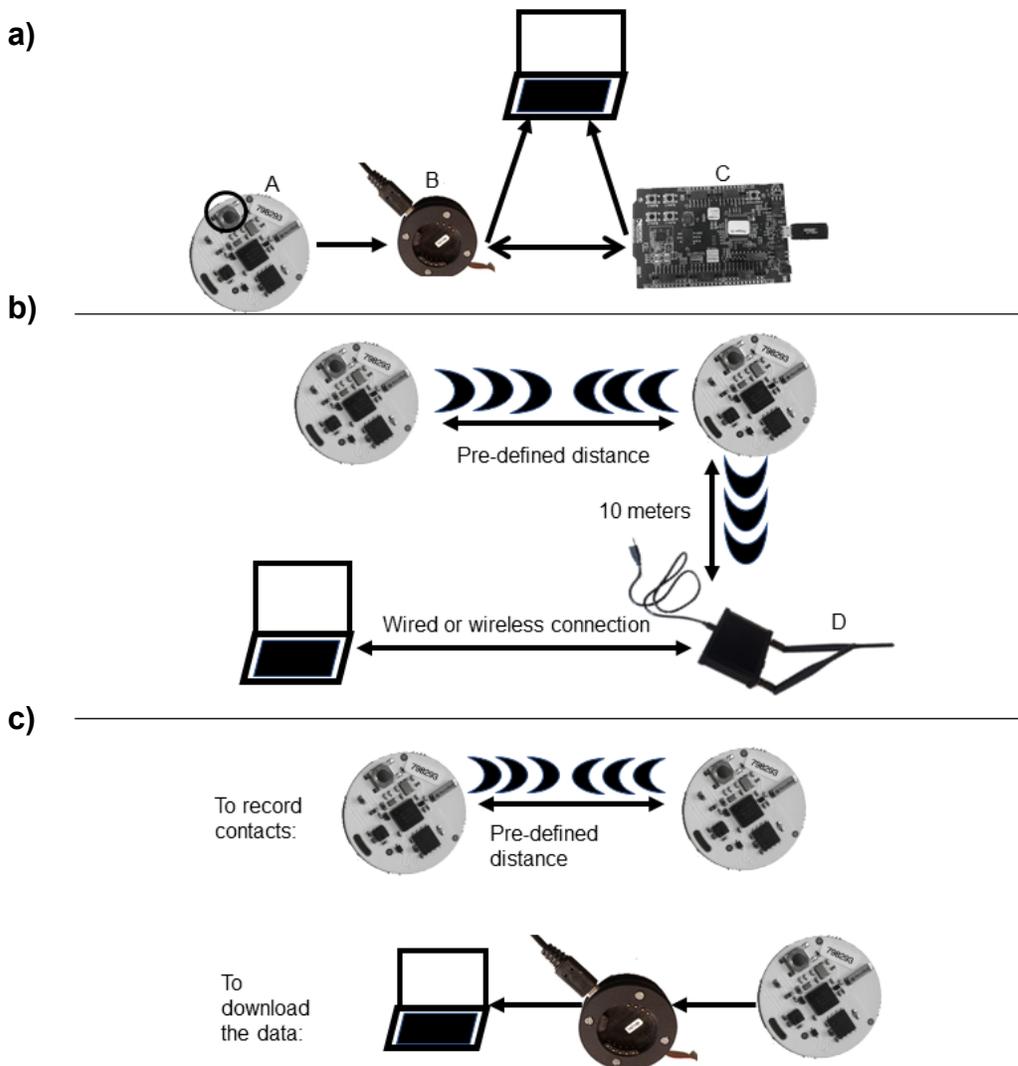


Figure 2.1:The RFID tag system and components. Panel a) shows the components used to flash the tags. The components required to flash a tag are: a computer, a tag (labelled A), a programmer-device (labelled B), and a Nordic board (labelled C). Panel b) shows the traditional RFID setup. In addition to a computer and tag, the communicating tags must be within 10 m of the reader (labelled D) in order to transmit the data to the computer. Panel c) depicts the modified system, in which the tags operate independently of the computer (top line). The data is downloaded to the computer with the use of the programmer-device. The button on the tag is indicated with a black circle in the top panel.

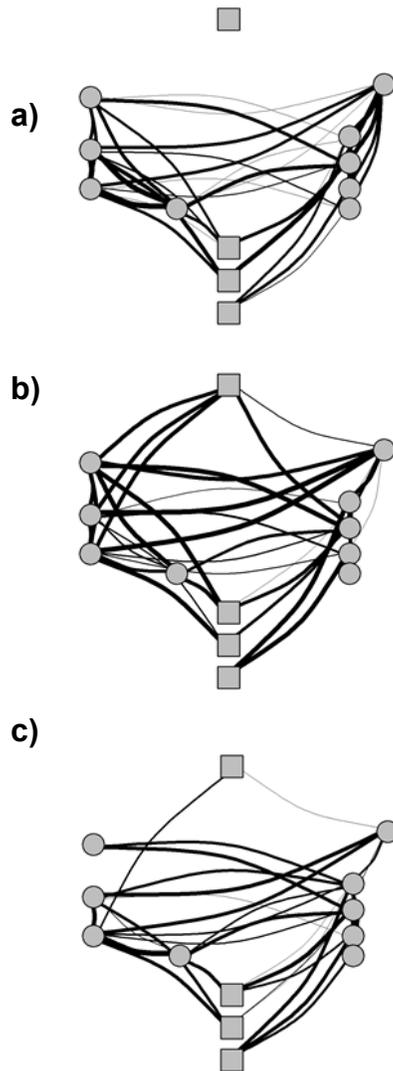


Figure 2.2: Selected horse and pasture contact networks. The horses are represented by circular nodes and the pastures are represented by square nodes. Panel a) represents the contacts that occurred on day 1 of the study, panel b) represents the contacts that occurred on day 4 of the study, and panel c) represents contacts that occurred on day 5 (Saturday) of the study. The thickness of the edges between each pair of nodes represents the contact duration for the respective day. Therefore, a larger edge represents a longer contact duration.

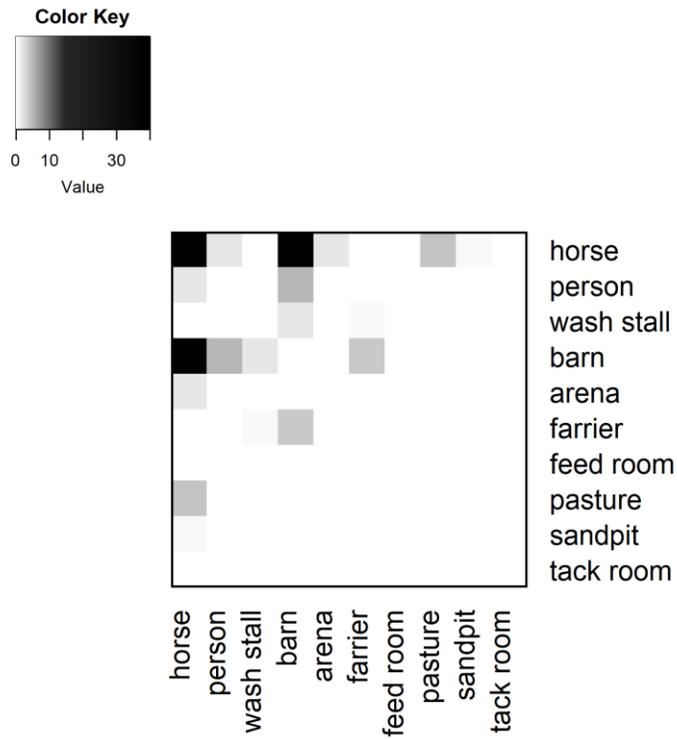


Figure 2.3: Sample heat map of aggregated tag contacts. The heat maps show the amount of contact that tags (based on individual tag assignment) had with each other. Darker colours indicate a higher duration of contact (hours) throughout the day.

CHAPTER 3

Quantification of contact patterns using radio-frequency identification proximity tags on an equine facility in Ontario, Canada: a pilot study

3.1 Abstract

Disease transmission models are often created on the assumption of homogenous mixing, which has the potential to incorrectly represent the contact patterns of the population. One method of correcting for this oversimplification is by using contact network data to inform the contact structure of the model. Different methods exist for collecting the data required to create a contact network, including radio-frequency identification (RFID) tags. The implementation of the traditional RFID system is challenging for outdoor areas, due to the requirement of an external power source. Therefore, a modified system in which the data is stored in the tags' onboard memory was used to collect and store contact pattern data between horses on an equine facility.

The RFID tags were deployed at an equine facility in Ontario over a 7-day period. The tags were attached to the horses' halters, which were kept with the horses at all times. At the end of the study period, the tags were collected and the data was downloaded to a computer at a different location. The data was analyzed using traditional network analysis methods including the computation of selected centrality measures (weighted and unweighted degree, and the eigenvector centrality). Other analyses included quantifying the ability to generalize the data over the study period, assessing the ability

to predict the contact patterns without the use of the tags, and assessing the assumption of homogenous mixing.

The modified tags provided a feasible method of collecting contact data in large outdoor areas. The study resulted in the collection of large amounts of contact data which were aggregated into 24-hour periods. Few trends were observed in the centrality measures. In particular, the horse that had the highest eigenvector centrality also had the highest weighted degree. This horse varied throughout the study week. The contact data collected did not support the homogenous mixing assumption, and was similar throughout the study period, resulting in a high similarity index. Lastly, a static network in which horses that shared a barn or pasture were assumed to come in contact, provided relatively similar networks to those created with the data collected from the tags.

3.2 Introduction

In 2016, it was estimated that there were 291,561 horses in Canada, across 39,164 farms (1). This number, reported from the Agricultural Census provides an estimate of the Canadian equine population size. However, the data do not provide specific details on the types of movements or contacts that occurred between the horses. The Livestock Identification and Traceability program in Canada captures some movement data from a number of agricultural animals including some cattle and pigs, however, it has not been implemented for the equine industry (2).

The collection and reporting of quantitative data regarding contact and/or movement patterns for the Canadian equine population is gaining interest. A study by Spence et al., (2017) reported contact pattern data that was collected using a survey instrument to document movement patterns of horses that attended an Equine Canada sanctioned show in 2014 and this area of work is ongoing (Spence et al., in press). Although some quantitative data is starting to populate the peer-reviewed literature, there remains a lack of data related to the types of contact that occurs within individual equine facilities.

In Ontario, Canada, horse movement on an individual facility is generally scheduled (e.g. turnout, feed times, and training are routine activities). The specific routine typically depends on the facility and the primary equine discipline of the facility (i.e. racing vs. riding horse) and anecdotally it is thought that the type of discipline will subsequently affect contact patterns. On a typical facility, horses spend a portion of their day turned out in a pasture that they may or may not share with other horses. Additionally, horses are generally assigned a stall in the barn, which may or may not be in the vicinity of their pasture mates. However, there are also opportunities for contact to occur outside of routine, scheduled barn activities, for example, when a horse is cross-tied in the barn aisle for a veterinary examination or exercised in a common area. These types of activities can result in variable contacts with other horses over time, resulting in the heterogeneity of contact patterns on a facility.

In general, heterogeneity in contact patterns can arise from variation in the number of contacts, variation in the specific individuals contacted, and variation in the duration and frequency of the contacts (3). Contact heterogeneity can impact a population in a

multitude of ways. For example, in the case of infectious disease transmission, direct contact has the potential to contribute to the transmission of a pathogen from an infectious individual to a susceptible individual contacted (4). The role of contact on the spread of a pathogen within a population varies based on the mode of pathogen transmission, for example, respiratory infections such as equine influenza can be transmitted through close contact such as nose-to-nose contact (4,5).

Contact heterogeneity within a population can be represented as a contact network consisting of nodes and edges (6). Nodes represent individuals or locations in the population of interest, and edges represent the interaction between the individuals (nodes) (7). Both the nodes and edges can have characteristics that represent heterogeneity, such as contact duration or frequency. Edge and node characteristics and relationships can be used to calculate network metrics such as those describing the importance of individuals in the network (e.g. weighted and unweighted degree, and eigenvector centrality) (6). Example network properties are displayed in Figure 3.1.

Data for contact networks are often collected and displayed as a static representation of a system, since time-varying contact data can be challenging to collect (8). Static models are appropriate representations for systems that are thought to have minimal variability over time such as livestock movement over a single season (4). However, for systems that tend to exhibit changes, such as the changes in livestock movement across different seasons, static representations can either over- or under-represent the contact events (4). Such variation in contact events arises from the natural variability in human and/or animal behavior and is often unpredictable. In these situations, it is more

realistic to model the population's contact structure using a dynamic network such as the example in Figure 3.1.

While contact networks can be an effective method of describing the contacts that occurred within a population, they can be challenging to generate. This challenge includes defining a contact and obtaining sufficient data to create a representative network (4). Various methods have been used to collect data about contact events within populations including diaries or surveys, direct or indirect observation, and radio-frequency identification (RFID) tags (4,9). Each data collection method has strengths and weaknesses which can result in different biases (Table 3.1).

Active RFID tags have been used to collect contact pattern data for both human (10–13) and animal populations (14–16). Human studies have focused on using RFID tags to quantify contact patterns amongst school children, as well as hospital medical staff and patients (8,11,12). Animal studies, have used RFID tags to study the behavioral and mating patterns of animals (14,15), such as observing the contact between white tailed deer dams and their offspring (14).

The contact networks generated using the data collected with RFID tags can be used to inform disease transmission models (17). Disease transmission models can be used to study the effect of different biosecurity measures which are measures taken to prevent the introduction and spread of infectious diseases in livestock settings (18) such as vaccination and quarantine procedures (19). The assumption of homogenous mixing, where each individual has an equal chance of coming in contact (20), is commonly used

in disease transmission models, however, due to the possible variation in contacts, this assumption may not be realistic for many populations, including equine populations.

Although RFID tags have successfully been used in past studies to collect contact data, they have yet to be used within facilities housing livestock species such as horses.

Equine facilities, as well as other outdoor livestock systems, present an additional challenge for data collection because the accessibility to electrical power can be limited or non-existent. Due to the requirement of electrical power needed to operate the RFID reader in the traditional OpenBeacon system, the original firmware was modified to eliminate the need for an external power source when using the RFID tags. This study represents a test case to examine the overall utility of the updated RFID tags when deployed within an equine facility. Specifically, the objectives of this study were to:

- 1) Quantify the dynamic equine contact network within an equine facility located in southwestern Ontario;
- 2) Determine if the measured contact network satisfies the assumption of homogenous mixing;
- 3) Assess the possibility of generalizing contact patterns, by day, within the study period;
- 4) Quantify the need for proximity logging technology for identifying contact patterns within an equine facility; and to,
- 5) Assess the utility of using active RFID tags to collect equine contact data in an agricultural setting.

3.3 Methods

This study was approved by the University of Guelphs' Research Ethics Board (REB #16AP009) and Animal Care Committee (AUP #3518).

3.3.1 RFID tags

Active RFID tags (OpenBeacon, Cambridge, United Kingdom (21)) (Figure 3.2) were used to collect equine contact data on a single equine facility over the course of a 7-day period. The devices detect and record contact information such as participant identifiers (ID) and the contact event times when tags come within a defined distance of each other. In the conventional configuration, the RFID tags send the collected contact data to an RFID reader which in turn transmits the data to a storage device such as a computer (22). Since the conventional RFID tag setup could not be implemented in the farm setting for this study, the tag firmware was modified to make use of the previously unused onboard memory for data storage. In brief, the modifications to the firmware enabled the storage of data on the tags which subsequently provided the ability to download the data at a different location at a later point in time. Therefore, the current study was used to test the utility of the updated RFID tags. All of the contact data were collected and stored on the modified RFID tags and then downloaded to a laptop computer at the end of the collection period. All data were stored in a MySQL database.

3.3.2 Pilot study

The study took place on an equine facility in southwestern Ontario, Canada over a 7-day period in November 2016. All horses in the facility resided in a single building, hereafter referred to as a "barn". Horses were trained and exercised in both outdoor

and indoor training areas (Figure 3.3). The facility boarded 9 horses, 4 of whom were used for lessons. The remaining 5 horses were used for dressage competition and field hunting.

All 9 resident horses were enrolled in the study. The tags were placed in a protective plastic cover (Avery Products Corporation, Whitby, Ontario, Canada) which was sealed closed with duct tape (Logix Insulated Concrete Forms, Port Hope and Cobourg, Ontario, Canada). In order to maximize the likelihood that contacts would be recorded, the tags were attached to the nose piece of the horses' halters (Figure 3.2) with vet wrap (3M™, United States). The tags are unable to send or receive transmissions through body masses (e.g. horse skull). Therefore, placing the tags on the nose piece of the halter enabled transmissions between horses standing side by side. In addition, the placement of tags on the nose piece meant that close proximity nose-to-nose contacts (e.g. contact sufficient to transmit a respiratory pathogen) would be recorded.

The horses' halters were kept with the assigned animal at all times. However, in order to ensure that the horses did not chew on the tags, the halters were removed from the horses and kept on a horse-inaccessible hook on the pasture fence when the horses were turned-out in the pasture. In order to maximize participation and minimally disrupt facility activities, the tags remained on the halter and were not switched between halter and bridle when horses were being exercised. Hence, the halters were placed on a hook near the mounting block when the horses were being exercised in the riding or training area. Each horse's halter was stored on the front of the stall door when the horse occupied the stall in order to comply with the standard operating procedures of

the facility. Furthermore, the removal of the halter when the horse was in the stall, allowed the tags to capture contacts when other horses were in the center aisle, within 2 meters (m) of the stall door. The horses' regular daily schedules were carried out, including turnout and training sessions.

A survey completed by the barn manager captured information regarding each horse's stall and pasture location, as well as their primary use and schedule, e.g. training, competition discipline, etc.

Upon completion of the data collection period, the tags were collected from the facility. This process involved removing the batteries at the tag locations in order to stop the signal transmission process.

3.3.3 Data analysis

The data collected at the facility were transferred off of the tags to a computer and then downloaded to a MySQL (Oracle Corporation, Redwood Shores, California, USA) database. The data from the database were aggregated into 24-hour time periods by adding together the duration of all contacts between two individuals for each day of the study. The data captured during the initial set-up and the tag collection on the last day of the study were discarded as these data were considered to be non-representative of the daily operations of the facility. The data were exported as comma separated value (CSV) files (one for each day of the study deployment), and imported into R version 3.3.0 (23) for analysis with the igraph package (24).

Horse contact patterns were compared for each day of the study to form a dynamic representation of the network over the 7-day study period. The descriptive measures calculated for each day of the study include the network density and the clustering coefficient. The network density represents the number of contacts in the network out of the total number of possible contacts. The clustering coefficient represents the number of groups of triangles out of the total possible number of triangles, where triangles represent 3 individuals who are all in contact (6). The clustering coefficient is an important metric to consider in social network analysis as it describes a population's propensity to form groups, thereby influencing the possible spread of diseases within the network (25,26).

The horse's relative importance in the network was examined using the centrality measures: unweighted and weighted degree, and eigenvector centrality. These measures were chosen based on their ease of applicability to the transmission of diseases through the facility. The unweighted degree represents the total number of unique contacts a horse had with other horses and was calculated by summing the number of edges incident to the respective node (7). Therefore, the degree can range from 0 to $N - 1$, where N represents the total number of nodes in the network.

The weighted degree represents the total duration of contact a horse had with all other horses in the network on each day of the study. The weighted degree, hereafter referred to as "strength", was calculated by summing the weight of all edges incident to the respective node (27). The minimum possible strength is 0 (for an isolated horse), while the maximum possible strength varies based on the number of contacts each node had

with other nodes, and the duration of each contact. Therefore, the maximum possible strength can surpass 24 hours, as a horse may come in contact with many other horses for long periods of time.

Lastly, the eigenvector centrality is similar to the degree, however, it takes into account the adjacent node's centrality, thus presenting a more comprehensive idea of second degree contacts a horse may have (28). The eigenvector centrality can range from 0 to 1, where a higher value indicates a more connected node.

The contact patterns were visualized with network diagrams. In the networks, nodes (representing the horses) are connected with an edge representing the contact between the horses. The width of the edge varies according to the duration of contact between the two horses on the respective day, with wider edges representing a longer duration of contact. For ease of visibility, the edge weights were log transformed and multiplied by 2. Heat maps were created to visualize the duration of the contact between horses, as well as to compare the amount of contact between horses in different aisles of the barn.

The Jaccard similarity index ($\frac{|A \cap B|}{|A \cup B|}$) compares the presence and absence of contacts between individuals in each network (29). The index was used to address the objective of assessing the generalizability of the contact patterns within the study period.

Specifically, the Jaccard similarity index was computed to compare the presence and absence of contacts between each horse on each day of the study. The resulting similarities were studied graphically using a heat map in which the cell color density

represents the similarity in contacts between each pair of days. Darker colored cells represent a greater similarity in contacts than lighter colored cells.

To address the objective of assessing the possibility of predicting the contact patterns without the use of RFID tags, the facility manager was asked to complete a questionnaire. Specifically, the questionnaire asked questions regarding the stall and pasture locations for each participating horse during the study week. This information was used to create a network (hereafter referred to as a “survey-based network”) in which it was assumed that all horses that shared a pasture or barn came in contact with each other. In the event that a horse was moved between pastures during the study week, the pasture in which the horse was turned out for the majority of the week was used in generating the survey-based network. The Jaccard similarity index and a classification table were used to compare the survey-based network to a weeklong network created by combining the data from each study day. Specifically, in the classification table, the network specified by the data collected with the tags was used as a gold standard, and the survey-based network was used as a predictive test. The classification table was used to calculate the sensitivity (sn), specificity (sp), and positive and negative predictive values (PPV and NPV respectively) of the survey-based network.

3.3.4 Study Assumptions

Some simplifying assumptions were made during this study. First, the tags were programmed to include a washout period of 30 seconds. This means that if two previously interacting tags failed to detect each other for 30 consecutive seconds, the

tags recorded that the contact event had ended. The premature termination of a contact event has direct implications for the role of contact in the transmission of diseases, both in terms of contact frequency and contact duration. The washout period impacts the duration of the contact event, for example, in the case that the horses changed position and the tags were unable to communicate, cutting the contact event short, and the frequency of the contact events, when, for example, a horse may be temporarily moved interrupting the tag-to-tag contact. In this case, the contact would be counted as two contact events instead of one.

A realistic definition of an effective contact was required to parameterize future network models (Chapter 5). A distance of 2m was chosen based on the current knowledge of common equine infectious pathogens and the authors' longer-term goal of modeling equine strangles and equine influenza in future work. While equine influenza likely has a larger transmission distance than 2m (30), strangles is often spread through fomites, and requires direct contact (31). We therefore chose the definition of a contact to be when horses come within 2m of each other as it represents a reasonable contact distance for both strangles and equine influenza transmission.

Lastly, the halters were kept on the pasture fence when the horses were turned out (in order to assure that the horses did not chew on the tags) (Figure 3.2). Since horses are social creatures, we assumed that when the horses were in their pastures, they would congregate, and therefore, to account for this contact, all halters were kept together on the pasture fence.

3.4 Results

3.4.1 Quantifying contact patterns within the facility

Similar contact patterns were seen on each day of the study and therefore, only results for three days of the study are presented in the main body of the text (Figure 3.4). A comprehensive set of results is presented in the supplementary documentation (Figures S3.1-S3.3). For each day of the study, edges can be seen between horses that resided in the same aisle of the barn, as well as between horses that were turned out in the same pasture. In general, the edge width is thicker, implying a greater duration of contact between horses that shared a pasture. The contact duration differed by day and is displayed in both the graphs and heat maps (Figure 3.4). Although there was contact between horses in the different aisles, visually, there were no obvious trends that appeared related to the duration of contact between the barn aisles (Figure 3.4).

The network density for each study day varied, with a minimum density of 0.50 and a maximum density of 0.75. The clustering coefficient ranged between 0.59 and 0.75 throughout the study period (Table 3.2). The network centrality indicates that the horse with the highest degree varied throughout the study week, both in terms of the individual horse and in terms of the horse's main use and aisle (e.g. teaching, pleasure riding etc.). The degree centrality ranged from 2-8 with the average daily degree ranging from 4 to 6 (Figure 3.5).

The horse with the highest strength and eigenvector centrality also varied over the study period, however, one horse in particular had the highest strength and eigenvector centrality on 4 of the 7 study days. The horse with the highest strength also had the

highest eigenvector centrality for each study day. The horses' strength, ranged from ~0.63 hours to 41 hours. The eigenvector centrality ranged from less than 0.01 to 0.70 (Figure 3.5).

3.4.2 The assumption of homogenous mixing

The heat maps used to assess the homogenous mixing assumption for each study day indicate that horses did not come in contact with all the other horses over the course of the week (Figure 3.4). Furthermore, the duration of each contact event between horses was inconsistent (Figure 3.4). The contact patterns within this facility do not appear to satisfy the assumption of homogenous mixing.

The weeklong static network does not suggest homogenous mixing and is similar in shape to each of the individual, day long heat maps (Figure 3.6).

3.4.3 Generalizability of the contact patterns

All days shared at least 53.6% of identical contacts. Days 2 and 7 were the most similar with ~73.9% of identical contacts followed by days 2 and 6 which had ~72.0% of identical contacts. Days 4 and 6 were the least similar with 53.6% of identical contacts (Figure 3.7).

3.4.4 Comparison of the tag-acquired network to the survey-acquired network

The survey-based network and the week-long tag-acquired network were reasonably similar. The tag-acquired networks shared 97.2% of contacts with the survey-acquired network. When the survey acquired network was compared to each day of the tag acquired network, the most similar networks had ~75% of common contacts and

occurred on day 1 of the study. The smallest similarity between the survey acquired network and the tag acquired network occurred on day 7 of the study with 50% of similar contacts, while days 1-4 all had at least 61% of similar contacts (Figure S3.4).

The classification table resulted in a high sensitivity ($sn = 1$) and specificity ($sp = 0.82$) when comparing the survey-based network to the week-long tag-acquired network. The positive and negative predictive values were 0.97 and 1.0 respectively.

When comparing the survey-based network to each day of the tag-acquired network, the classification table resulted in high sensitivities ($sn = 1$) and low specificities ($sp \leq 0.33$) for each study day (Table S3.1). Furthermore, both the positive predictive values and negative predictive values were relatively high (≥ 0.5 and 1 respectively) (Table S3.1).

3.5 Discussion

This study fills an important gap in the current knowledge regarding contact patterns between horses within an equine facility in southwestern Ontario. The study goals were to address the current lack of information regarding equine contact patterns using different data collection methods and network analysis tools. This work has demonstrated the utility of using modified OpenBeacon RFID tags to quantify contact patterns within an equine facility. While network analysis has previously been used to study animal movement and contacts, such as cattle movement during the 2001 foot and mouth disease epidemic in the UK (6,32,33), traditional methods of data collection such as surveys, databases, and observation are more common methods for collecting

this type of data. As tools for quantifying heterogeneity in contact patterns are becoming more advanced and less cost-prohibitive, they represent opportunities for obtaining a more reliable and comprehensive set of contact data in different types of settings (34). This in turn, enables the parameterization of increasingly realistic disease transmission models which can better inform decisions regarding improved disease prevention and control strategies (34).

Although the advancement in technology enables the quantification of more detailed contact networks, it also raises questions regarding how to best utilize the increased quantity of data (34). This study represents a starting point for understanding the utilization, limitations, and potential of using modified RFID tags in animal agricultural settings, and provides an improved understanding of the strengths and limitations of the data that they provide.

While there were no observable trends with respect to the centrality measures, it is important to note that the degree centrality ranged between 2 and 8. The degree range is important as it helps to describe the range of outcomes that could potentially arise from the introduction of an infectious disease into the study population (26). If the infectious horse has a degree of 2, then, at most the infectious horse can infect two other horses, potentially limiting the spread of the disease. Alternately, if the degree of the infectious horse is 8, then there is a higher likelihood of the disease spreading to the entire study population. Thus, both the degree distribution and the position of the initially infected horse in the network have the potential to define the range of outcomes on the network.

A similar outcome can be inferred from the eigenvector centrality measure. While the eigenvector centrality ranges from >0.00 - ~ 0.70 , if a horse with a higher eigenvector centrality was to become infectious, it could have catastrophic effects on the adjacent horses in the network, which also have relatively high degrees.

Trends in the contact patterns visualized in the networks, as well as the heat maps, were similar throughout the study period. This is to be expected as horses are generally kept on a specific schedule which includes feed times, vet and farrier visits, turnout, and training. Chance events due to human interaction with the horses may cause the schedules to vary resulting in stochasticity in the networks from day to day. The study was conducted in November which is not a busy competition month in Ontario, hence there was little to no movement off the facility. The observed contact patterns are likely to change during the height of the competition season (summer months) when horses are constantly entering and leaving the facility and using the outdoor trails.

Horses tended to come in contact with other horses in their aisle of the barn, as well as horses with whom they shared a pasture. The variation in contact duration between horses and between days is a strong indicator that contact patterns within the facility did not conform to the homogenous mixing assumption. Therefore, using the homogenous mixing assumption to parameterize a disease transmission model for this population would grossly under- or over-estimate contact events and any resulting recommendations for improved biosecurity practices.

The question of contact pattern generalizability depends on the research question and the required data sensitivity. In this study, ~53% of contacts were repeated throughout the study period, and approximately 50% of contacts were found in the survey-based network. This is important as it demonstrates that approximately half of the contacts that occurred are predictable. Furthermore, the ability to generalize from the data allows the required length of an empirical study to be shortened, reducing the time spent at the facility, as well as the costs of conducting the study.

The comparison of a survey-based network in which each horse comes in contact with other horses in its barn, as well as horses with whom it shares a pasture, yielded >50% of similar contact events when compared with the tag acquired networks. Furthermore, the classification table indicated that the survey-based network was able to correctly predict all the contacts that occurred ($sn = 1$). However, the low specificity ($sp \leq 0.33$) indicates that the survey-based network over-estimated contacts, as it failed to identify contacts that did not occur. The high sensitivity and low specificity are likely due to the barn management style. While the horses were turned out in 3 different pastures, they resided in a single barn. Therefore, the survey-based network represented a complete network in which all horses came in contact. The complete network structure of the survey-based network led to high negative predictive values ($NPV = 1$) indicating that the survey-based network correctly identified individuals who did not come in contact. Furthermore, the positive predictive values were relatively high indicating that the majority of contacts that did occur were represented in the survey-based network. This

was supported by the Jaccard similarity index which was identical to the positive predictive values.

While the contact pattern predictability results are important in terms of data inference without the use of tags, they come with the following caveat; the Jaccard distance index and the classification table were computed on binary networks. Therefore, the results do not provide insight into the duration or frequency of the contact events. If the transmission of a disease is dependent on the duration or frequency of contact events, then the predictability results are inapplicable in the study of contact patterns and their effect on disease transmission. However, if a single contact event is sufficient for the transmission of a pathogen, for example, in the case of highly transmissible infectious diseases, then the predictability results provide a good approximation of the contact rate. This outcome can then be used in a disease transmission model to study realistic outcomes in the presence of various prevention and control strategies without the use of technology such as RFID tags.

While this is the first study that aims to quantify equine contact patterns using active RFID tags in an animal agricultural setting, similar studies have been carried out with humans in other settings such as schools and hospitals (9,10). The study of equine contact patterns is in some ways comparable to the school and hospital studies as in each example; the day-to-day activities of the participants are highly structured and well defined. Schools, hospitals, and equine facilities are run on a schedule that is generally repeated over the course of a time span (e.g. daily schedules of activities, events, and movements). For this reason, some studies infer contact patterns for longer periods of

time based on data collected for a short period of time (12). The assumption of the generalizability of contact patterns over a period of time, while used in previous studies, has been substantiated by the current study for this study facility. The confirmation of the ability to generalize the contact patterns is important as it enables researchers to shorten the required study period. The ability to shorten the required study period has many implications such as minimizing different selection and information-based biases arising from participant compliance (e.g. forgetting to wear the tag) and minimizing the participant time requirement. Furthermore, in the case that it is not possible to conduct a complete or longer deployment, the contact patterns can be generalized based on minimal data. Although the lack of complete data may impact the accuracy of the results resulting from the disease transmission model, the ability to generalize the results to obtain some idea of the heterogeneous mixing patterns still provides an improvement compared to the assumption of homogeneous mixing.

Technological advancement raises questions regarding how best to incorporate such vast amounts of data into network studies. For example, the current study which was conducted on 9 horses over a 7-day period resulted in 211,906 collected data points over the study period. The data generated with RFID tags can be intensive and time consuming to analyze and interpret. Possible methods for dealing with large volumes of data include aggregating the data by various time periods, e.g. day long intervals, or by participant type (e.g. horse vs. horse owner). These methods and outcomes are discussed in detail in the paper by Machens et al., (2013), and should be further investigated for different study settings. In their study, Machens et al., (2013), used

data collected in a pediatric hospital with RFID tags to study the effect of different data aggregation schemes. The authors explored the effect of 6 different aggregation schemes on the results of a disease transmission model. The resulting disease transmission models show different results for the different aggregation schemes, implying that an aggregation scheme should be chosen with the population and population characteristics in mind. In the current study, the data was aggregated into 24-hour intervals representing days. The horses in the study facility spent a large portion of their day in their respective pastures and spent the night in their assigned stall. Thus, while using smaller aggregation intervals such as hourly aggregation intervals might provide different study outcomes in this population, it would likely have minimal effect on overall study results.

3.5.1 Benefits and limitations

Certain limitations exist with respect to the study design. First, the facility represented a small network size (N=9). However, this study served as a pilot study that was used to refine the technological aspects of the data collection methodology. With the exception of the initial study set-up and deployment, and tag collection at the end of the deployment, the methodology requires minimal researcher involvement, limiting interruptions in the daily running of the facility.

While the potential lack of researcher involvement in the day-to-day running of the study limits certain biases such as the Hawthorne Effect, in which participants act differently than they usually would due to being observed (35), it provides the opportunity for other biases to arise. Non-compliance with the study design, such as forgetting to hang the

halter in the correct location could cause different forms of information bias, such as misclassification bias, which could affect the results by incorrectly recording contacts that never occurred (or occurred to a lesser degree) or failing to record contacts that did occur.

A second source of potential misclassification arises from the study design. When the horses were in their pasture, the halters and tags were placed on a halter hook on the pasture fence. The data collected during this time would indicate that the horses in the pasture were all within 2 meters of each other. While horses tend to be social creatures, the assumption that the horses congregate while in the pasture might be incorrect, causing the collection of contacts that never occurred.

The tags are unable to communicate through body mass such as a horse's skull. Hence the tags were placed on the nose piece of the halter. Therefore, a contact event was only recorded if the contact occurred when the horses were within 2 meters of each other and were face to face or side by side (with their faces next to each other). While this precludes non-face to face contacts such as nose-to-tail contacts, which are bound to occur, the authors' interest is related to directly transmitted pathogens which are spread via respiratory droplets, in particular, equine influenza and strangles. These pathogens are both expelled and contracted through the respiratory tract. Hence, this limitation has minimal effect on our overarching objective of creating a network to be used in a network model.

One of the study objectives was to assess the suitability of active RFID tags for the collection of contact data in animal agricultural settings. As previously mentioned, the traditional RFID tag setup involves the use of power-operated RFID reader devices. In order to use the RFID tags in this study, the tag firmware was modified to make use of the onboard data storage. The firmware is freely available from the corresponding author upon request. This modification enabled the collection of data over the weeklong study period when horses were both in the barn and outside. For the most part, this implementation was successful for the collection of contact data, however there was an issue related to the tag batteries holding a charge for the duration of the study period. This issue may be exacerbated by colder temperatures. Overall, however, the implementation of RFID tags within the equine facility proved to be easy and convenient to carry out in a setting with limited power in the outdoor areas.

3.5.2 Future work

Future work building on the results of this pilot study should consider including multiple and larger equine facilities as well as different facility types (racing vs. riding facility). The collection and analysis of data from multiple, and/or larger facilities would enable more comprehensive comparisons of the network characteristics allowing for the further identification of similarities and differences that exist in the quantified equine contact patterns in different settings. This will subsequently provide more detailed information on equine contact patterns and enable improvements to be made in the parameterization of disease transmission models for horses. Overall, the resulting

disease transmission models could be used to infer more relevant biosecurity recommendations for equine facilities in southwestern Ontario.

3.6 References

1. Statistics Canada. Table 004-0224 - Census of Agriculture, other livestock on census day, every 5 years [Internet]. CANSIM (database). [cited 2017 Oct 27]. Available from: <http://www5.statcan.gc.ca/cansim/>
2. Canadian Food Inspection Agency. Canadian Food Inspection Agency [Internet]. Requirements for Livestock Producers: Livestock Identification and Traceability Program. 2016 [cited 2017 Apr 27]. Available from: <http://www.inspection.gc.ca/>
3. Bansal S, Read J, Pourbohloul B, Meyers LA. The dynamic nature of contact networks in infectious disease epidemiology. *J Biol Dyn*. 2010;4(5):478–89.
4. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci*. 2015;370(1669):20140107–20140107.
5. Myers C, Wilson WD. Equine influenza virus. *Clin Tech Equine Pract*. 2006;5(3):187–96.
6. Dubé C, Ribble C, Kelton D, McNab B. A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound Emerg Dis*. 2009;56(3):73–85.
7. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis*. 2009;56(4):109–20.
8. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*. 2011;6(8):e23176.
9. Mastrandrea R, Fournet J, Barrat A. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One*. 2015;10(9):e0136497.
10. Voirin N, Payet C, Barrat A, Cattuto C, Khanafer N, Régis C, et al. Combining high-resolution contact data with virological data to investigate influenza transmission in a tertiary care hospital. *Infect Control Hosp Epidemiol*. 2015;36(3):254–60.
11. Barrat A, Cattuto C, Tozzi AE, Vanhems P, Voirin N. Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin Microbiol Infect*. 2014;20(1):10–6.
12. Gemmetto V, Barrat A, Cattuto C. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis*. 2014;14(1):695–705.
13. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, van den Broeck W, et al. Close encounters in a pediatric ward: Measuring face-to-face proximity and

- mixing patterns with wearable sensors. *PLoS One*. 2011;6(2):e17144.
14. Walrath R, Van Deelen TR, VerCauteren KC. Efficacy of proximity loggers for detection of contacts between maternal pairs of white-tailed deer. *Wildl Soc Bull*. 2011;35(4):452–60.
 15. Ji W, White PCL, Clout MN. Contact rates between possums revealed by proximity data loggers. *J Appl Ecol*. 2005;42(3):595–604.
 16. Lavelle MJ, Fischer JW, Phillips GE, Hildreth AM, Campbell TA, Hewitt DG, et al. Assessing risk of disease transmission: Direct implications for an indirect science. *Bioscience*. 2014;64(6):524–30.
 17. Stehlé J, Voirin N, Barrat A, Cattuto C, Colizza V, Isella L, et al. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Med*. 2011;9(1):87.
 18. Nordmann BD. Issues in biosecurity and biosafety. *Int J Antimicrob Agents*. 2010;36(SUPPL. 1):S66–9.
 19. Nöremark M, Widgren S. EpiContactTrace: an R-package for contact tracing during livestock disease outbreaks and for risk-based surveillance. *BMC Vet Res*. 2014;10(1):71.
 20. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface*. 2007;4(16):879–91.
 21. Bitmanufactory. Open hardware for open beacons [Internet]. [cited 2017 Apr 5]. Available from: <https://www.bitmanufactory.com/>
 22. Cattuto C, van den Broeck W, Barrat A, Colizza V, Pinton JF, Vespignani A. Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One*. 2010;5(7):e11596.
 23. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.r-project.org/>
 24. Gabor C, Tamas N. The igraph software package for complex network research. *InterJournal* [Internet]. 2006;Complex Sy:1695. Available from: <http://igraph.org>
 25. Schank T, Wagner D. Approximating clustering coefficient and transitivity. *J Graph Algorithms Appl*. 2005;9(2):265–75.
 26. Eubank S, Guclu H, Kumar VS, Marathe M V, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. *Nature*. 2004;429(6988):180–4.
 27. Opsahl T, Agneessens F, Skvoretz J. Article node centrality in weighted networks:

- Generalizing degree and shortest paths. *Soc Networks*. 2010;32(3):245–51.
28. Borgatti SP, Everett M, Johnson JC. *Analyzing social networks*. SAGE Publications Limited; 2013.
 29. Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. *J Am Soc Inf Sci Technol*. 2007;58(7):1019–31.
 30. Davis J, Garner MG, East IJ. Analysis of local spread of equine influenza in the park ridge region of Queensland. *Transbound Emerg Dis*. 2009;56(1–2):31–8.
 31. Gutierrez MPA. Strangles : the most prevalent infectious respiratory disease in horses worldwide. *CES Med Vet y Zootec*. 2013;8(1):143–59.
 32. Webb CR. Farm animal networks: Unraveling the contact structure of the British sheep population. *Prev Vet Med*. 2005;68(1):3–17.
 33. Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *J R Soc Interface*. 2006;3(10):669–77.
 34. Machens A, Gesualdo F, Rizzo C, Tozzi AE, Barrat A, Cattuto C. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect Dis*. 2013;13(1):185–200.
 35. I. Dohoo, W. Martin HS. *Veterinary epidemiologic research*, 2nd Edition. University of Prince Edward Island; 2014.
 36. Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *Science (80-)*. 2009;323(5916):892–6.
 37. Salathe M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci*. 2010;107(51):22020–5.
 38. Craft ME, Caillaud D. Network models: An underutilized tool in wildlife epidemiology? *Interdiscip Perspect Infect Dis*. 2011;2011:1–12.
 39. Ruiz-Garcia L, Lundadei L. The role of RFID in agriculture-applications, limitations and challenges. *Comput Electron Agric*. 2011;79(1):42–50.

3.7 Tables

Table 3.1: Summary of strengths and weaknesses of selected contact pattern collection methods.

Method	Strengths	Weaknesses
Observation	<ul style="list-style-type: none"> • Less susceptible to error than survey data (36) 	<ul style="list-style-type: none"> • Difficult to observe all individuals simultaneously (4)
Diaries and surveys	<ul style="list-style-type: none"> • Allows the for the specification of the type and frequency of the contact events (9) 	<ul style="list-style-type: none"> • Recall and non-response bias (9) • Focused on the ego network (11) • Costly (9) • Often have a small sample size (37)
RFID* tags	<ul style="list-style-type: none"> • Relatively inexpensive and easy to use (9) • Detect close-range face-to-face encounters (22) 	<ul style="list-style-type: none"> • Technological constraints (e.g. power supply) (38) • Radio-technology limitations e.g. radio-wave interference (39) • Only provide information about the individuals wearing a tag (9)

* RFID: radio-frequency identification

Table 3.2: Descriptive measures for the contact network of horses residing on a single equine facility in southwestern Ontario, Canada. Both the density and clustering coefficient were always larger than 0.5.

	Day								
	1	2	3	4	5	6	7	Mean	Standard Deviation
Density	0.75	0.61	0.61	0.61	0.56	0.58	0.50	0.60	0.08
Clustering coefficient	0.75	0.63	0.64	0.61	0.59	0.62	0.59	0.63	0.05

3.8 Figures

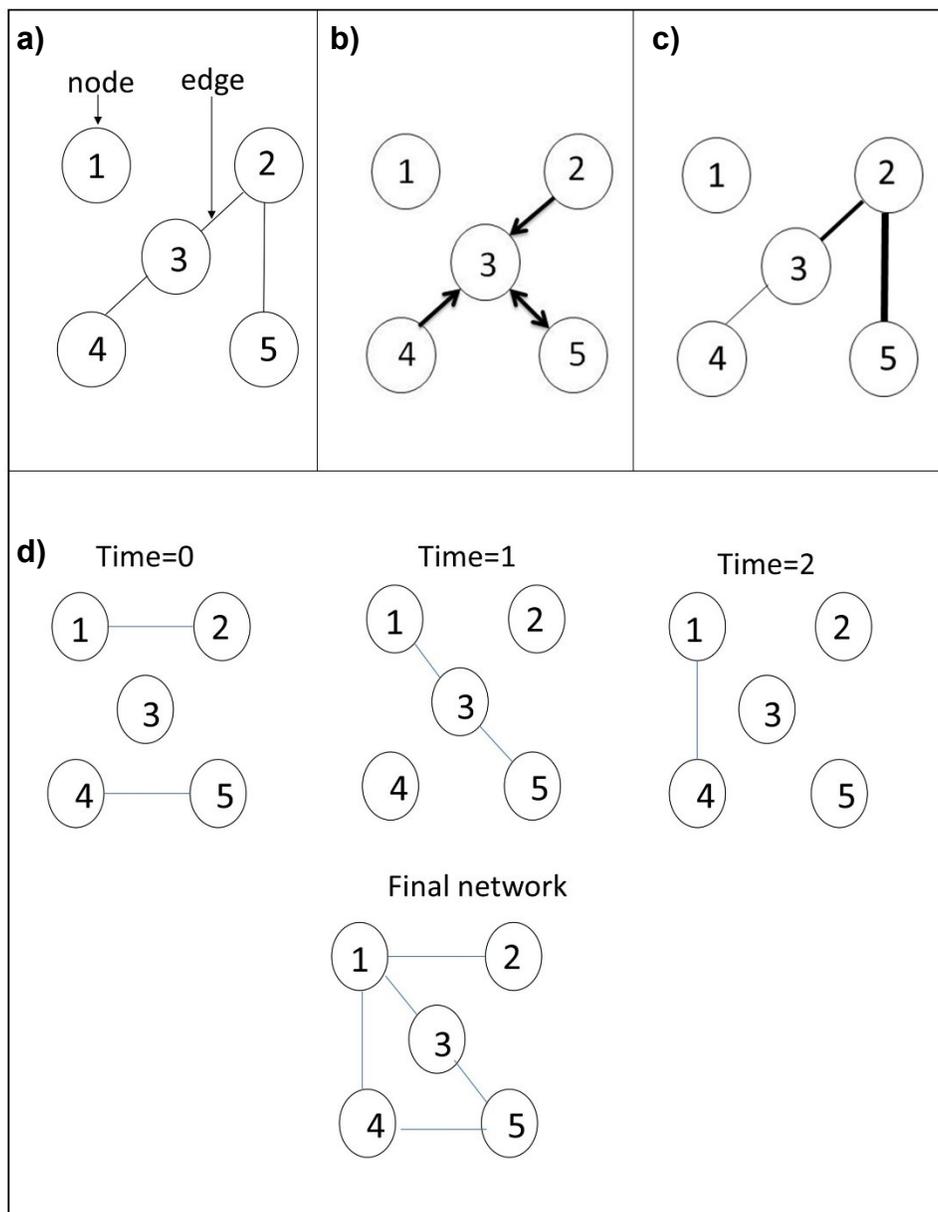


Figure 3.1: Graphical representations of contact network characteristics. Panel a) shows a simple, non-directed, non-weighted network. Panel b) shows a directed network. In this network, the contact originates at one node (e.g. node 4) and goes towards another node (e.g. node 3). Panel c) contains a weighted network in which edges with larger weights (larger line widths) imply more contact between the respective nodes. The network in panel d) shows both a dynamic network over a three-day period, as well as a static representation of the same network, marked as “final network”.

a)



b)



c)



Figure 3.2: OpenBeacon RFID proximity tag with onboard storage (a), horse wearing a tag on his halter (b), and halters with a tag on the nose piece hung on a pasture fence (c).

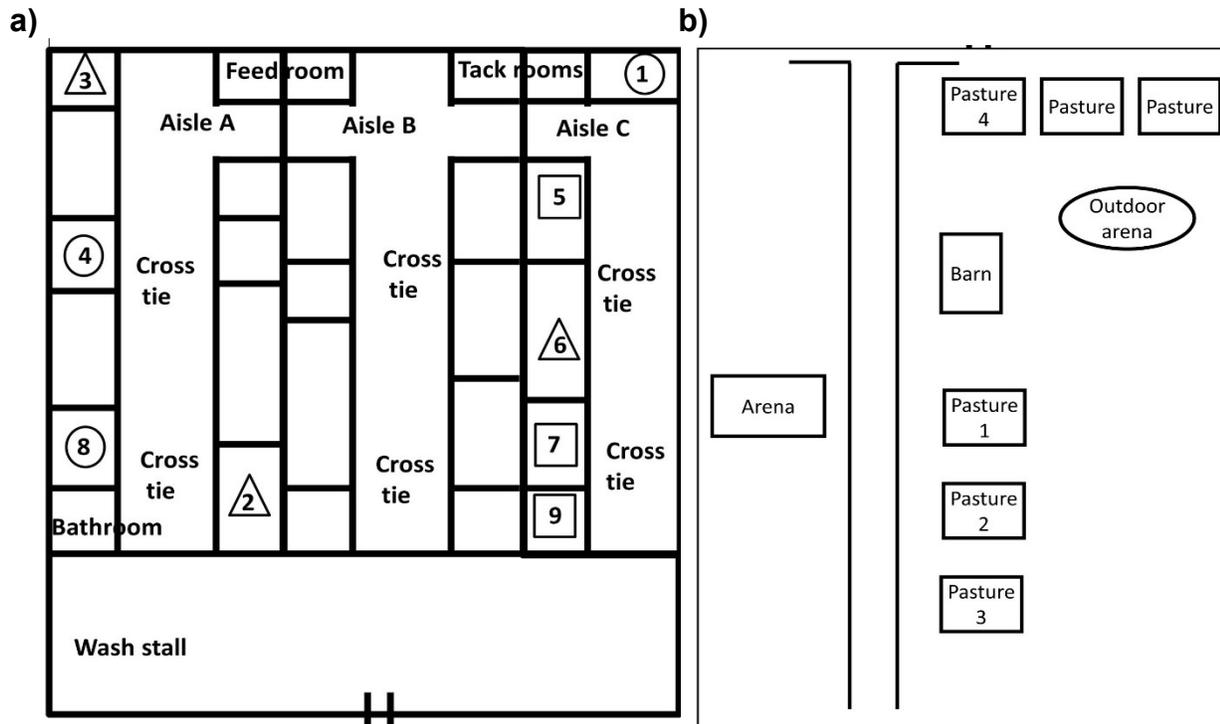


Figure 3.3: Map of the barn (a) and facility (b). Horses are represented by the numbers in the stall. The shape surrounding the numbers signifies the pasture in which the horse was turned out. The triangle represents pasture 1, the circle represents pasture 2 and the square represents pasture 3. The barn was separated into aisles, which were used to construct heat maps to quantify the amount of contact between the horses in the different aisles (Figure 3.4). The horses were turned out in pastures 1-3, except for on the weekend when at least one horse was moved to pasture 4. Each pasture shape in the barn contains a number which refers to the horse IDs in the heat maps in Figure 3.4.

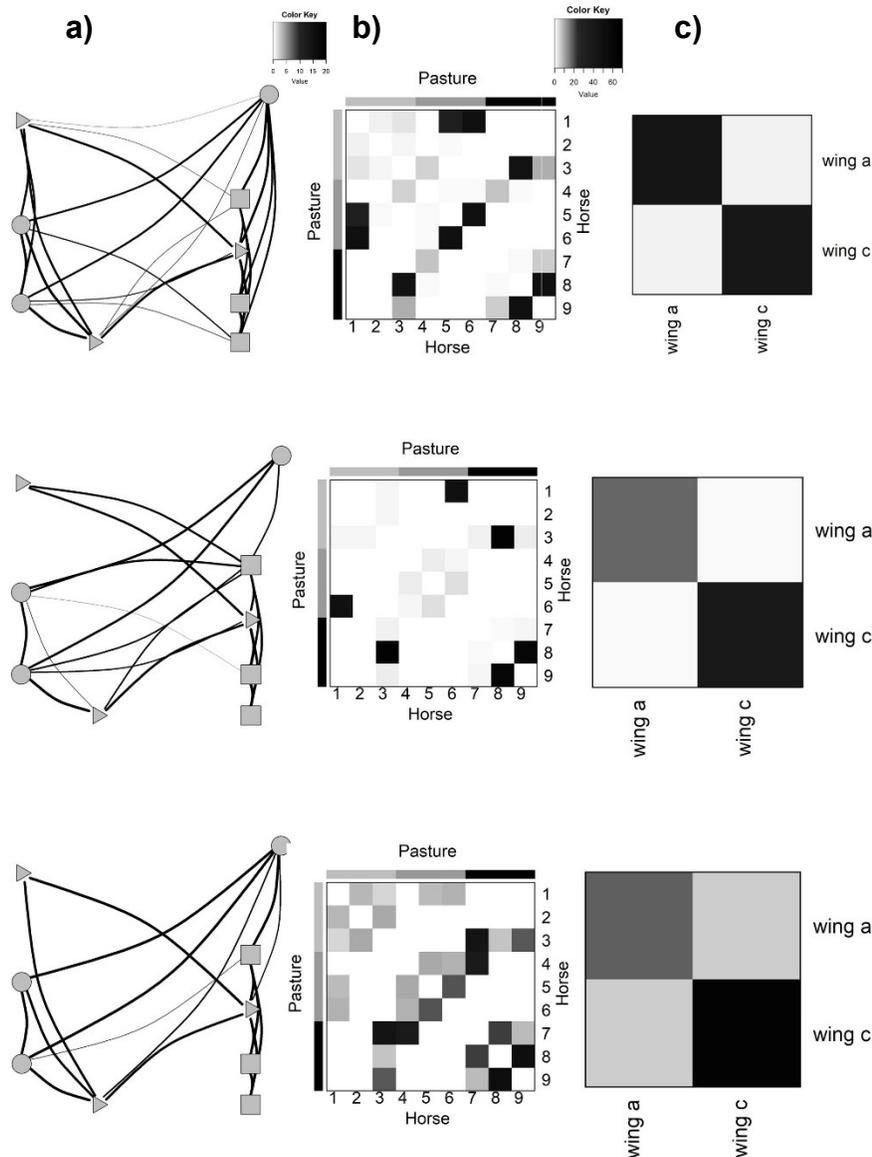


Figure 3.4: Summary study results. Networks in which horses are laid out in relative locations of the stalls in the barn (Figure 3) are found in (a). Horses' nodal shapes are coded by pasture number. The edge width represents the contact duration ($\log(\text{seconds})^2$) between respective horses. (b) shows heat maps of the duration of contact (hours) throughout the day. Horses are numbered in conjunction with the numbering system in Figure 3. Lastly, (c) shows heat maps for the contact between horses in each wing. Only three days of results are shown (day 1 (top row), day 5 (middle row), and day 7 (bottom row)), as the results were similar for each day of the study. The full set of figures can be found in the supplementary information in Figures S3.1-S3.3.

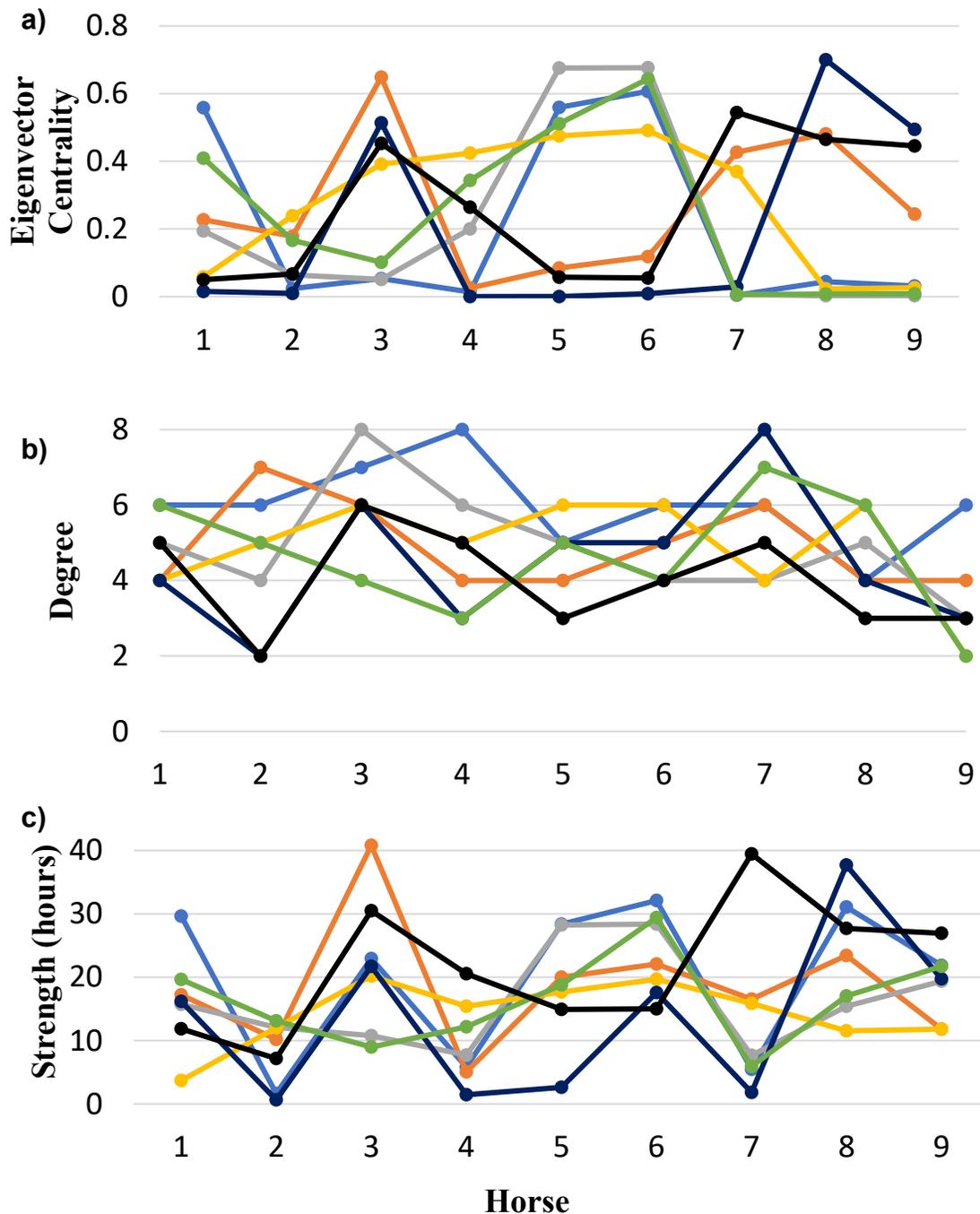


Figure 3.5: Selected centrality measures for each horse on each study day. Centrality measures include eigenvector centrality (a), degree (b), and strength (c). Each day, a different horse was the most central horse (day 1=light blue, day 2= orange, day 3=grey, day 4= yellow, day 5=dark blue, day 6= green, day 7= black). The eigenvector centrality ranged between 0.0 and 0.7, while the degree and strength ranged between 1 and 8, and 0 and 40 respectively. No trends were evident in the centrality plots.

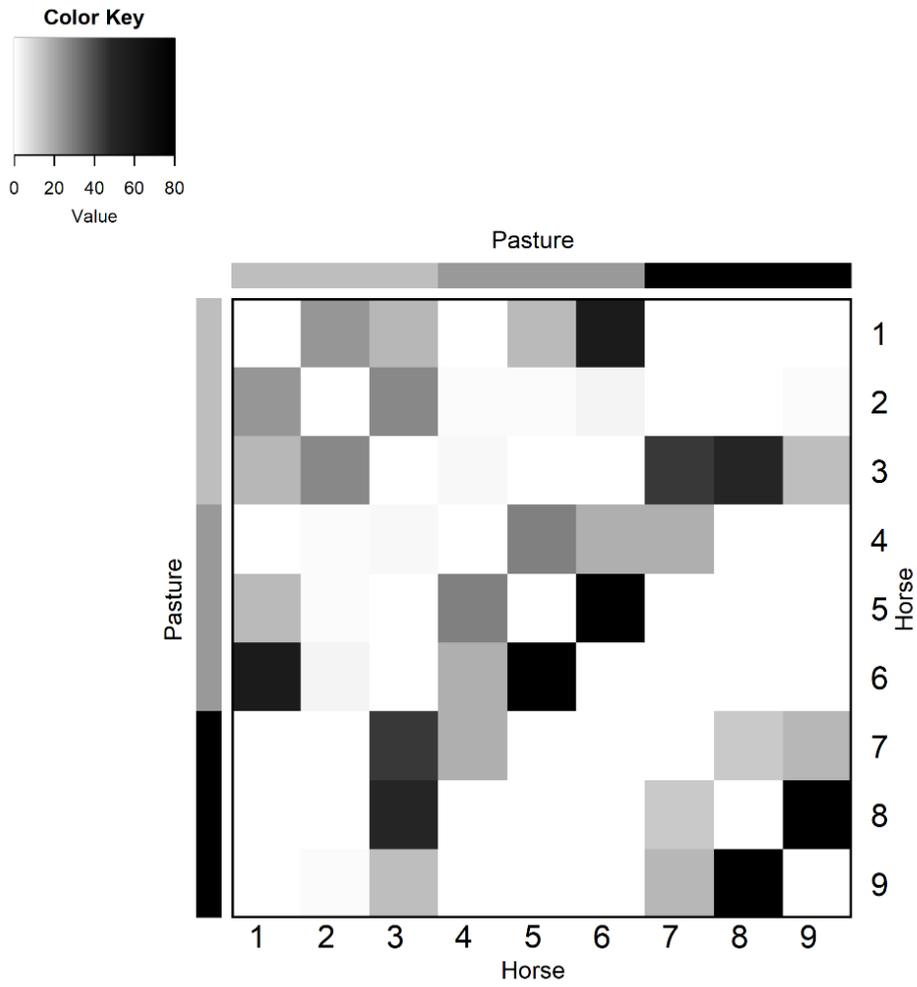


Figure 3.6: Heat map of the 7 combined dynamic networks. The heat map is not suggestive of homogenous mixing. The colored panels on the left and top axes represent the different clustering of horses by pasture.

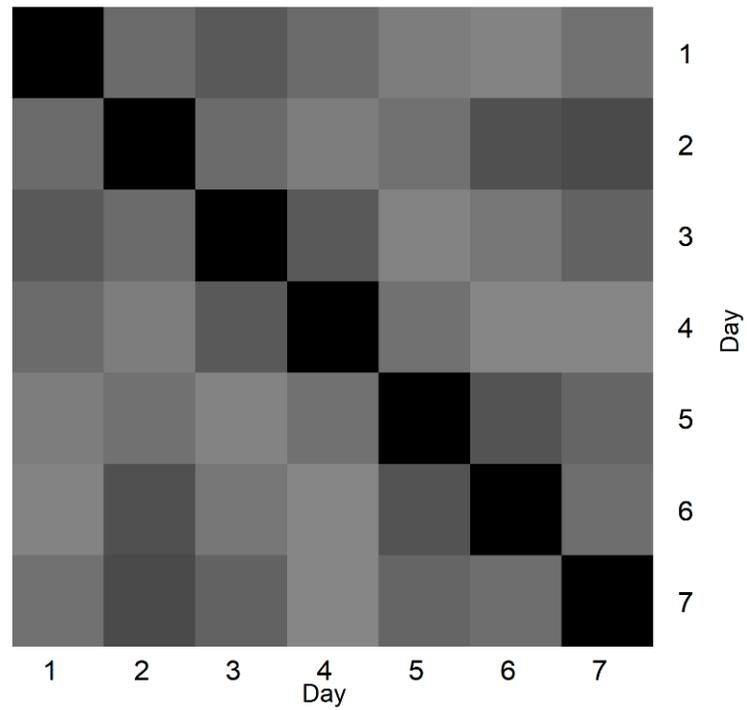
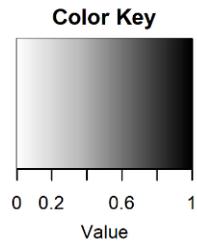


Figure 3.7: Heat map of the Jaccard similarity index. The heat map compares the similarity in contacts between participants on each day of the study. All days share at least ~54% of identical contacts.

CHAPTER 4

Comparison of the dynamic networks of four equine boarding and training facilities

4.1 Abstract

Contact networks can be analyzed to assess the potential for disease spread throughout the network. The lack of Canadian facility-level equine contact data makes the characterization of the equine contact structure difficult. Therefore, the purpose of this study was to use empirical contact data to characterize and compare equine network characteristics between equine facilities in Ontario. Contact pattern data from 4 equine facilities were collected using radio-frequency identification tags. The collected data were used to form 7 static contact networks (1 for each study day) for each facility. The assumption of homogenous mixing, where each individual in a population has an equal probability of coming in contact, was assessed for each network, since homogenous mixing is often used to describe mixing patterns in disease transmission models. At the facility level, neither the day-long static networks, nor a combined, week-long network were representative of homogenous mixing. The Jaccard Similarity Index indicated that 11-62% of the contacts were repeated throughout the study period. A network generated with survey-based data enabled the prediction of 8.7-79.6% of the contacts that were recorded with the RFID tags. With respect to the node centrality, the normalized node degree ranged from 0.0 to 0.96, with a mean of 0.31. The node strength ranged from 0 to 1 with a mean of 0.38. For both the node degree and node strength, nodes that had a high centrality score relative to other nodes always had a

high centrality score relative to the other nodes in the network and vice versa. A significant ($p < 0.05$), weak positive correlation existed between the node degree and strength ($0.41 < r < 0.54$). The normalized betweenness centrality ranged from 0.00 to 1.00, with a mean of 0.11. Lastly, an exponential random graph model was used to quantify the relationship between the distance between the horses' stalls and edge formation. The distance parameter was not significant for all of the facilities. To conclude, the non-homogenous nature of the contact patterns, coupled with the large range of the centrality measures indicate the importance of using empirical data to understand processes such as disease spread potential within equine populations. Although the collection of a full set of data is optimal, the study results suggest an ability to infer contact networks using observational data in situations where little-to-no data exist. This study serves as a starting point for the characterization of equine contact networks in Ontario.

4.2 Introduction

The underlying contact structure between individuals in a population is highly associated with processes such as disease transmission (1,2). Given the relationship between contact networks and the processes of interest, the use of accurate and representative networks is essential to the system analysis and the resulting outcomes. Network accuracy largely stems from data quality and quantity. However, in animal agricultural populations, within-farm contact patterns are seldom recorded (3). This scarcity of data extends to Canadian equine population and makes the classification of equine contact patterns through network analysis challenging.

Historically, the collection of comprehensive datasets that describe the interactions between individuals (human or animal) has been challenging due to the cost and difficulty of collecting the data (4). More recently, however, the use of radio-frequency identification (RFID) tags has enabled the collection of high resolution data (5). This technology is relatively inexpensive, requires minimal time and intervention from the researcher, and is capable of collecting large amounts of contact data. Modifications conducted by the authors to the OpenBeacon RFID system (Chapter 2) enabled the data collected by the tags to be stored to the tags' 8 megabyte (MB) internal flash memory. The data can subsequently be downloaded and used to create dynamic contact networks.

Contact networks explicitly quantify the contact events between individuals in a population (6,7). Contact networks consist of nodes which represent individual entities such as people or animals, and edges which represent the contact between the nodes (8). The networks are often generated in one of two ways: 1) empirical data can be used to simulate larger contact networks (9), and 2) the empirical data can be used to replicate the networks that occurred at the time of data collection (10). In the first method, descriptive data is collected from a sample of the population of interest. The collected data can then be used to extrapolate a network composed of the same characteristics as the sampled data (11). In the second method, the collected data is deemed sufficient to create a contact network representative of the study population. Therefore, the network only consists of observed data, and no extrapolation of the

contacts is necessary. This paper focuses on the second type of contact network, one in which sufficient observational data is available to form a representative contact network.

Contact networks can be analyzed using descriptive methods, such as network analysis (8), and statistical methods, for example exponential random graph models (12).

Descriptive methods include metrics that describe the network's properties such as the number of nodes and edges in a network, and centrality measures that describe a node's importance in the network. Most centrality measures can be normalized to enable the comparison of different networks (8). Depending on the type of data collected (e.g. temporal or cross sectional), different statistical tool sets can be used to analyze the networks. Exponential random graph models (ERGMs) describe the significance and relationship between the dependent and independent variables in the ERGM (13). In ERGMs, the dependent variable is a network and the independent variables represent relationships between the nodes (12,13). These relationships can range from structural descriptors, such as triangulation in the network, to node or edge specific descriptors, such as an individual's sex. Similarly to ordinary logistic regression, the regression coefficients represent the log-odds that an edge will form based on the model (14).

Given the gap in knowledge with regards to the structure and characterization of the Canadian equine population, the objective of this research was to use empirical contact data to characterize and compare the contact pattern structure within equestrian facilities in Ontario. The network analysis included the assessment of: a) the degree to which the networks could be generalized, b) the ability to predict the networks from

observational data using the tag collected data as a gold standard, c) selected centrality measures to assess for network-level correlations, as well as patterns between the networks, d) whether the networks were representative of the assumption of homogeneous mixing, and, e) a statistical analysis to determine if the horses' relative stall location was important with respect to the formation of edges in the network.

4.3 Methods

This study was approved by the University of Guelph's Research Ethics Board (REB certificate #16AP009) and the Animal Care Committee (AUP # 3518).

4.3.1 Study design and data collection process

Four equestrian facilities located in southwestern Ontario, Canada, participated in the study. Consent was obtained from the participants according to the REB. The facilities ranged in size from 20-28 horses (Table 4.1). Each facility consisted of one barn, multiple pastures, and indoor and outdoor riding arenas. Horses were either housed in the barn or in a pasture.

The study was conducted at each facility for a 7-day period between May and July 2017. Active RFID tags (Bitmanufactory, Cambridge, United Kingdom) were used to record the contact events that occurred between horses wearing the tags. A detailed description regarding the operation of the tags has been documented Chapter 2. Briefly, the tags contain 8 MB of onboard flash memory. When two tags were within 2 meters (m) of each other, the tags recorded the tag identification number assigned to the

participants of the contact event, as well as the start and end times for each contact event. The recorded data was stored to the internal flash memory.

Each RFID tag was placed in a plastic cover (Avery Products Corporation, Whitby, Ontario, Canada) which was sealed closed with duct tape (Logix, Port Hope and Cobourg, Ontario, Canada). The tags were attached to the nosepiece of the horses' halters with Vetwrap (3M™, United States). Employees and trainers were instructed to keep the halters with the logging device with the assigned horse at all times for the duration of the study period. Horses tend to be social and curious animals by nature (15). Therefore, the halters were placed on a halter hook located on the pasture fence when a horse was in its pasture in order to prevent the horses from chewing on the tags when unsupervised. The standard operating procedure at one of the facilities required that the horses wear the halters at all times. Consequently, at this facility, duct tape was used to attach the tag to a livestock tag (TSC, Ontario, Canada) which was subsequently attached to the halter with a carabiner. This enabled easy removal of the tag from the halter.

A survey completed by the facility manager was used to collect data regarding the facility's day-to-day schedule, as well as the horses' demographic information. For each participating horse, data were collected regarding the horse's stall number and location, the primary pasture in which the horse was turned out during the study week, as well as the horse's breed, sex, age, and primary use, e.g. competition, pleasure, school horse, etc.

At the end of the study period, the tags were removed from the horses' halters and were transported to the university where the data were downloaded to a computer and stored in a MySQL (Oracle Corporation, Redwood Shores, California) database. The data were subsequently cleaned and aggregated by 24-hour periods, excluding the time elapsed during the deployment setup and the final tag collection on the first and last days of the study. The data were output as edge-lists in 7 comma separated value (CSV) files, providing 1 file for each day of the study. The CSV files were imported into R version 3.3.0 (16) for analysis.

4.3.2 Network description

Network analyses using the igraph package, version 1.0.1 (17) were performed for each study day for each facility. The analyses included both visual and analytical aspects (described below). The results of the analyses were used to compare the network characteristics for each facility.

The Jaccard similarity index (JSI) was used to assess the network generalizability and reproducibility. The JSI was calculated by comparing which contacts were repeated in each of the networks containing the same node set ($\frac{|A \cap B|}{|A \cup B|}$ (18)). Since the JSI compared the presence or absence of an edge between two nodes in different networks, it did not take the edge weight into consideration. The network generalizability was assessed by computing the JSI for each day of the study, per facility, and was visualized using a boxplot.

The objective of assessing the predictability of the networks was met by using the JSI to compare a survey-based network to the networks formed with the tag collected data, which was used as the gold standard. The survey-based network for each facility was formed by assuming that all horses that shared a barn or pasture (as specified in the study survey) came in contact with each other. In the case that a horse was reported to be housed in more than one pasture during the study period, the first pasture listed in the survey was chosen to represent the horse's primary pasture for the survey-based network.

At the facility level, both the unweighted and weighted degree (hereafter referred to as "degree" and "strength" respectively) were computed for each horse in each network. The degree refers to the number of distinct horses that a horse of interest contacted in each network and is calculated by summing the number of edges incident to the respective horse's node. The normalized degree, used for comparing networks, was obtained by dividing the degree by $(n - 1)$, where n represents the number of nodes (horses) in the network. A node's strength represented the total duration (hours per day) of contact a node had with all the other nodes in the network. A function was developed to normalize the strength of a node by dividing a node's strength by the maximum node strength in the network. Due to the non-normal nature of the data, a comparison of the horse-level degree and the horse-level strength for each study day was conducted using the spearman correlation coefficient for each horse on each day of the study.

The normalized betweenness (the number of times a node was on the shortest path between two other nodes) was calculated for each node. Betweenness was calculated

as the sum of the number of shortest paths through a node of interest divided by the total number of shortest paths (8). This number was normalized by multiplying the betweenness by $\frac{2}{(n-1)(n-2)}$. Since betweenness cannot be calculated between nodes found in different weak components, defined as a subgroup of nodes in which all nodes are indirectly linked (19), the metric was calculated for the nodes within each component.

The assumption of homogenous mixing was studied at two levels of aggregation: 1) the data were analyzed by study day, and 2) the data were aggregated to form a single, static, week-long network. Heat maps were formed for each network and were interpreted to assess whether the networks conformed to the assumption of homogeneous mixing. The heat maps were also used to compare the relative duration of contact between each pair of horses on each study day.

4.3.3 Statistical analysis

The R Statnet suite of packages version 2016.9 (9) was used to conduct a regression analysis on the networks using the ERGM (20) package. The model, which takes the form of $y \sim \langle term1 \rangle + \langle term2 \rangle + \langle term3 \rangle + \dots$ (12), included the *edge* term which is analogous to the intercept in an ordinary logistic model, and the *edg cov* term, which is used to study relational predictors such as distance. The *edg cov* term was used to describe the effect of the distance between horses' stall locations (hereafter referred to as "distance") on the probability of edge formation. The stall locations were only used if the horses were housed in the stalls overnight. For horses that were housed

in the pasture (and hence were outside all day), the pasture location was used as a proxy for stall location. The distance was calculated as the Euclidean distance between two horses' stall or pasture locations based on a map of the facility.

4.4 Results

4.4.1 Descriptive results

Horse enrolment at each of the 4 participating facilities was 100% (Table 4.1). The horses ranged in age from 1-year to 37-years old. Three stallions participated in the study, with the majority of horses being geldings (n=62) and mares (n=33). Horses tended to be used for competition purposes, pleasure riding, and as school horses (Table 4.1).

For each network, both the number of edges and the density varied over the study period, resulting in densities which were seldom greater than 0.5 (Table 4.2). The network order (the number of nodes in the network) of the study facilities ranged from 20.0-28.0. Different numbers of components were seen throughout each facility's networks. While each facility had multiple components, some daily networks were only composed of a single component (Table 4.2).

4.4.2 Assessment of the assumption of homogeneous mixing

Due to the similarity in the heat maps from each study day, only the heat map from day 1 of the study is included here (Figure 4.1). The remaining heat maps can be found in the supplementary material (Figures S4.1-S4.4). Each cell in the heat map represents the duration of the contact that occurred between individual pairs of horses in the

network. Although contact was observed between horses from different pastures, the majority of the observed contact occurred between horses that shared a pasture. This contact is found along the diagonal of the heat map. Similar contact patterns were observed on each study day for each facility, with the main difference being the contact duration between each pair of horses. These patterns were seen for both the day-long networks, as well as the week-long, aggregated network (Figure 4.2).

4.4.3 Assessment of network generalizability

Boxplots of the JSI, used to assess network generalizability, indicated that there was a large amount of overlap between the facilities (Figure 4.3). The minimum amount of generalizability of contacts from day to day occurred on Facilities 2 and 3 and was characterized by ~11% of contact similarity during the study period. The most similar days in terms of contacts occurred on Facility 4 with ~62% of identical contacts during the study week. The average JSI was 43.8% (SD=0.07), 27.1% (SD=0.13), 31.2% (SD=0.14), and 51.5% (SD=0.07) for facilities 1, 2, 3, and 4 respectively.

4.4.4 Assessment of network predictability

The Jaccard similarity indices indicated that the survey-based networks were able to correctly predict ~8.7% - 79.6% of contacts (Figure 4.4). The survey-based networks predicted an average of 33.8% (SD=0.05), 33.1% (SD=0.20), 33.0% (SD=0.13), and 58.6% (SD=0.09) of contacts on Facilities 1, 2, 3, and 4 respectively. There was minimal variation in the data within each facility's networks.

4.4.5 Computation and comparison of centrality measures

The minimum normalized degree was 0.00 for each facility while the maximum normalized degree varied between facilities and was generally less than 0.85 (Figure 4.5). Often, unless a horse's degree was 0.00, the degree trends were similar across days. In other words, horses who had higher degrees relative to other horses, tended to have higher relative degrees on most study days. The node strength varied between horses and study days (Figure 4.6). Similar patterns to those found with the relative degree were found with respect to the relative strength. A positive trend was observed in the plots used to test whether a correlation existed between degree and strength. While the correlation coefficient was always significant ($p < 0.05$), it presented a weak to moderate correlation between the degree and strength for each horse ($0.41 < r < 0.54$) (Figure 4.7).

Lastly, the normalized betweenness varied for each horse throughout the study week (Figure 4.8). The betweenness generally ranged between 0.0 and 0.8, with one node having a betweenness of 1.0. The horse with the highest betweenness was different on each day of the study. No trends were observed for the betweenness between facilities.

4.4.6 Regression analysis

The ERGM models indicated that for 2 of the 4 facilities (Facilities 3 and 4), the distance parameter was always statistically significant (the p-values ranged from < 0.0001 to ~ 0.005) with a negative coefficient (Tables 4.3-4.6). No patterns were evident for Facilities 1 or 2.

4.5 Discussion

The analysis of contact networks has applications ranging from the description of the flow of information through a network to the description of possible disease transmission routes (21). Furthermore, the networks are often used in conjunction with other analytical methods such as disease transmission models. However, the paucity of contact data within the Ontario equine community poses a challenge for the characterization and use of the contact network structure. The current study aims to fill this gap in knowledge by collecting data and comparing the contact networks from 4 equestrian facilities.

Due to the difficulty in collecting contact pattern data, studies often use the assumption of homogeneous mixing to characterize the contact rate in disease transmission models (22–24). The results of the current study however, imply that this assumption is insufficient to describe the types of contacts that occurred within the participating equine facilities. While the heat maps describing the contact durations for each facility were slightly different in shape, all of the heat maps led to the same conclusion: the contacts between the horses did not conform to the homogeneous mixing assumption.

Furthermore, the majority of the contacts occurred between horses that shared a pasture. This was true when looking at both the individual, day-long networks and the aggregated, week-long networks. These findings suggest that heterogeneous mixing patterns should be used when developing models for disease transmission within an equestrian facility in order to more accurately capture the risk of disease spread within the facility.

Some network-based analyses may require larger amounts of contact data than are available. This could occur, for example, when using the contact data to parameterize a disease transmission model for a disease whose natural history spans a longer time period than the data collection period. One method of addressing this issue is to repeat the empirical contact patterns for the required length of time (25). This data replication process assumes that the contact patterns are similar during the given period. The current study substantiated this assumption by measuring the similarity of the contact events over the study period. The JSI, used to assess the similarity in contacts between study days suggests that between 11% and 62% of individual contacts were repeated over the study period. Therefore, in situations where minimal data is available, it is possible to generalize and extrapolate the contact events from the collected data. In addition to the ability to generalize the contacts, the current study showed that knowledge of the facility logistics can help predict the contacts that may occur. Both the predictability and generalizability of networks increase the possibility of planning intervention strategies in advance of an outbreak in order to minimize the epidemic impact.

Centrality measures are often used to infer the extent to which a disease can spread in a network and have been used in the past to infer such data in animal populations (26–28). The overall trends in the relative degree and strength provide a platform on which it is possible to base disease control and prevention inferences. Depending on the disease, contact duration and/or frequency can contribute to the transmission probability (7). Therefore, it is important to study both the network degree (for cases when contact

duration does not contribute to disease transmission probability) and strength (for cases where contact duration does contribute to disease transmission probability).

The degree centrality is directly related to the number of direct contacts that each node had in the network (8). Given the trends in the degree centrality, where the horses who had higher degrees relative to other horses tended to have higher degrees during the entire study week, it is possible to know in advance which horses are likely to be more instrumental in disease transmission based on minimal amounts of data. The normalized degree for each facility ranged from 0.0 to ~0.8 indicating the range of possible outcomes from the introduction of a pathogen into a population. In the best-case scenario, a horse with degree 0.0 can become infectious, in which case the chances of the disease spreading to other horses by direct contact is minimal.

Alternatively, if a horse with larger degree becomes infectious, then it is more likely that the pathogen will spread throughout the facility.

A high betweenness indicates that the respective node may represent a cut-point (29). There was no one node with the highest betweenness centrality on all days. The large range in betweenness makes predictive inferences due to node betweenness difficult in these populations.

An ERGM model was used to understand the factors that influenced the probability of the formation of edges in the networks. Although the distance predictor did not show meaningful results across the facilities, the variable was always significant for both facilities 3 and 4. Both of these facilities had minimal “noise” around the diagonal in the

heat maps (Figures 4.1 and 4.2) compared to the other two facilities, suggesting that the majority of contact occurred between horses that shared a pasture. This is also supported by the fact that most horses in Facility 4 were outdoor horses, and therefore together for both day and evening hours. While the small network sizes precluded the inclusion of a pasture variable in the model, it should be considered as an important source of contact.

4.5.1 Limitations, benefits, and opportunities for future work

Although horses in Ontario can, and do leave the facility year-round, the frequency with which they leave changes based on the time of year and their usage. Therefore, since the study was conducted during the summer months, the contact patterns collected during the study and the resulting network characteristics may differ from the contact patterns during the winter when horses seldom leave the facility and may stay indoors due to low temperatures.

The fact that horses leave the facility provides an opportunity for the introduction of external pathogens. This is confirmed by a study by Spence et al., (2017) which examines the impact of disease spread between facilities when horses congregate at a competition. This external influence should be taken into consideration when considering the routes and characteristics of disease spread.

Multiple sources of misclassification bias may have influenced the results of the study. First, if a horse was turned out in more than one pasture over the study period, then the survey-based network in which a horse was assigned to just one pasture may over- or

under-estimate the contacts predicted. Second, if a halter was not kept with the horse at all times, for example, if the horse owner forgot to keep the halter with the horse, the final count and duration of contacts that occurred will be incorrect.

Lastly, the small size of each facility limited the ability to use statistical techniques to study the factors affecting the probability of edge formation. Animal-based disease transmission models, and the collection of contact pattern data with RFID tags are growing in popularity necessitating the development of analytical tools that can address a multitude of data challenges inherent to these types of datasets.

Regardless of these limitations, the networks created provide insight into the types of contacts that occur within equestrian facilities in Ontario. Furthermore, the networks provide an improved ability to model more realistic outcomes with various types of computer simulations.

4.6 References

1. Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005;2(4):295–307.
2. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings DAT. Close encounters of the infectious kind: Methods to measure social mixing behaviour. *Epidemiol Infect*. 2012;140(12):2117–30.
3. Chen S, White BJ, Sanderson MW, Amrine DE, Ilany A, Lanzas C. Highly dynamic animal contact network and implications on disease transmission. *Sci Rep*. 2014;4:4472–9.
4. Isella L, Stehlé J, Barrat A, Cattuto C, Pinton JF, Van den Broeck W. What's in a crowd? Analysis of face-to-face behavioral networks. *J Theor Biol*. 2011;271(1):166–80.
5. Bioglio L, Génois M, Vestergaard CL, Poletto C, Barrat A, Colizza V. Recalibrating disease parameters for increasing realism in modeling epidemics in closed settings. *BMC Infect Dis*. 2016;16(1):676–91.
6. Perkins SE, Cagnacci F, Stradiotto A, Arnoldi D, Hudson PJ. Comparison of social networks derived from ecological data: Implications for inferring infectious disease dynamics. *J Anim Ecol*. 2009;78(5):1015–22.
7. Krause J, Croft DP, James R. Social network theory in the behavioural sciences: Potential applications. *Behav Ecol Sociobiol*. 2007;62(1):15–27.
8. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis*. 2009;56(4):109–20.
9. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: Software tools for the Statistical Modeling of Network Data. 2003.
10. Craft ME, Caillaud D. Network models: An underutilized tool in wildlife epidemiology? *Interdiscip Perspect Infect Dis*. 2011;2011:1–12.
11. Welch D, Bansal S, Hunter DR. Statistical Inference to advance network models in epidemiology. *Epidemics*. 2011;3(1):38–45.
12. Morris M, Handcock MS, Hunter DR. Specification of exponential-family random graph models: Terms and computational aspects. *J Stat Softw*. 2008;24(4):1548–7660.
13. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *J Stat Softw*. 2008;24(3):nihpa54860.
14. Goodreau SM, Handcock MS, Hunter DR, Butts CT, Morris M. A statnet tutorial. *J*

- Stat Softw. 2008;24(9):1–27.
15. Lloyd AS, Martin JE, Bornett-Gauci HLI, Wilkinson RG. Horse personality: Variation between breeds. *Appl Anim Behav Sci.* 2008;112(3–4):369–83.
 16. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.r-project.org/>
 17. Gabor C, Tamas N. The igraph software package for complex network research. *InterJournal* [Internet]. 2006;Complex Sy:1695. Available from: <http://igraph.org>
 18. Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. *J Am Soc Inf Sci Technol.* 2007;58(7):1019–31.
 19. Dubé C, Ribble C, Kelton D, McNab B. A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound Emerg Dis.* 2009;56(3):73–85.
 20. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Krivitsky PN, Morris M. *ergm*: Fit, simulate and diagnose exponential-family models for networks. The Statnet Project (<http://www.statnet.org>). R package version 3.8.0. 2017.
 21. Holme P. Network reachability of real-world contact sequences. *Phys Rev E.* 2005;71(4):46119.
 22. Daly JM, Newton JR, Wood JLN, Park AW. What can mathematical models bring to the control of equine influenza? *Equine Vet J.* 2013;45(6):784–8.
 23. Park AW, Wood JLN, Daly JM, Newton JR, Glass K, Henley W, et al. The effects of strain heterology on the epidemiology of equine influenza in a vaccinated population. *Proc R Soc London B Biol Sci.* 2004;271(1548):1547–55.
 24. Hamede RK, Bashford J, McCallum H, Jones M. Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: Using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecol Lett.* 2009;12(11):1147–57.
 25. Gemmetto V, Barrat A, Cattuto C. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect Dis.* 2014;14(1):695–705.
 26. Kiss IZ, Green DM, Kao RR. The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. *J R Soc Interface.* 2006;3(10):669–77.
 27. Firestone SM, Christley RM, Ward MP, Dhand NK. Adding the spatial dimension to the social network analysis of an epidemic: Investigation of the 2007 outbreak of equine influenza in Australia. *Prev Vet Med.* 2012;106(2):123–35.
 28. White LA, Forester JD, Craft ME. Using contact networks to explore mechanisms

of parasite transmission in wildlife. *Biol Rev.* 2017;92(1):389–409.

29. Makagon MM, McCowan B, Mench JA. How can social network analysis contribute to social behavior research in applied ethology? *Appl Anim Behav Sci.* 2012;138(3–4):152–61.

4.7 Tables

Table 4.1: Descriptive statistics for each of the 4 participating equestrian facilities. The facilities were located in southwestern Ontario, Canada, and boarded 20-28 horses that ranged in age, sex, breed, and use. Each facility had a 100% enrolment rate.

			Facility 1	Facility 2	Facility 3	Facility 4
		Number of participating horses /total horses boarded within the facility	20/20	28/28	26/27 *	24/24
		Number of barns	1	1	1	1
		Number of pastures **	7	8	6	5
		Study start date	May 2017	May-June 2017	June 2017	July 2017
Number of horses by:	Sex	Gelding	8	18	19	17
		Mare	9	10	7	7
		Stallion	3	0	0	0
	Breed	Thoroughbred	1	9	4	4
		Quarter horse	1	3	1	8
		Hanoverian	0	0	5	0
		Pony	1	4	2	2
		Crossbreed	0	11	7	5
		Warmblood	15	0	3	4
		Other	2	1	4	0
		Unknown	0	0	0	1
	Use	School/ program horses	0	20	9	6
		Competition	11	6	10	6
		Pleasure riding	1	2	1	3
		Breeding	1	0	0	0
		In training	4	0	1	1
		Other (i.e. retired)	3	0	5	8
	Age range (years)	<1 year	0	0	0	0
		1-4	5	0	0	1
		5-9	7	5	10	3
10-19		7	17	12	7	
20-37		1	6	5	13	

*The 27th horse was enrolled in the study but was off the facility for the duration of the study. Therefore, it is not included in the node count.

** This number only includes pastures that housed horses that participated in the study.

Table 4.2: Description of the networks for each day of the study. Contact data was collected over a 7-day period at 4 equestrian facilities in southwestern Ontario. The data was aggregated and used in the development of 7 networks describing the contacts that occurred on each study day. All of the facilities had multiple components, however, the number of components varied by day. Each network had between 20-28 nodes. The network densities were generally less than 0.5. Densities greater than 0.5 are bolded.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Facility 1							
Nodes	20						
Edges	73	76	60	72	65	58	46
Density	0.384	0.4	0.316	0.379	0.342	0.305	0.242
Number components	2	2	1	2	3	3	5
Number of components of size ≤ 3	1	1	0	1	2	2	4
Facility 2							
Nodes	28						
Edges	226	236	187	82	64	33	50
Density	0.598	0.624	0.495	0.217	0.169	0.087	0.132
Number of components	1	2	4	5	9	11	16
Number of components of size ≤ 3	0	1	3	4	8	10	15
Facility 3							
Nodes	26						
Edges	102	124	117	86	50	39	28
Density	0.314	0.382	0.36	0.265	0.154	0.12	0.086
Number of components	1	1	2	2	6	7	11
Number of components of size ≤ 3	0	0	0	1	5	6	10
Facility 4							
Nodes	24						
Edges	98	130	120	58	83	87	61
Density	0.355	0.471	0.435	0.210	0.301	0.315	0.221
Number and size of components	1	1	2	3	3	4	4
Number of components of size ≤ 3	0	1	1	1	2	3	3

Table 4.3: ERGM^o analysis for each facility. The analysis was carried out using the R ERGM (20) package in the Statnet suite version 2016.9 (9) and studied the effect of the distance between the horses' stalls or pasture location on the resulting network. The distance variable was not always statistically significant. Statistically significant values (p<0.05) are indicated with bold text.

	Day 1		Day 2		Day 3		Day 4		Day 5		Day 6		Day 7	
	Edge	Distance												
Facility 1														
Coeff.*	0.12	-0.06	0.30	-0.07	0.31	-0.11	-0.50	0.00	0.24	-0.09	-0.23	-0.06	-1.06	-0.01
Error[‡]	0.34	0.03	0.34	0.03	0.36	0.04	0.35	0.03	0.35	0.03	0.36	0.03	0.39	0.04
p^Ω	0.74	0.06	0.38	0.03	0.39	0.00	0.15	1.00	0.49	0.01	0.52	0.07	0.01	0.82
Facility 2														
Coeff.*	1.18	-0.05	0.56	-0.00	-0.36	0.02	-0.95	-0.02	-1.29	-0.02	-2.69	0.02	-1.99	0.01
Error[‡]	0.24	0.01	0.23	0.01	0.22	0.01	0.26	0.02	0.28	0.02	0.41	0.02	0.33	0.02
p^Ω	<0.00	0.00	0.01	0.79	0.10	0.08	0.00	0.16	<0.00	0.24	<0.00	0.35	<0.00	0.70
Facility 3														
Coeff.*	0.76	-0.14	0.76	-0.11	0.39	-0.08	-0.21	-0.07	-0.55	-0.11	-0.87	-0.10	-1.30	-0.10
Error[‡]	0.27	0.02	0.25	0.02	0.24	0.02	0.26	0.02	0.31	0.03	0.34	0.03	0.38	0.04
p^Ω	0.00	<0.00	0.00	<0.00	0.11	<0.00	0.40	0.00	0.07	0.00	0.01	0.00	0.00	0.01
Facility 4														
Coeff.*	3.77	-0.24	2.45	-0.12	1.34	-0.08	1.31	-0.17	1.27	-0.11	1.45	-0.12	1.44	-0.17
Error[‡]	0.51	0.03	0.36	0.02	0.29	0.01	0.33	0.02	0.30	0.02	0.31	0.02	0.33	0.02
p^Ω	<0.00													

^oERGM: exponential random graph model
^{*} The log odds coefficient of the model

[‡]the standard error of the model
^Ω the model p-value

4.8 Figures

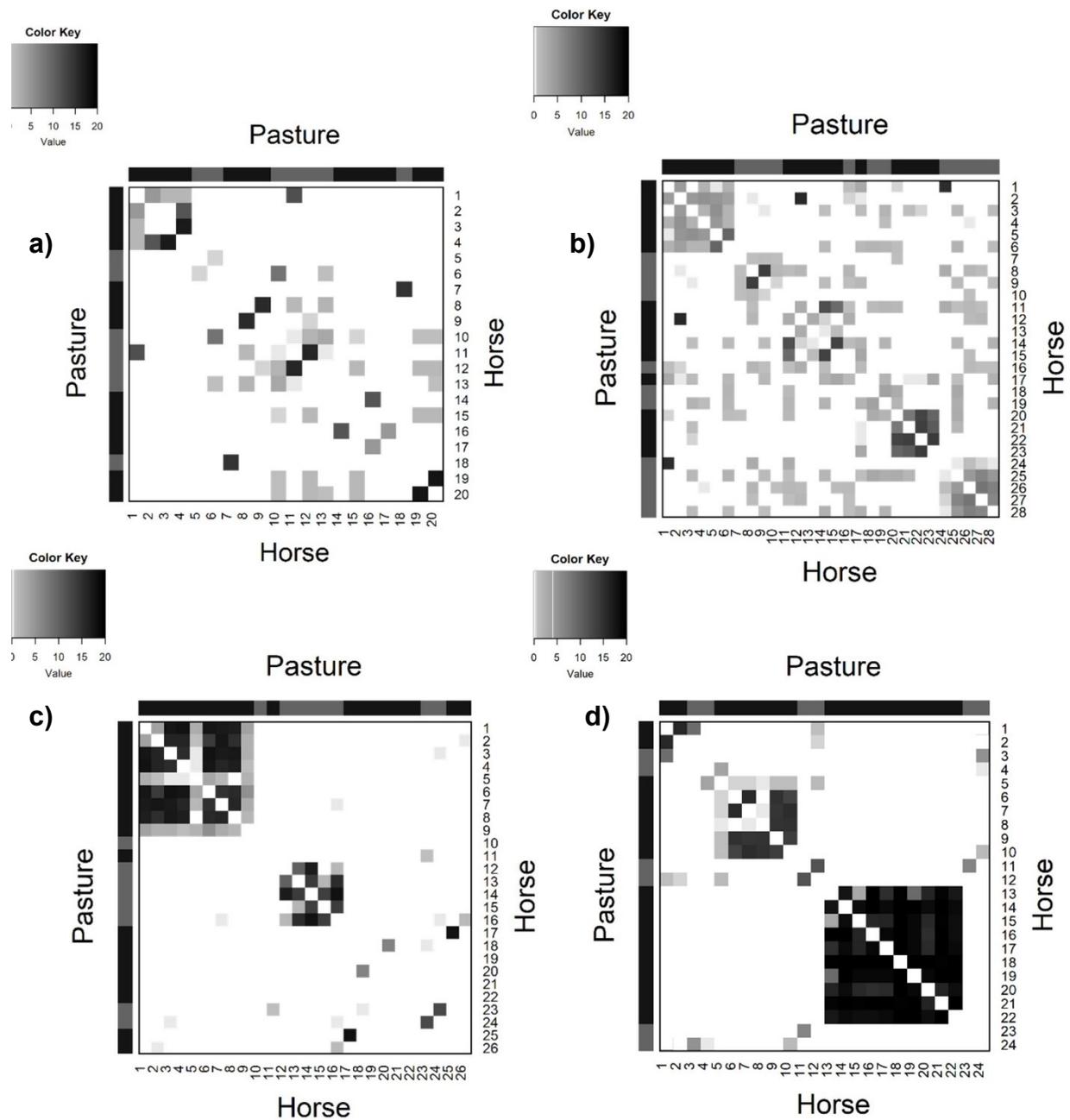


Figure 4.1: Heat maps of the contact that occurred on each facility on day 1 of the study. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. Panels a) and b) show the heat maps for facilities 1 and 2 respectively while panels c) and d) represent facilities 3 and 4 respectively. A strong contact duration (hours per day) exists on the diagonal of the heat map between horses who shared a pasture. There is no evidence of homogeneous mixing at this level of aggregation.

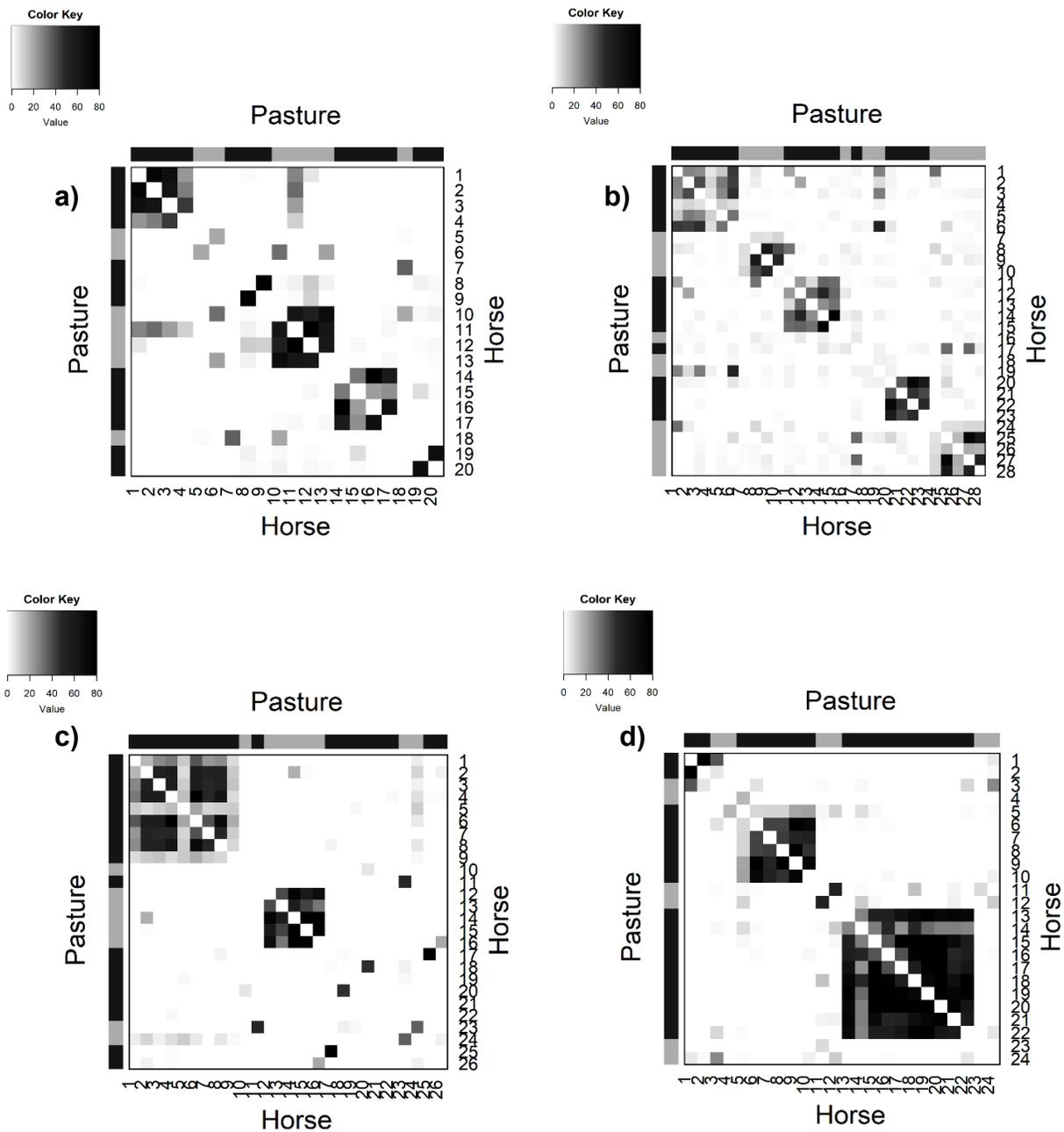


Figure 4.2: Aggregated, week-long heat maps for each facility. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. Panels a) and b) contain the heat maps for facilities 1 and 2 respectively while panels c) and d) represent facilities 3 and 4 respectively. Heat maps are visually similar to the daily heat maps in Figure 4.1. There is a strong trend on the diagonal between horses who shared a pasture. There is no evidence of homogeneous mixing at this level of aggregation.

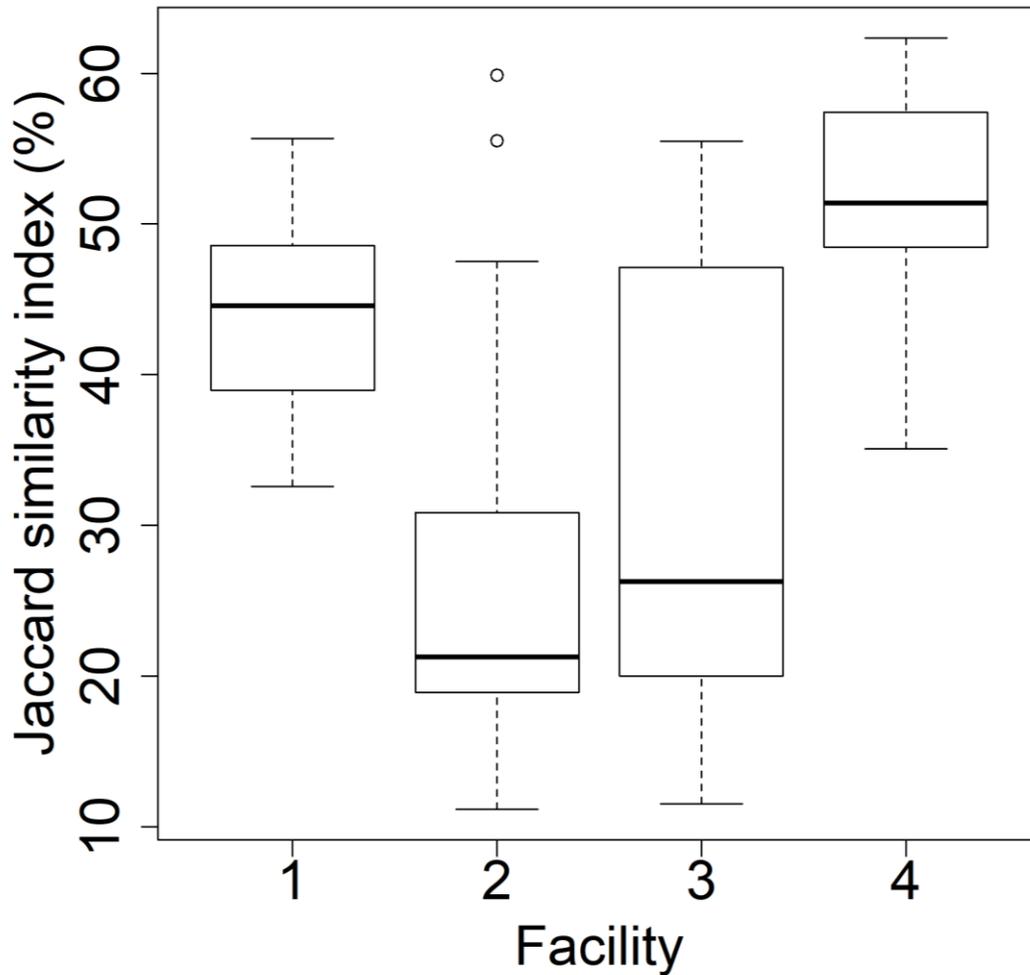


Figure 4.3: Boxplot for Jaccard similarity index to quantify the similarity in contacts between participating horses on each day of the study for each facility. The similarity in contacts can be used to generalize the contacts that occurred over the course of the study period. Roughly 11-62% of contacts were repeated over all the networks.

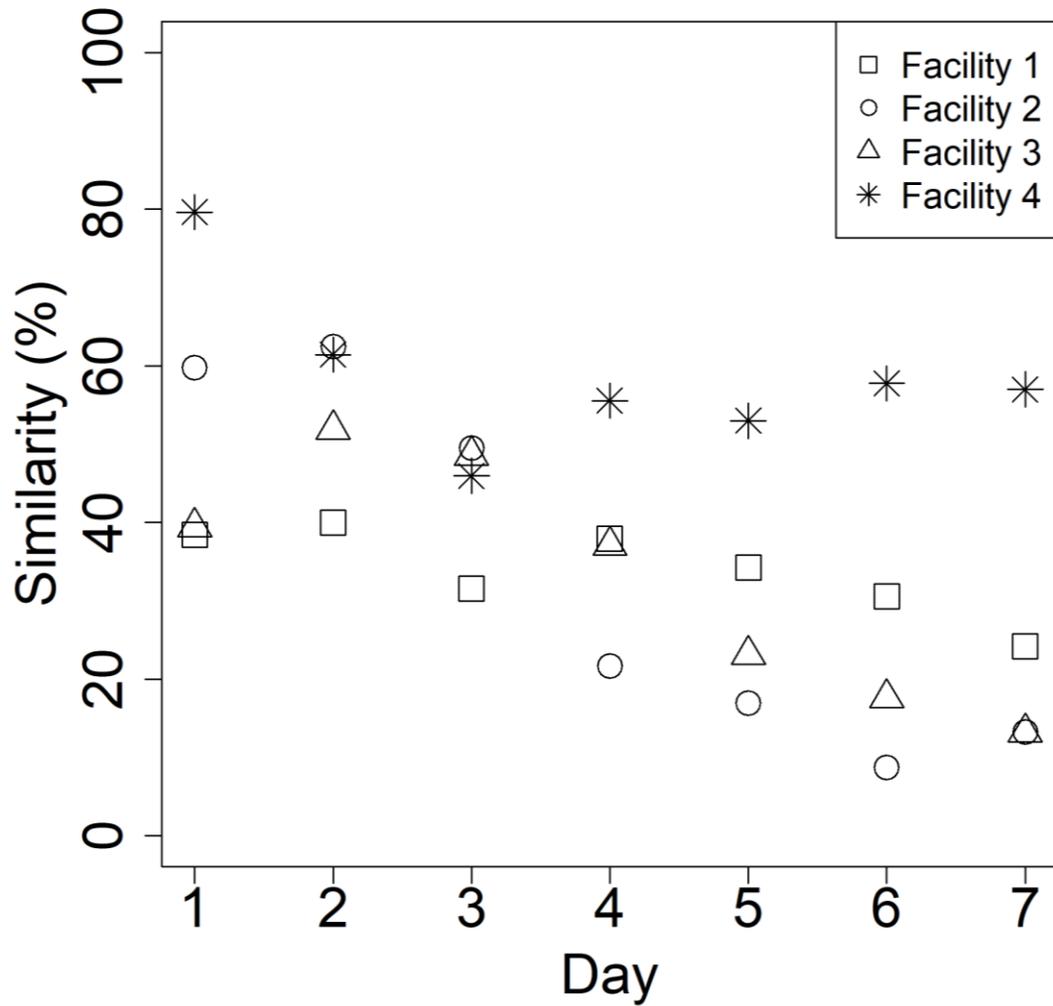


Figure 4.4: Jaccard similarity indices for the comparison of the survey-based networks and the networks formed with the logger acquired data. Survey-based networks were formed by assuming that all horses that shared a pasture or barn came in contact. The survey-based networks were able to correctly predict 8.7%-78.7% of contacts that occurred on the four equestrian facilities.

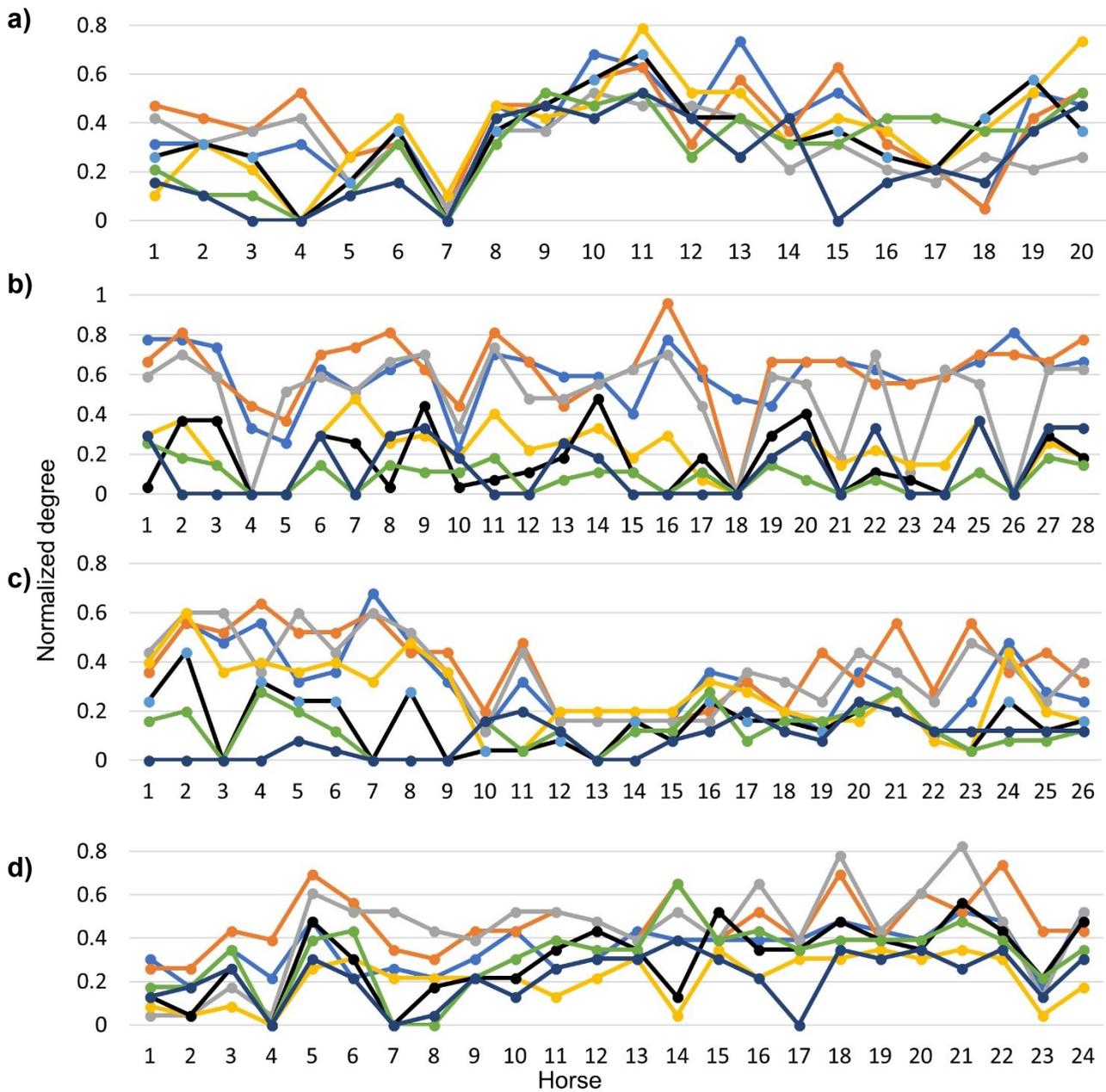


Figure 4.5: Normalized degree centrality for each horse during the study week. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. Different coloured lines represent the different study days (blue=day 1, orange=day 2, grey=day 3, yellow=day 4, black=day 5, green=day 6, and navy=day 7). Overall, the same trends occurred on each of the study days. In other words, with the exception of isolates, horses that had a larger degree compared to other horses had a larger degree on each day of the study compared to the other horses.

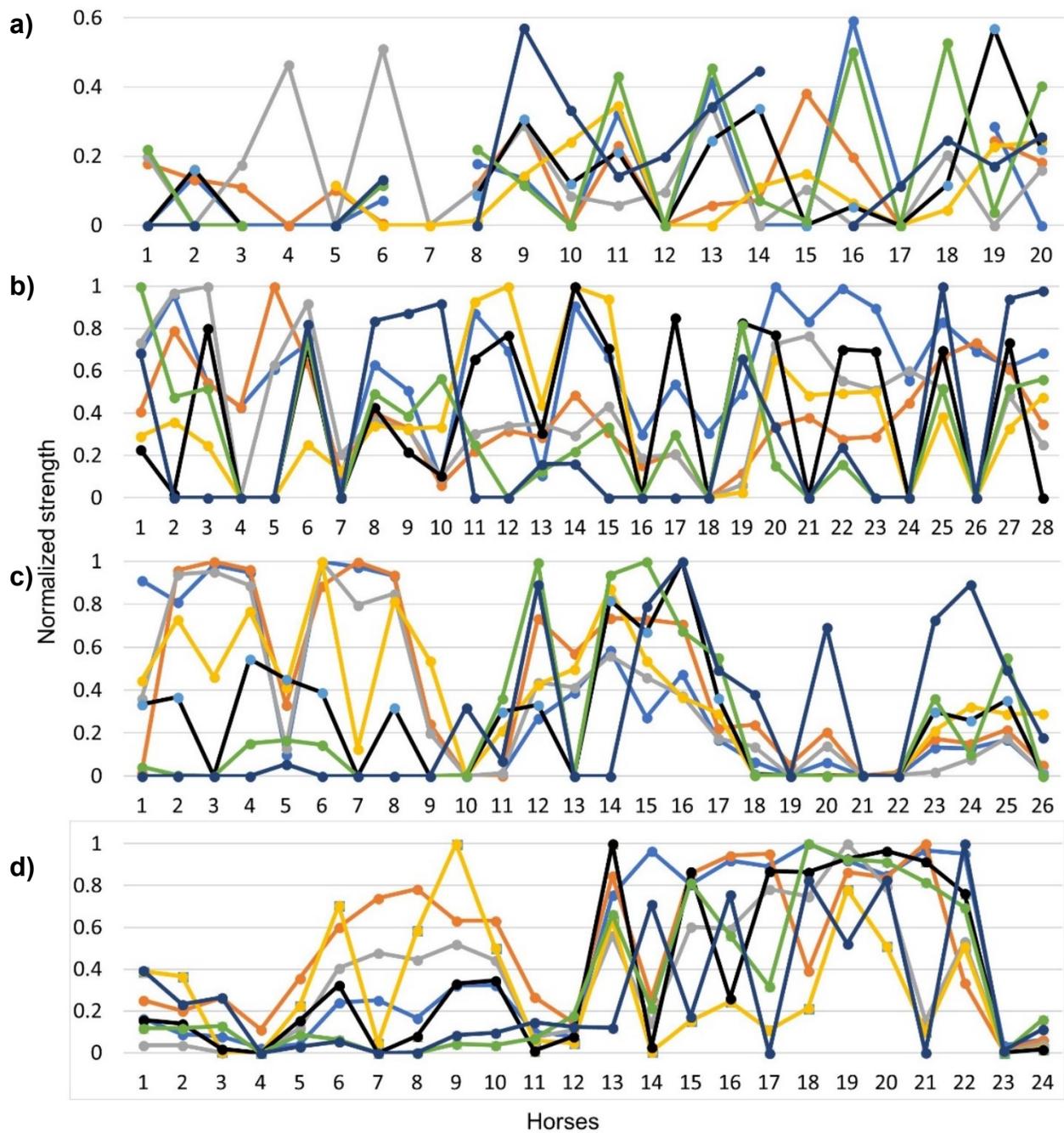


Figure 4.6: Normalized strength for each horse during the study week. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. Different coloured lines represent the different study days (blue=day 1, orange=day 2, grey=day 3, yellow=day 4, black=day 5, green=day 6, and navy=day 7). Similar trends occurred as those in Figure 4.5.

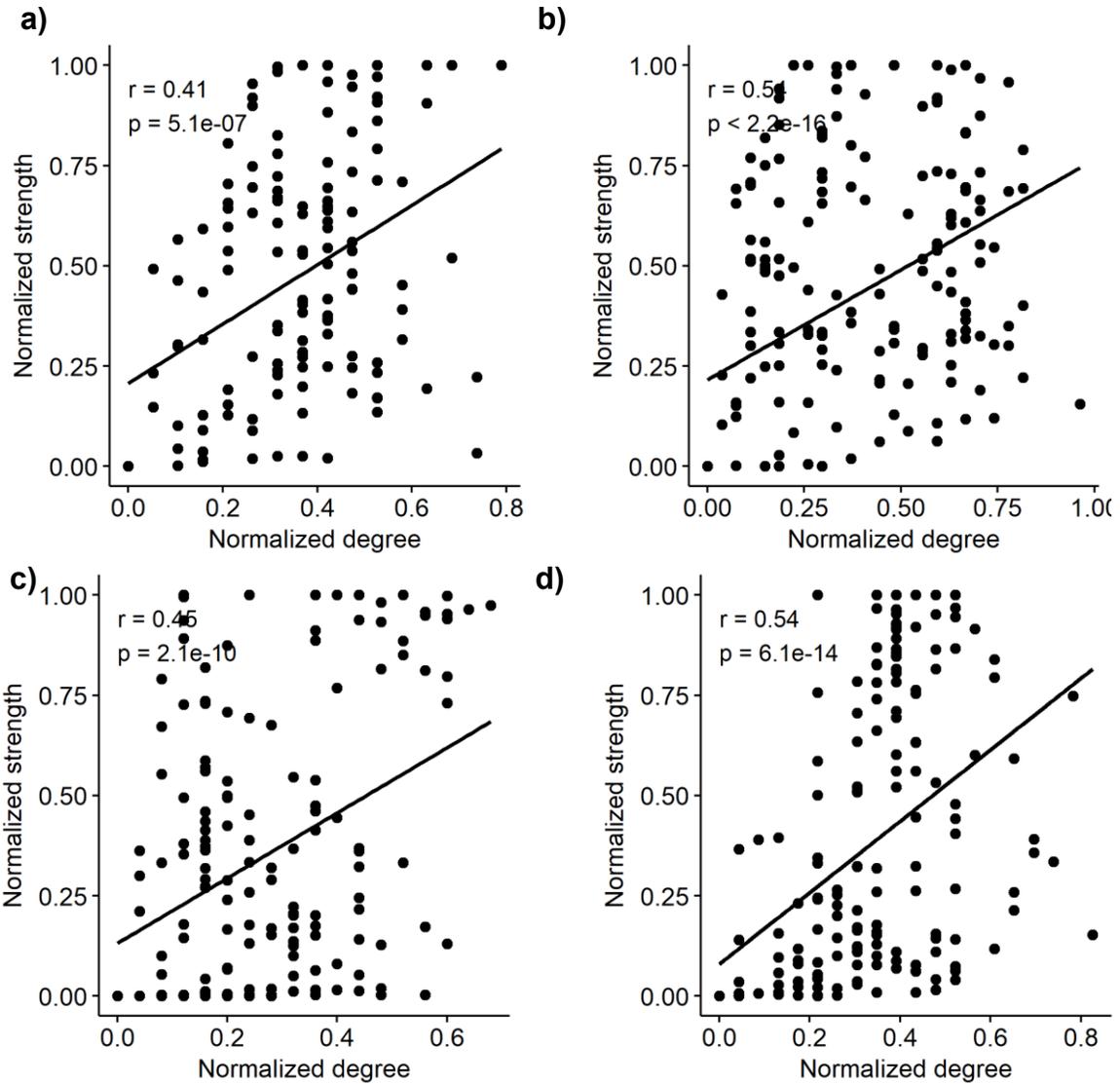


Figure 4.7: Comparison of the normalized degree and normalized strength over the study period. The line on each graph represents the regression line. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. A statistically significant, weak to moderate positive correlation, computed with the spearman correlation coefficient (r) was observed for each facility.

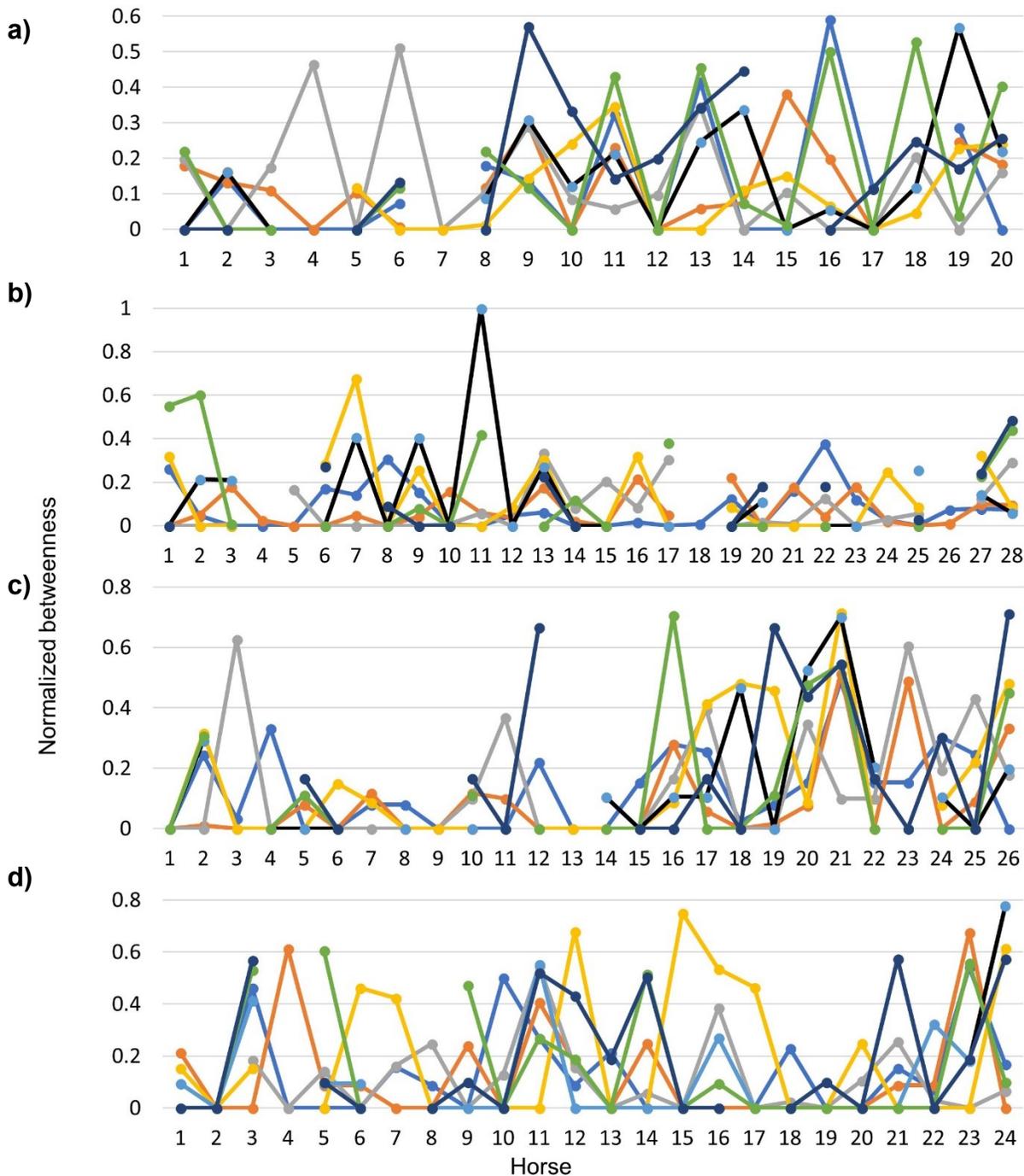


Figure 4.8: Normalized betweenness for the horses in each component. Panels a) and b) represent facilities 1 and 2 respectively, and panels c) and d) represent facilities 3 and 4 respectively. Gaps in the graphs represent isolates (for which it is impossible to calculate the betweenness centrality). No trends were observed for the betweenness between the facilities.

CHAPTER 5

Comparing the effects of non-homogenous mixing patterns on epidemiological outcomes in equine populations: A mathematical modelling study

5.1 Abstract

Disease transmission models often assume homogenous mixing. This assumption, however, has the potential to misrepresent the disease dynamics for populations in which contact patterns are non-random. A disease transmission model with an SEIR structure was used to compare the effect of empirical equine contact networks to theoretical networks using equine influenza as a case study, as well as the effect weighted and unweighted networks on the disease dynamics. Incidence curves generated with the unweighted empirical networks were similar in epidemic duration (17-18 days) and peak incidence (30.8-46.4%). An unweighted complete network resulted in a larger peak incidence (95.8%) and a similar epidemic duration (17 days) to the unweighted empirical networks. An unweighted network in which each node had an average degree of 4 resulted in the smallest peak incidence (12.5%), and the longest duration (19 days). In contrast, the all incidence curves resulting from the weighted networks had different size peaks and epidemic durations. The incidence curves for both the weighted and unweighted the empirical networks were bimodal, while the incidence curves for the weighted and unweighted theoretical networks were unimodal. The incorporation of vaccination and isolation in the model caused a decrease in the cumulative incidence, however, this effect was only seen at high levels of vaccination and isolation for the complete network. This study highlights the importance of using

weighted empirical networks to describe contact patterns within populations that are unlikely to exhibit homogenous mixing such as equine populations.

5.2 Introduction

Disease transmission models are used to examine disease dynamics and the associated effects of implementing different intervention strategies in the population (1). In the models, infectious individuals can infect susceptible individuals at a rate, β , which comprises the probability of infection and the contact rate (2). Often, disease transmission models assume that the entire population has an equal probability of coming into contact (3). This assumption is referred to as the assumption of homogenous mixing, and, if not representative of the actual contacts that occurred in the population, can result in incorrect model predictions (3). One method of correcting for the assumption of homogenous mixing is to incorporate heterogeneous population mixing.

Heterogeneous contact patterns can be integrated in the model by defining contact networks (4). Contact networks describe the rate and frequency of the contacts that occurred between individuals in the population of interest. The simulation of disease transmission over a contact network is referred to as a network epidemic model (2). The nodes in network epidemic models represent individuals, while the edges between the nodes represent contacts between the individuals through which disease transmission can occur (5). While empirical networks describing the effective contacts in a population can be useful, it can be challenging to collect representative contact data required to

form the networks (6,7). Different methods have been used to simplify the process of simulating disease dynamics on networks including the use of summary statistics (i.e. network centrality) to generate representative theoretical networks (5). Common theoretical networks include random networks, small-world networks, and scale free networks. Random networks are characterized by random edge formation between nodes. Random networks tend to have small path lengths and minimal clustering (7). While small-world networks also have short path lengths, they are characterized by high clustering. Lastly, scale free networks follow a power law distribution such that most nodes have few connections while a few nodes have many connections (7,8).

Equine influenza (EI) is a highly transmissible respiratory disease caused by the equine influenza virus (9). The disease is most commonly spread by direct contact with an infectious horse, but transmission is also possible via indirect mechanisms such as: aerosols, wind, and fomites (9–11). While EI has a morbidity rate of up to 100% (10), it has a low mortality rate, and tends to be self-limiting (12).

Prevention and control strategies for equine influenza include vaccination of susceptible horses, quarantine of exposed horses, and/or isolation of infectious horses (13). There are currently two types of vaccines available in Canada; a vaccine containing the killed virus and a vaccine containing a modified live virus (14). Neither type of vaccine confers full protective immunity against infection (12). Since the vaccine efficacy is highly dependent on an individual's antibody levels against the glycoprotein haemagglutinin(10), the efficacy is improved when the vaccine strain matches the challenge strain (10). There are currently no regulatory requirements with respect to EI

vaccination or biosecurity measures in Canada. It is, however, advised that susceptible horses get vaccinated, and that horses are quarantined for a two-week period prior to being introduced into a new facility (14). Therefore, the population level immunity against EI is likely to differ based on the population of interest.

Although vaccination, quarantine, and isolation are recommended biosecurity practices for the prevention and containment of EI (9,15), existing disease transmission models for EI have focused primarily on vaccination as an intervention. These studies have modeled the transmission of EI in both vaccinated and unvaccinated populations (16–21). Furthermore, most published models have assumed homogenous mixing within the population, which may or may not correctly characterize the types of contact that generally occur within equine populations. The results of the existing studies have highlighted the importance of vaccination as a method for reducing the EI disease burden in equine populations.

The purpose of this study was to assess the suitability of using the assumption of homogenous mixing in equine disease transmission models, as well as to assess the effect of using both vaccination and isolation as intervention strategies. These research objectives were addressed by comparing the epidemiological outcomes of a network epidemic model, informed with empirical contact networks from 4 equine facilities, as well as theoretical networks. Given the different network structures, the study hypothesis was that the model simulations would result in different outcomes and trends for each network examined.

5.3 Methods

This study was approved by and conducted in accordance with the University of Guelph's Research Ethics Board (REB #16AP009) and the Animal Care Committee (AUP #3518). Informed consent was obtained from all study participants prior to the study deployment.

5.3.1 Empirical networks

Active radio-frequency identification (RFID) tags were used to collect contact pattern data from 4 equine facilities in southwestern Ontario, Canada, over a 7-day period between May and June 2017. Each facility contained 1 barn in which the horses' stalls were located, multiple pastures, and both indoor and outdoor riding areas. The facilities boarded between 20-28 horses and horses were either housed in the barn or in a pasture overnight (Table 5.1). The RFID tags were attached to the horses' halters which were kept with the animals at all times. Horses are gregarious animals and also curious by nature (22,23). Hence, to prevent accidental damage to the loggers and to ensure horse safety, the halters were placed on the pasture fence when the horses were in the pasture.

A detailed description of the data collection methods can be found in Chapter 2. Briefly, modified OpenBeacon RFID tags (Bitmanufactory Ltd., Cambridge, United Kingdom) were used to collect and record contact pattern data over the 7-day study period. When tags came within 2 meters of each other, the tags recorded the individuals that were in contact and the time of contact. A contact distance of 2 meters was used as a proxy for close contact sufficient for transmission of EI. The tags were collected at the end of the

study period, and the data were downloaded to a laptop computer where they were stored in a MySQL (Oracle Corporation, Redwood Shores, California) database. The data were aggregated into 24-hour periods and output as comma separated value (CSV) files which were imported into R version 3.3.0 (24) for analysis. The Statnet (25) suite of packages, in particular, the NetworkDynamic (26) and EpiModel (27) packages were used for the data preparation, model building, and simulation phases of the study.

The data collected with the RFID tags were used to generate a dynamic contact network for each of the 4 facilities, hereafter referred to as Empirical Networks 1-4. The dynamic networks consisted of 7 static networks, each network representing the contacts that occurred within a single study day. The edge weight in the contact networks represents the total contact duration (hours) between the respective horses on the study day (within a 24 hour period).

5.3.2 Contact network structure

Four types of contact networks were used to quantify the effect of different network structures on the disease transmission model outcomes. The following networks were used: 1) empirical networks (as described above), hereafter referred to as *Empirical Networks 1-4*, 2) a *Random Mixing Network (RMN)*, 3) a network with a small average degree, hereafter referred to as a "*Minimal Network*", and 4) a *Complete Network* (Figure 5.1). The edge weights for each of the theoretical networks were assigned from a truncated normal distribution such that the weights were all greater than or equal to 0. The distribution was generated using the average edge weight of each day for each facility (mean= 28.35 hours), and the standard deviation for each day of the study for

each facility (standard deviation= 33.35). In addition, the RMN was generated by averaging the characteristics of the empirical networks, namely, the network size, defined as the number of nodes in the network (28), and the average number of edges in the network. The network size for the RMN was 24. The expected number of edges ($N_{edges} = 87.6$) was calculated using the mean degree, representing the number of edges incident to a node of interest (28), of the empirical networks. The Minimal Network was formed by assuming that horses came in contact with an average of 4 other horses per day: the horses whose stalls were on either side of the horse of interest, an average of one horse in the centre aisle and an average of 1 horse in a common riding area. The Minimal Network had an average of 48 edges and is representative of facilities in which horses have reduced contact in order to avoid the potential for injuries (23). Lastly, the Complete Network was generated using the average network size of the empirical networks ($n=24$) and ensuring that all nodes came in contact with every other node in the network. Therefore, the Complete Network had 276 edges ($\frac{n(n-1)}{2}$), where n represents the number of nodes in the network.

Two types of analysis were completed: 1) the weighted networks were used to inform the contact rate of a disease transmission model in which no intervention strategies were considered, and 2) unweighted networks were used to inform the contact rate of a disease transmission model in which different prevention and control strategies were implemented. For both of these objectives, the 7 day-long static networks were combined to create 1 week-long dynamic network (each daily network was used to inform the contact rate for a single day of the model simulation). The model simulations,

however, were run for a 6-week period, creating the need for additional contact pattern data. Therefore, given the repetitive nature of the equine schedule, the empirical week-long dynamic contact networks were repeated to form a 6-week dynamic network. A similar process was used to generate the contact structure for the weighted theoretical networks used in the first study objective. In contrast, new unweighted networks were generated for each day of the simulation for the RMN, Minimal, and Complete Networks resulting in a non-repeating series of contact patterns for the duration of the simulations used in objective 2 above.

5.3.3 Equine influenza disease transmission model

The stochastic disease transmission model incorporated a typical SEIR structure and included both vaccination and isolation intervention strategies. A deterministic analogue of the model can be found in Figure 5.2. Model parameters were obtained from the peer-reviewed literature (Table 5.2). The following assumptions were used in the model formation:

1. Horses had no pre-existing immunity to EI.
2. Vaccinated horses were considered to be immune to infection by the start of the simulation.
3. The implications of an imperfect vaccine are that vaccinated horses have a reduced probability of becoming infected, a reduced infectious period, and a longer latent period than unvaccinated horses (19,29). These implications meant that vaccinated horses could both shed the virus and become infected with the

virus (12). Therefore, the transmission rate, β , the latent period, σ , and the recovery rate, γ , satisfied the following conditions: $\beta_v < \beta$, $\sigma < \sigma_v$, and $\gamma_v < \gamma$.

4. The probability of transmission (p) was reduced by 50% if at least one of the participating horses in the contact event (either a susceptible horse or an infectious horse) was vaccinated (16,19,29).
5. The model did not include waning immunity since waning immunity is not expected to occur until approximately 6 months after vaccination and the model was only run for 1.5 months (14,30).
6. The isolation period was always longer than the infectious period. Therefore, at the end of the isolation period, all horses were assumed to be fully recovered (i.e. horses could no longer shed the virus).
7. Infectious, vaccinated horses were assumed to be asymptomatic, and therefore were not isolated (31).
8. Given the projected impact of the weighted and non-weighted networks, different transmission probabilities were implemented for the 2 types of networks. For the base model in which weighted networks were used to inform the contact rate, the transmission probability (τ_1) was calculated using the transmission rate, $\beta = 1.85$ (19), combined with the contact duration for each empirical network averaged over each study day (\bar{k}). Therefore, $\beta = \bar{k}\tau_1$, τ_1 was calculated to be 0.06. In contrast, the transmission probability, τ_2 , for the vaccination and isolation models in which the contact rate was informed by an unweighted network, τ_2 , was assumed to be 100% (19).

The model formed with the unweighted networks incorporated 3 different prevention and control strategies that were compared to the baseline (no-intervention) model. The interventions considered were: 1) vaccination, 2) isolation, or 3) a combination of vaccination and isolation. The effect of vaccination on the epidemic outcome was studied by varying the proportion of horses vaccinated at the start of the simulation. Similarly, the effect of isolation was evaluated by varying the proportion of infectious, symptomatic horses isolated throughout the simulation. The proportion of horses isolated was used as a proxy for different levels of symptom severity. For example, it was assumed that horses with less severe symptoms would be less likely to be identified as infectious, and consequently, were less likely to be isolated. Horses were isolated for 14 days and were assumed to be fully recovered after the isolation period.

At the start of the model, a single, random horse was assigned an “infectious” status. Additionally, random horses were assigned a status of “vaccinated” according to the specified proportion to vaccinate. The occurrence of each transmission or transition event was determined using a binomial distribution. The probability of transmission for each contact was calculated as: $1 - (1 - k\tau)$, where k represents the respective edge weight and τ is the transmission probability. If an infectious, unvaccinated horse came in contact with a susceptible, unvaccinated horse, the infectious horse could transmit the disease to the susceptible horse with a probability p (Figure 5.2). If either the infectious horse or the susceptible horse was vaccinated, then the probability of transmission, p , was reduced by 50% (16,19,29). Infected horses could progress from

the exposed class to the infectious class at the end of the latent period which was 2.52 days for vaccinated horses and 1.75 days for unvaccinated horses.

Once infectious, unvaccinated horses could be isolated. All infectious horses moved to the recovered compartment at the end of the isolation or infectious period respectively. Recovered horses were assumed to be non-infectious with complete immunity to re-infection for the duration of the model time horizon.

The model, which was run over a 6-week period, was simulated 5000 times for the no-intervention model, and 10,000 times for the vaccination and isolation model using the initial conditions: $(S, E, I, R) = (N - 1, 0, 1, 0)$. For ease of comparison, the model results at each time point were averaged across all the simulations. Models were analyzed with respect to the cumulative disease incidence defined as the sum of the largest integer less than or equal to the simulated incidence at each time point divided by the total population size, and the daily incidence, and prevalence curves.

5.4 Results

5.4.1 No-intervention model using a weighted network

The prevalence curves resulting from the Empirical Networks were the most similar to the curve obtained using the minimal network (Figure 5.3 and Table 5.3). The smallest peak prevalence values (17- 26%) resulted from 3 of the 4 empirical networks (Empirical Networks 1,2, and 4). The epidemic duration for these curves was the longest (21- 24 days). Empirical Network 3 resulted in a similar epidemic curve to the Minimal Network (peak prevalence of 37%, and an 18-day epidemic duration). The Complete Network resulted in the largest peak prevalence (54%) and the shortest

epidemic duration (16 days). Furthermore, the Complete Network peaked before any of the other networks (day 3).

The incidence curves for the Empirical Networks were all bimodal while the incidence curves for the theoretical networks were all unimodal (Figure 5.3 and Table 5.3). The peak incidence for Empirical Networks 1, 2 and 4 over the simulation period was the smallest (5-15%), followed by the Minimal Network (16%), and Empirical Network 3 (25%). The Complete Network had the largest incidence peak (45%), preceded by the Random Mixing Network (37%). The generation of new cases was the shortest for the Complete Network (4 days) followed by the Random Mixing and Minimal Networks (6 and 8 days respectively). The generation of new cases was the longest for the (9-15 days).

5.4.2 Vaccination and isolation models informed with unweighted networks

5.4.2.1 Cumulative incidence

Similar trends were observed in the heat maps describing the cumulative incidence, defined as the total number of new infectious cases divided by the number of nodes in the network, in all the networks other than the Complete Network (Figure 5.4). In the majority of the networks, the implementation of an isolation-only strategy, in which infectious horses were isolated for a 14-day period, resulted in a distinct decrease in the cumulative incidence with an increase in the proportion of infectious horses isolated. However, an increase in the proportion of vaccinated horses in a vaccination-only strategy, had minimal effect on the cumulative incidence until vaccination rates of 50% had been reached. For the majority of the mixed vaccination-isolation intervention

strategies, an increase in the isolation rate had a greater effect on reducing the cumulative incidence than an increase in the vaccination rate. The incorporation of any isolation strategy in addition to a 100% vaccination rate had minimal effect on the cumulative incidence and resulted in a cumulative incidence of 50-64%. In contrast, increasing the vaccination rate when high levels of isolation (~50-100%) were used caused a decrease in the cumulative incidence. This trend was evident until a ~25-50% vaccination rate had been achieved, after which the cumulative incidence increased again.

The trends in the cumulative incidence for the RMN and the Minimal Network were similar to those of the empirical networks, however, an increase in the vaccination rate resulted in a more distinct decrease in the cumulative incidence than in the empirical networks. Furthermore, the non-linear effect of vaccination used in conjunction with an isolation program was less obvious in the Minimal Network than in the other networks. The implementation of a vaccination and/or isolation program decreased the cumulative incidence for both the RMN and the Minimal Network. In contrast, the implementation of any intervention program on the Complete Network was ineffectual unless a 100% vaccination or isolation rate had been achieved. The implementation of a 100% vaccination and/or isolation rate decreased the cumulative incidence to 87.5%. The Complete Network also resulted in a non-linear trend in cumulative incidence when a 100% isolation rate was implemented in addition to a vaccination program.

Lastly, the incorporation of an intervention program in the model resulted in the smallest cumulative incidence in the Minimal Network, followed by the Random Mixing Network.

The intervention strategies had a similar effect on the empirical networks. The Complete Network was the least affected by the implementation of any intervention strategy.

5.4.2.2 Epidemic curves

The prevalence curves for the empirical networks were similar in shape, with the exception of Empirical Network 2 which was bimodal (Figures 5.5 and S5.1). In all the empirical networks, an increase in the vaccination and isolation rates resulted in a decrease in the disease prevalence. The difference in the prevalence with increasing vaccination rates was the smallest at a 100% isolation rate. The prevalence within each isolation level decreased with an increase in the vaccination rate. For the remaining networks, an increase in the isolation (up to a level of 75%) and vaccination rates caused a decrease in the prevalence. However, for isolation levels of at least 75%, the RMN and Minimal Network presented minimal to no difference in prevalence with an increase in the vaccination rate.

In all cases, other than when an isolation program targeting 75-100% of the infectious population was implemented, the incidence curves for the empirical networks all had at least 2 distinct peaks (Figures S5.2 and S5.3). Furthermore, an increase in the vaccination rate caused a decrease in the first peak height, however, this effect was reversed for the second peak. In contrast, the incidence curves for the Complete, RMN, and Minimal Networks were unimodal. The peak incidence for the Complete Network was the largest in the absence of a vaccination scheme. However, the length of time in which new cases were being generated was smaller in the absence of a vaccination program than at larger vaccination rates. In general, an increase in the vaccination rate

caused a decrease in the incidence until vaccination rates of 75-100% had been achieved. Similar trends were observed in the RMN as the Complete Network, however, the RMN exhibited a smaller incidence for all strategies than the Complete Network. Lastly, the shape of the incidence curve for the Minimal Network was vastly different from the remaining networks. While the RMN, Empirical, and Complete Networks had a distinct peak, the Minimal Network tended to plateau at the peak incidence.

5.5 Discussion

Limited data exists describing herd-level mixing patterns of agricultural animals (32). This data deficiency extends to the equine population making it difficult to develop realistic disease transmission models. Therefore, it is not surprising that few mathematical modelling studies have focused on facility-wide transmission dynamics of equine influenza. Furthermore, the models that have focused on EI tended to assume homogenous mixing patterns within the population of interest. Until recently, no published data existed to substantiate this assumption, however, a recent study by the authors, whose focus was the collection and analysis of contact pattern data, indicated that the equine mixing patterns within the participating equestrian facilities did not satisfy the homogenous mixing assumption (Chapter 4). The current study used data collected with RFID tags to quantify the effects of heterogenous mixing patterns on the dynamics of disease spread.

The networks used in this study were used to assess the effect of two characteristics on disease spread dynamics: 1) weighted vs. unweighted networks, and 2) empirical vs.

theoretical networks. It was hypothesized that the different network structures would result in distinct epidemiological outcomes. While this hypothesis was incorrect when comparing the dynamics of the unweighted networks, it was correct with respect to the weighted networks, in which all of the resulting epidemic curves had different peaks and durations. The similarity in the disease dynamics for the unweighted networks can be explained by the law of mass action. The law of mass action states that the number of new infections when susceptible individuals (S) come in contact with infectious individuals (I) is proportional to SI (33). The proportionality constant is often referred to as the transmission rate, β , and consists of the contact rate and the probability of transmission given an effective contact (34,35). In each participating facility, groups of horses were turned out together, forming a subpopulation in which homogenous mixing could occur. During the turnout period, the horses' halters and the attached RFID tags were stored together on the pasture fence, inferring that the horses were in contact for the duration of the turnout period. The horses remained in their pastures for the majority of the day, until they were returned to their stalls in the evening. The horses were removed from their pastures for training and pleasure riding, and during this time, could come in contact with horses from the other pastures. Therefore, it is likely that the law of mass action occurred on a smaller scale, within the pastures, and could be the reason for the similarity in the epidemic outcomes observed. In contrast, in the weighted networks, the added characteristic of edge weight contributed to the population level heterogeneity and likely influenced the difference in the disease dynamics observed between the empirical networks.

The degree distribution directly correlates with disease propagation due to its relationship with the transmission rate (36–38). Furthermore, the number of direct contacts that each node had in the unweighted empirical networks produced consistent patterns in the relative degree of each horse compared to the other horses in the network (Chapter 4). Since it has been shown that local structure slows disease spread when compared to random mixing (39), the underlying structure in the degree distribution of the empirical networks can explain the discrepancy in the shape of the incidence curves between the theoretical and the weighted empirical networks.

With respect to the effect of incorporating different interventions in the model, the RMN produced similar results to the unweighted empirical networks. This is unsurprising since the RMN was generated by averaging the characteristics of the empirical networks. In contrast, the Minimal Network had a smaller degree than the empirical networks, resulting in a reduced transmission rate, which in turn caused a decrease in the disease incidence. Therefore, horses were more likely to be isolated before they could transmit the disease to other horses. Conversely, the high node degree in the Complete Network caused an associated increase in the transmission rate. The increased transmission rate directly affects the law of mass action, resulting in an increased disease incidence. Given the high contact rate, it follows that the implementation of an imperfect vaccine would be ineffective unless used at high levels. Furthermore, the high contact rate caused the disease incidence to increase to 100% by the second day of infection, making it impossible to isolate the horses before they infected other horses.

The implementation of a vaccination and isolation program in combination in the model resulted in unexpected patterns in the cumulative incidence. The cumulative incidence generally decreased with an increase in the respective intervention strategy. However, higher isolation rates used in addition to a vaccination program lead to a non-linear trend in the cumulative incidence with increasing levels of vaccination. The non-linear trends in the cumulative incidence are likely attributable to a backward bifurcation. A detailed description of a backward bifurcation can be found in (40). Briefly, bifurcation analysis relates the stability of a system to the basic reproduction number, R_0 (40,41). An $R_0 < 1$ is generally indicative of possible disease eradication, while an $R_0 > 1$ implies disease persistence (41,42). Consequently, a health program should endeavour to maximize the realistic intervention strategies in order to decrease R_0 to less than 1. However, a backward bifurcation implies that decreasing the R_0 to less than 1 will not necessarily eliminate a disease (40,41). Known model structures can cause a backward bifurcation, including the incorporation of an imperfect vaccine, multiple groups such as multiple susceptible groups with different characteristics, and differential susceptibility to infection (40,41,43). Given the trends in the cumulative incidence with respect to the vaccination rate it is likely that the model is portraying a backward bifurcation. Therefore, the model results indicated a specific level of vaccination which will minimize the cumulative incidence, however, a higher or lower vaccination rate will cause an increase in the infection incidence. Mechanistically, this non-linear relationship can be explained by the presence of the imperfect vaccine. At high levels of vaccination, a large proportion of the population is likely to be asymptomatic shedders, inhibiting the

ability to isolate the infected horses, and increasing disease transmission within the population despite high vaccine coverage in the population.

With respect to the theoretical networks, the cumulative incidence of the unweighted Minimal Network was the smallest, and the cumulative incidence of the unweighted Complete Network was the largest. The relative cumulative incidence for each of the unweighted theoretical networks can be attributed to the network degree which was smallest for the Minimal Network, and largest for the Complete Network. These results correspond to the study by May et al. (2001), in which the authors examined the effect of increasing the average degree with respect to the total number of individuals that became infected during the course of the epidemic. The authors concluded that an increase in the network connectivity corresponded to an increase in the number of individuals infected. Therefore, in networks with a high connectivity, all individuals became infected (39).

The epidemiological results of this study were consistent with the results of Glass et al. (2002) in which the authors used a stochastic model to describe the effects of vaccination on the epidemiology of EI. Glass et al. (2002) concluded that vaccine uptake decreased the epidemic size. The model was validated with data obtained from a New York race track. The validation data indicated that the incidence curve was bimodal with new cases occurring for a duration of 1 month. The fitted model, however, was unimodal with the development of new cases for a duration of 20 days. The incidence curves from the current study were also bimodal (for the weighted empirical networks), with new cases occurring for 9-15 days. One possible explanation for the

discrepancy in the duration of the generation of new cases is the stochastic nature of the model. The study results were averaged over 5000 simulations, meaning, that some of the longer epidemic durations were likely consistent with the results from the Glass study. Furthermore, race facilities are managed differently than sport facilities, and these differences may contribute to the difference in results. For example, in order to avoid injury, horses at a racetrack may be turned out individually as opposed to the group turnouts practiced in the participating equine facilities.

5.5.1 Study limitations

While the study results imply an ability to generalize biosecurity strategies for equine facilities, there are limitations. First, the network sizes were relatively small. In their study, Glass et al. concluded that the population size did not affect the model outcomes, however, this result might not hold for heterogenous mixing patterns. Second, the assumption that horses were in contact during turnout might have over-estimated the contact rates for the horses that shared a pasture. Tracking the actual contacts between the horses that shared a pasture might result in slightly different results, as the number of horses that each horse contacted might decrease. Lastly, the complexity of incorporating time varying networks in the model made the application of certain analyses difficult, including the calculation of R_0 , a bifurcation analysis, and the use of sensitivity analyses such as Latin Hypercube Sampling and Partial Rank Correlation Coefficients.

Regardless of these limitations, the importance of the underlying contact network characteristics on the epidemiological outcomes of disease spread is evident. The

similarity in the management styles and schedules of the participating facilities, and the associated model results enables the generalization of prevention and control strategies for this type of facility. Furthermore, this study emphasizes the importance of using detailed empirical networks to describe the contact structure for populations such as equine populations which are unlikely to exhibit homogenous mixing.

5.6 References

1. Lofgren E, Halloran ME, Rivers CM, Drake JM, Porco TC, Lewis B, et al. Opinion: Mathematical models: A key tool for outbreak response. *Proc Natl Acad Sci*. 2014;111(51):18095--18096.
2. Craft ME. Infectious disease transmission and contact networks in wildlife and livestock. *Philos Trans R Soc B Biol Sci*. 2015;370(1669):20140107–20140107.
3. Perisic A, Bauch CT. Social contact networks and disease eradicability under voluntary vaccination. *PLoS Comput Biol*. 2009;5(2):e1000280.
4. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *J Theor Biol*. 2005;235(2):275–88.
5. Pellis L, Ball F, Bansal S, Eames K, House T, Isham V, et al. Eight challenges for network epidemic models. *Epidemics*. 2015;10:58–62.
6. Krivitsky PN, Morris M. Inference for social network models from egocentrically sampled data , with application to understanding persistent racial disparities in HIV prevalence in the US. *Ann Appl Stat*. 2017;11(1):427–55.
7. Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005;2(4):295–307.
8. Barabasi A-L, Bonabeau E. Scale-Free Networks. *Sci Am*. 2003;288(5):60–9.
9. Myers C, Wilson WD. Equine influenza virus. *Clin Tech Equine Pract*. 2006;5(3):187–96.
10. vanMaanen C, Cullinane A. Equine influenza virus infections: an update. *Vet Q*. 2002;24(2):79–94.
11. Constable PD, Hinchcliff KW, Done SH, Gruenberg W. *Veterinary medicine*. 11th ed. W.B. Saunders; 2017.
12. Cullinane A, Elton D, Mumford J. Equine influenza - surveillance and control. *Influenza Other Respi Viruses*. 2010;4(6):339–44.
13. OIE World Organisation for Animal Health. Equine influenza: general disease information sheets [Internet]. [cited 2018 Feb 4]. Available from: <http://www.oie.int>
14. Wright B, Kenney D. Influenza in horses [Internet]. 2011 [cited 2018 Mar 13]. Available from: <http://www.omafra.gov.on.ca/>
15. Equine Guelph. Equine biosecurity risk calculator [Internet]. [cited 2018 Jan 5]. Available from: http://www.equineguelph.ca/Tools/biosecurity_calculator.php
16. Glass K, Wood JLN, Mumford J a, Jesset D, Grenfell BT. Modelling equine influenza 1: a stochastic model of within-yard epidemics. *Epidemiol Infect*.

2002;128(3):491–502.

17. Park AW, Wood JLN, Daly JM, Newton JR, Glass K, Henley W, et al. The effects of strain heterology on the epidemiology of equine influenza in a vaccinated population. *Proc R Soc London B Biol Sci.* 2004;271(1548):1547–55.
18. De La Rua-Domenech R, Reid SWJ, González-Zariquiey AE, Wood JLN, Gettinby G. Modelling the spread of a viral infection in equine populations managed in Thoroughbred racehorse training yards. *Prev Vet Med.* 2000;47(1–2):61–77.
19. Park AW, Wood JLN, Newton JR, Daly JM, Mumford JA, Grenfell BT. Optimising vaccination strategies in equine influenza. *Vaccine.* 2003;21(21–22):2862–70.
20. Baguelin M, Newton JR, Demiris N, Daly JM, Mumford JA, Wood JLN. Control of equine influenza: scenario testing using a realistic metapopulation model of spread. *J R Soc Interface.* 2010;7(42):67–79.
21. Garner MG, Cowled B, East IJ, Moloney BJ, Kung NY. Evaluating the effectiveness of early vaccination in the control and eradication of equine influenza--a modelling approach. *Prev Vet Med.* 2011;99(1):15–27.
22. Lloyd AS, Martin JE, Bornett-Gauci HLI, Wilkinson RG. Horse personality: Variation between breeds. *Appl Anim Behav Sci.* 2008;112(3–4):369–83.
23. Majecka K, Klawe A. Influence of paddock size on social relationships in domestic horses. *J Appl Anim Welf Sci.* 2018;21(1):8–16.
24. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available from: <https://www.r-project.org/>
25. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. *statnet: Software tools for the Statistical Modeling of Network Data.* 2003.
26. Butts CT, Leslie-Cook A, Krivitsky PN, Bender-deMoll S. *networkDynamic: Dynamic extensions for network objects* [Internet]. 2016. Available from: <https://cran.r-project.org/package=networkDynamic>
27. Jenness S, Goodreau SM, Morris M. *EpiModel: Mathematical modeling of infectious disease dynamics* [Internet]. 2018. Available from: <https://cran.r-project.org/package=EpiModel>
28. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis.* 2009;56(4):109–20.
29. Daly JM, Newton JR, Wood JLN, Park AW. What can mathematical models bring to the control of equine influenza? *Equine Vet J.* 2013;45(6):784–8.
30. Daly JM, Murcia PR. Strategic implementation of vaccines for control of equine

- influenza. *Equine Vet J*. 2018;50(2):153–4.
31. Lewis NS, Daly JM, Russell CA, Horton DL, Skepner E, Bryant NA, et al. Antigenic and genetic evolution of Equine Influenza A (H3N8) Virus from 1968 to 2007. *J Virol*. 2011;85(23):12742–9.
 32. Chen S, White BJ, Sanderson MW, Amrine DE, Ilany A, Lanzas C. Highly dynamic animal contact network and implications on disease transmission. *Sci Rep*. 2014;4:4472–9.
 33. Heesterbeek H. The law of mass-action in epidemiology: a historical perspective. *Ecol Paradig*. 2005;81–104.
 34. McCallum H, Barlow N, Hone J. How should pathogen transmission be modelled? *Trends Ecol Evol*. 2001;16(6):295–300.
 35. Allen LJ, Brauer F, Van den Driessche P, Wu J. *Mathematical epidemiology*. Cachan JMM, Groningen FT, Paris BT, editors. Springer; 2008.
 36. Newman MEJ. Spread of epidemic disease on networks. *Phys Rev E*. 2002;66(1):16128.
 37. May RM. Network structure and the biology of populations. *Trends Ecol Evol*. 2006;21(7):394–9.
 38. Scott J. *Social network analysis*. SAGE Publications Limited; 2017.
 39. May RM, Lloyd AL. Infection dynamics on scale-free networks. *Phys Rev E*. 2001;64(6):066112–6.
 40. Gumel AB. Causes of backward bifurcations in some epidemiological models. *J Math Anal Appl*. 2012;395(1):355–65.
 41. Brauer F. Backward bifurcations in simple vaccination models. *J Math Anal Appl*. 2004;298(2):418–31.
 42. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface*. 2005;2(4):281–93.
 43. Arino J, McCluskey C. C, van den Driessche P. Global results for an epidemic model with vaccination that exhibits backward bifurcation. *SIAM J Appl Math*. 2003;64(1):260–76.

5.7 Tables

Table 5.1: Demographic information for the participating equestrian facilities. The facilities ranged in size, and discipline. Vaccination rates for equine influenza ranged from 46-100%.

	Facility 1	Facility 2	Facility 3	Facility 4
Number of horses boarded at the facility	20	28	27	24
Number of participating horses	20	28	26*	24
Study date	May 2017	May-June 2017	June 2017	July 2017
Facility discipline	Dressage Eventing Jumping	Dressage Jumping	Dressage	Therapeutic riding
Proportion of horses vaccinated for equine influenza (%)	85	100	46.2	100

*Although all horses were enrolled in the study, one horse left the facility for the duration of the study, and therefore was not included in the study

Table 5.2: List of parameters and variables used in the network epidemic model. Parameter values were obtained from the peer-reviewed literature. The parameter symbol represents parameters found in the deterministic analogue (Figure 5.2).

Parameter/ variable	Interpretation	Value	Citation
ρ	Proportion of the population that was vaccinated	0-1 (%)	Assumed
p_v	Infection probability for vaccinated horses	$0.5*\tau$	(19,29)
p, τ_1	Infection probability for unvaccinated horses used in the unweighted networks	1	(19)
τ_2	Infection probability for unvaccinated horses used in the weighted networks	0.06	Calculated
c	Number of acts per encounter	1	Assumed a binary network
σ_v	Latent period for vaccinated horses	2.52 days	(19,29)
σ	Latent period for unvaccinated horses	1.75 days	(19,29)
γ_v	Recovery period for vaccinated horses	2.5 days	(19,29)
γ	Recovery period for unvaccinated horses	4.8 days	(19,29)
θ	Duration of isolation	14 days	Assumed
η	Proportion of infectious horses isolated	0-1 (%)	Assumed
S	Susceptible population		
E	Exposed population, i.e. horses that are infected but not yet infectious		
I	Infectious population		
R	Recovered population		
Q	Isolation compartment for infectious, symptomatic horses		

Table 5.3: Key epidemiological results for the disease transmission model of equine influenza. The weighted, no-intervention model had an SEIR structure and did not include any disease prevention or control interventions. The epidemic model was simulated on different weighted networks including 4 empirical networks (“Empirical Networks 1-4”), a network with the average characteristics (average degree, and network size) of the empirical networks (“Random Mixing Network”), a Complete Network in which all the nodes were connected to each other, and a network with an average degree of 4 (“Minimal Network”). The empirical networks were the most similar to the Minimal Network. The Complete Network had the largest epidemic peak and the shortest epidemic duration.

	Size of epidemic peak		Time of epidemic peak (day)		Duration (days)	
	Prevalence (%)	Incidence (%)	Prevalence curve	Incidence curve	Epidemic duration	Time until no new infections
Empirical Network 1	20.0	5.0	9.0	8.0	21.0	10.0
Empirical Network 2	17.8	10.7	10.0	8.0	24.0	15.0
Empirical Network 3	37.5	25.0	5.0	2.0	18.0	9.0
Empirical Network 4	26.9	15.3	8.0	3.0	21.0	10.0
Random Mixing Network	45.8	37.5	5.0	3.0	18.0	6.0
Complete Network	54.2	45.8	3.0	2.0	16.0	4.0
Minimal Network	37.5	16.7	6.0	4.0	19.0	8.0

5.8 Figures

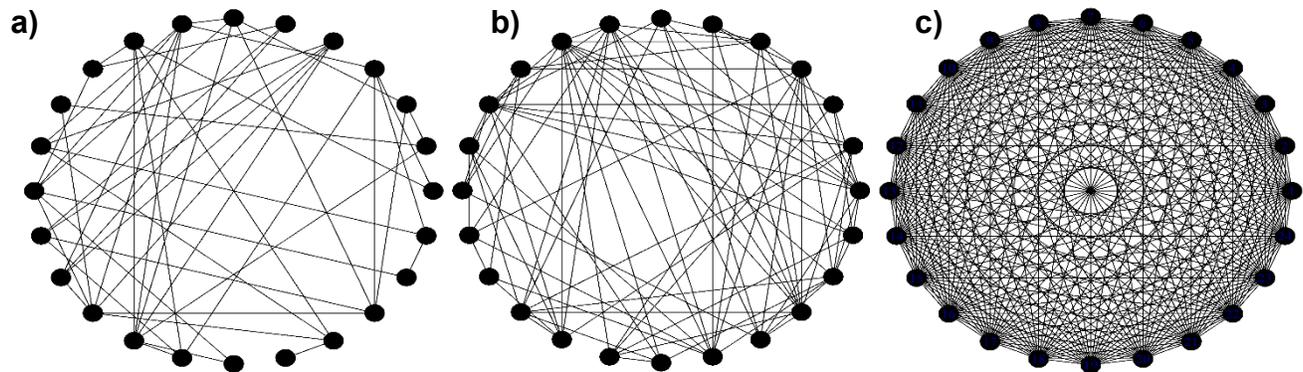


Figure 5.1: Sample theoretical networks used in the network epidemic model. Panel a) represents the Minimal Network which has an average of 48 edges. Panel b) represents the Random Mixing Network which has an average of 87.6 edges. Panel c) represents a Complete Network in which each node is connected to every other node.

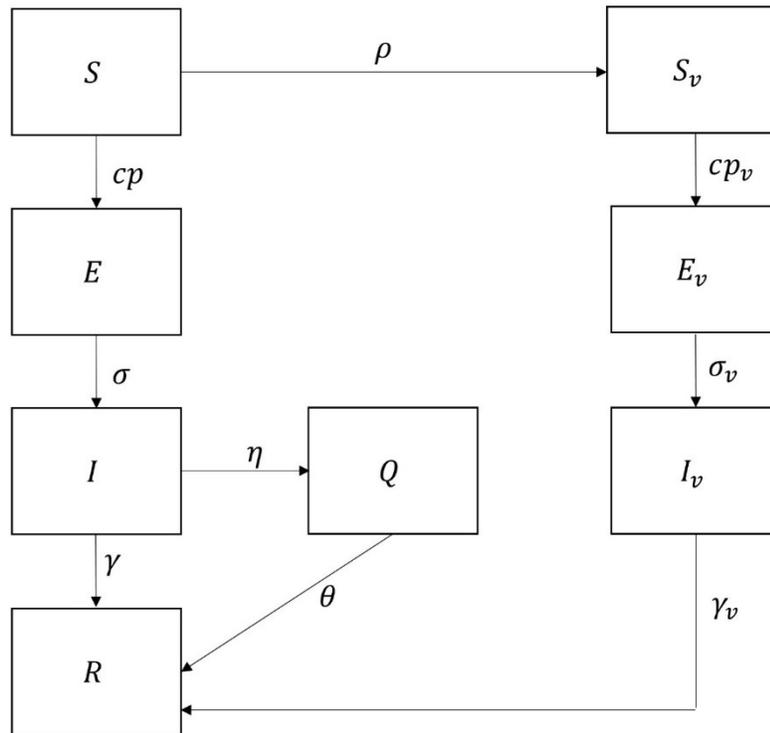


Figure 5.2: Deterministic analogue of the study model. In this model, horses start out susceptible to infection (S). A proportion (ρ) of the susceptible horses are vaccinated (S_v). Both vaccinated and unvaccinated horses can come in contact. Furthermore, both vaccinated and unvaccinated horses can become infected (E), however the rate of infection differs between the vaccinated horses (transmission probability = p_v) and unvaccinated horses (transmission probability = p). The transmission rate of the pathogen between infectious horses (I) and susceptible horses is determined through the contact rate (c) and the transmission probability. At the end of the infectious period (σ and σ_v for vaccinated and unvaccinated horses respectively), horses can recover (R), where it is assumed that they have full immunity to re-infection. A proportion (η) of the unvaccinated, infectious horses (I) are isolated (Q). These horses remain in isolation for a pre-determined length of time (θ), before they move to the recovered compartment.

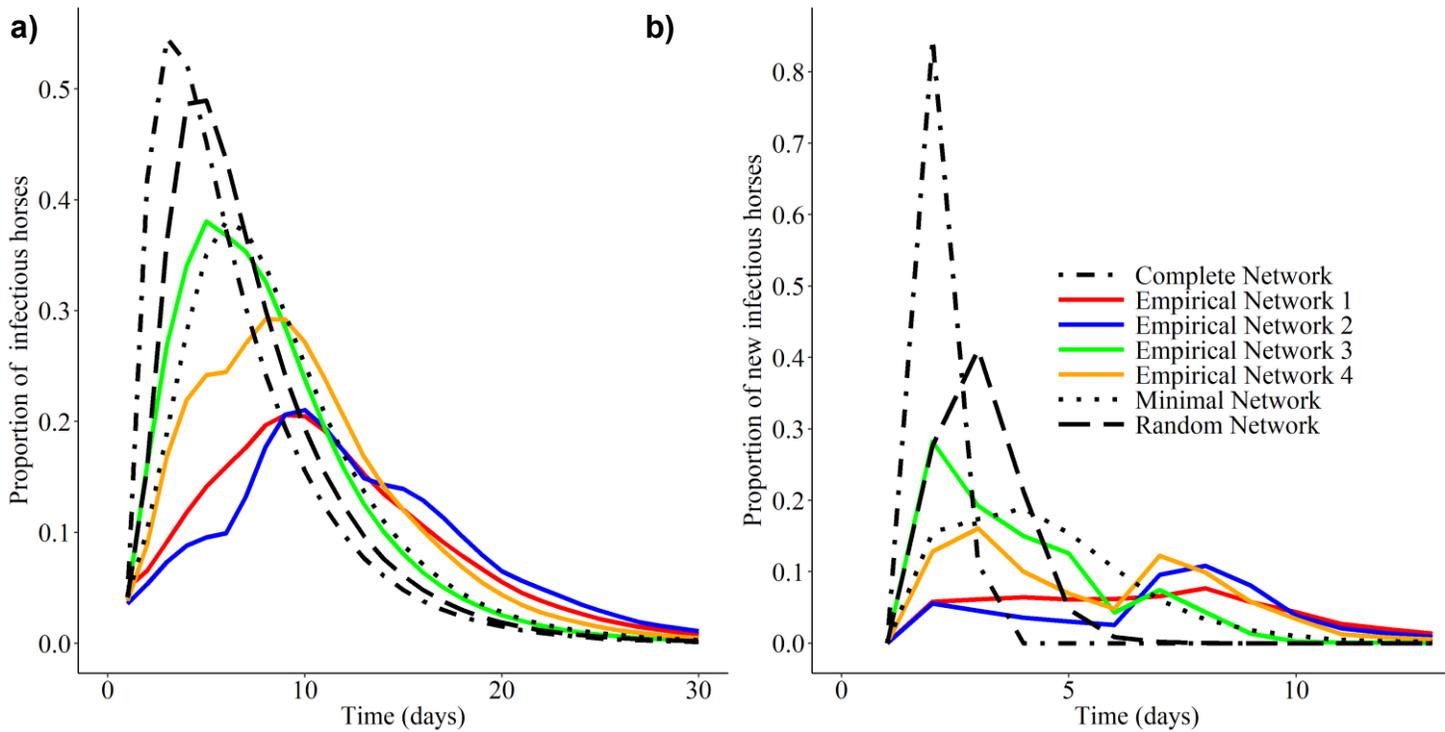


Figure 5.3: Epidemiological results for the model in the absence of interventions. Panel a) contains plots of the prevalence over time, and panel b) contains plots of the incidence over time. Similar shaped curves were produced for each empirical network. The peak incidence and prevalence for the Complete and Random Networks were larger than the Empirical Networks, and the epidemic duration was shorter than the empirical networks. The peak incidence and prevalence for 3 of the empirical networks were the smallest, and the epidemic duration for these networks was the longest.

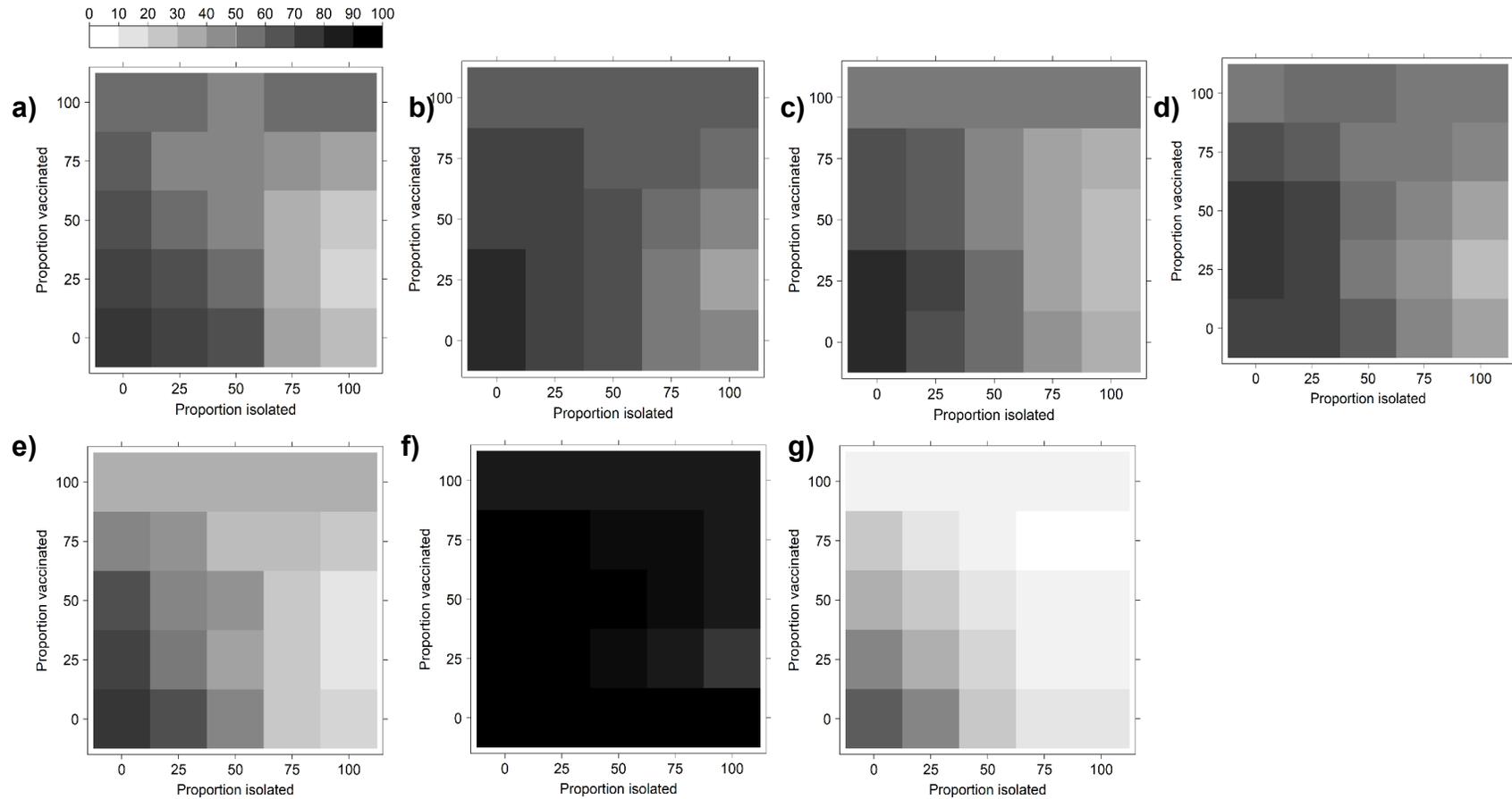


Figure 5.4: Heat maps of the cumulative incidence (%) when isolation and vaccination were implemented in the population. The heat maps represent Empirical Networks 1-4 (panels a-d), the Random Mixing Network (panel e), the Complete Network (panel f), and the Minimal Network (panel g). Increasing the proportion of the population isolated and/or vaccinated generally decreased the cumulative incidence. However, a non-linear trend in the cumulative incidence was observed for high isolation rates used in addition to a vaccination program. Other than a 100% vaccination and isolation rate, the implementation of any intervention strategy had minimal effect on the cumulative incidence for the Complete Network.

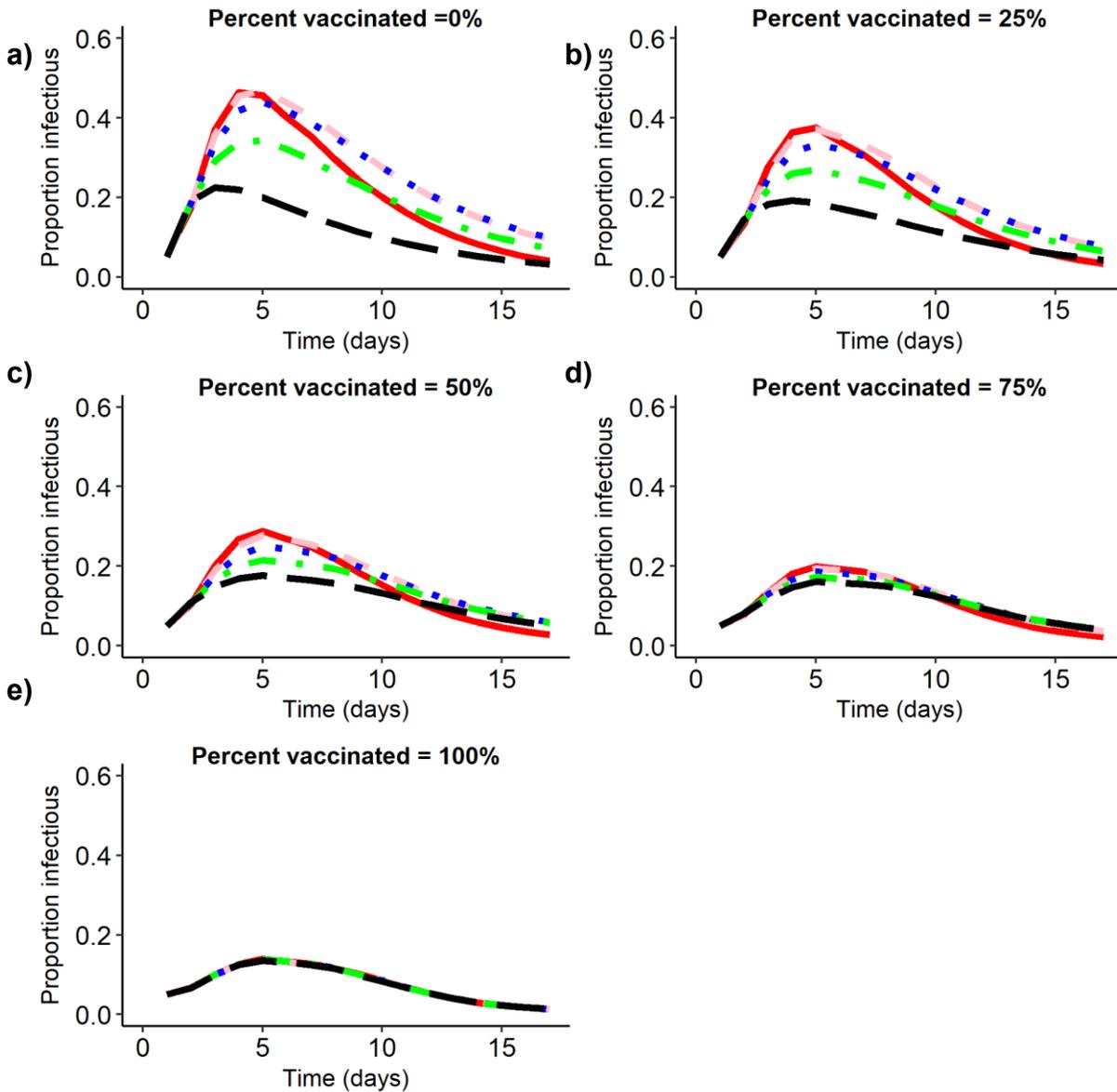


Figure 5.5: Prevalence plots for Empirical Network 1. Empirical Networks 2-4, the RMN, and the Complete and Minimal Networks resulted in similar trends. The plots show the effect of varying the proportion of the infectious population that is isolated as the proportion of the population that is vaccinated is kept constant. The proportion isolated is portrayed with different line colours: red solid line=0%, pink dotted line=25%, blue dotted line= 50%, green dotted line=75%, and black solid line=100%. Panels a)-e) represent a population in which 0, 25,50, 75, and 100% of horses are vaccinated. For vaccination rates of 0-50%, increasing the isolation rate decreased the disease prevalence. However, the difference in prevalence due to the increasing isolation rates decreased as the proportion vaccinated increased. By a 75% vaccination rate, there was no difference in the prevalence with increasing isolation rates.

CHAPTER 6

Discussion

The overarching rationale behind this body of work was to investigate the suitability of using the assumption of homogenous mixing to represent the contact rate in equine populations. Given the challenges associated with the collection of contact pattern data in animal agricultural populations, this thesis presents important concepts and results regarding the necessity for the utilization of empirical contact pattern data through the documentation of data collection and analysis methods. The thesis involved a series of studies, each building upon the previous study, as well as on the necessary knowledge for the formation of representative disease transmission models for equine populations.

The work presented in this thesis began by describing the modified radio-frequency identification (RFID) tag functionality and methods of use (Chapter 2). Network analysis techniques were used in conjunction with survey-based data to validate the plausibility of the contact data collected with the tags. Following the validation stage, the collected data were used in a pilot study to assess the feasibility of the study design and to document the types of analysis possible with the collected data (Chapter 3).

Subsequent to the initial test phase of the technology, the RFID tags were used to collect contact pattern data for 4 equine facilities in Ontario, Canada (Chapter 4). The collected data were used to characterize and compare the contact structure of the participating facilities. Given the non-homogenous nature of the empirical contact networks, the contact pattern data collected in Chapter 4 were used to inform a network epidemic model for equine influenza (Chapter 5). The network model was used to

validate the effect of the different contact structures on the resulting epidemiological outcomes. Lastly, the current chapter (Chapter 6) focuses on summarizing the main findings of the studies, as well as discussing the study limitations and opportunities for future work.

6.1 Summary of key findings

6.1.1 Feasibility of using the modified RFID tags in agricultural settings

The modified RFID technology presented a reliable, accessible alternative to other data collection methods. From the study participants' point of view, the technology was non-invasive, and did not affect the daily management of the facility (Chapter 2). The most challenging aspect of the study for the participants in maintaining compliance with respect to study design, was to remember to keep the halters with the horses when they were being moved from location to location. However, the participants reported that the study implementation caused no concern with respect to horse or human safety.

From a logistical standpoint, the tags were small and inexpensive. The study deployment required minimal researcher involvement, limiting researcher interference in the daily operating procedures of the facility. The most time-consuming aspect of the deployment involved the tag preparation for each facility and included: 1) assigning the tags to the participants, 2) creating a representative map of the facility, 3) attaching the tags to the horses' halters, 4) collecting the tags at the completion of the data collection period, 5) uploading the data onto the computer, and 6) clearing the tags' memory (Chapter 2).

In terms of the quality and quantity of the collected data, the tags recorded large amounts of data (up to 524,288 interactions) that were subsequently validated with survey-based information (Chapters 2 and 3). The majority of the collected data was in agreement with the survey-based data, however, there were some limitations noted, as well as some inconsistencies in the recorded data. For example, tags were fixed to key locations around the facility such as the wash stall, barn areas, and riding areas. Although these tags were stationary during the study period, one wash stall tag was recorded as having come in contact with a tag located within the barn area. This type of inconsistency did not occur often and is likely attributable to the limitations inherent within radio-wave technology, such as the proclivity for signal distortion and interference (1). Furthermore, many of the tag batteries lost charge by the end of the study period (7 days). For example, 12-56% of the tags deployed within the participating facilities recorded data for less than 6 days of the study, while 44-88% of the tags stopped recording data during the 6th study day. The short recording period can be attributed to minimal use (i.e. the tags that were located at the facility perimeter were only activated if a horse left the facility during the study period), while the short battery life can be attributed to the colder temperatures combined with the high communication frequencies (Chapter 2). High communication frequencies were especially common for the tags that were used by human participants, as these tags were stored together when not in use, resulting in a constant signal transmission and recording process.

6.1.2 Network analysis and the resulting network characteristics

The data were aggregated into 24-hour periods based on the assumption that most equine facilities have routine standard operating procedures that result in a 24-hr repetitive pattern, as well as the plausible variation in the contact patterns during this time (Chapters 2-4). The resulting networks were analyzed using both traditional network analysis techniques, such as the calculation of centrality measures, as well as non-traditional methods, such as the assessment of the generalizability and predictability of the contact patterns (Chapters 3-4). The comparison of the network characteristics between the larger facilities that boarded 20-28 horses, identified similar trends for each of the facilities (Chapter 4). Briefly, the network density seldom exceeded 0.5, implying that the majority of horses contacted less than 50% of the horses within the facility. Furthermore, while the week-long series of static networks (1 per day) for each facility contained multiple components, the static, day-long networks were not always comprised of multiple components. None of the networks resulted in contact patterns that would be considered representative of homogenous mixing within the population, with the majority of the recorded contact occurring between horses that shared a pasture. The contact patterns were relatively similar between study days for a given facility, with 11- 62% of identical contacts recurring within the study period. Similarly, survey collected data regarding a horse's stall and pasture location enabled the prediction of ~9-80% of the RFID recorded contacts. These ranges are large and are therefore not optimal for high accuracy predictions of contact patterns. Furthermore, the survey-based network that was used to predict the contact patterns was assumed to

be static throughout the week. This assumption is unrealistic as horses can be assigned to different pastures and stalls over time. Regardless of these limitations, given the importance of heterogeneous contact patterns on disease dynamics, the ability to predict even small amounts of contact pattern data is useful for situations where minimal data is available.

In terms of the node-level centrality, for a given facility, horses with a higher degree relative to the other horses in the network on any given day, always maintained a higher degree than the other horses throughout the study period (Chapter 4). The normalized degree ranged between 0.00-0.80, 0.00-0.96, 0.00-0.68, and 0.00-0.83 for each facility, while the betweenness ranged between 0.00- 0.60, 0.00-1.00, 0.00-0.71 and 0.00-0.78 for each facility. These results indicate the possibility of a large range of epidemiological outcomes given the introduction of a disease into the population. For example, the degree, which refers the number of direct contacts that a node of interest had in the network (2), directly relates to the routes of direct disease transmission. Consequently, if a horse with a degree of 0 were to become infectious (e.g. when the horse left the facility), then the horse would be unable to transmit the disease to any other horses. Alternatively, if a horse with a normalized degree of 0.8 were to become infectious, then there is a possibility that the horse would transmit the disease to 80% of the population. Betweenness represents the probability that the shortest path between any two nodes will pass through the node of interest (2). Nodes with a high betweenness are said to be cut-points, i.e. nodes whose removal will result in the segmentation of the network (3), thus inhibiting transmission between the network

segments. Therefore, the identification of nodes with a high betweenness, e.g. 0.78, can be used to control and minimize disease spread within the network.

Lastly, an exponential random graph model analysis did not produce consistent results for all 4 facilities. However, for 2 of the 4 facilities, the distance parameter, measuring the Euclidean distance between horses' stalls or pastures, was always negatively correlated with the probability of forming a contact. This result was statistically significant (Chapter 4). The lack of consistency between facilities makes it difficult to infer a relationship between the distance parameter and the probability of edge formation. However, in the two facilities for which the distance parameter was significant, the majority of the contact occurred between horses that shared a pasture. The relationship between the distance variable and the probability of edge formation for these two facilities implies a potential correlation between the horses that shared a pasture and the probability of edge formation. This probable correlation is important as most of the horses in the participating facilities shared a pasture with other horses. Moreover, the horses in each participating facility were in the pasture for the majority of the day, and some horses stayed in the pasture overnight. Since horses are gregarious animals (4), significant amounts of contact is likely to occur between the horses that shared a pasture, resulting in the potential for disease transmission between the horses within the pasture. Therefore, further research should be conducted to quantify the implications of shared pastures on the probability of edge formation and the associated disease dynamics.

6.1.3 Necessity of using the RFID tags to collect and record contact data

The work presented in this thesis evaluated different aspects of the utility of collecting and utilizing empirical contact data. Briefly, although the empirical networks were not representative of homogenous mixing, the associated network models resulted in similar epidemiological outcomes. Hence the interpretation of the results are variable. However, overall, the added structure within the empirical networks compared to theoretical networks contributed to the disease dynamics resulting from the network model simulations, demonstrating the importance of using non-homogeneous, empirical mixing patterns for populations that are unlikely to exhibit random mixing.

The Jaccard similarity index was used to assess the generalizability and predictability of the networks. With respect to the pilot study, 53.6-73.9% of the horse-to-horse contacts recurred over the study period (Chapter 3). Furthermore, interpretation of data detailing each horse's stall and pasture location enabled the prediction of 50-75% of the tag-recorded contact events (Chapter 3). Similar trends in contact event prediction were recorded in the larger facilities in which 32.6-55.7%, 11.1-55.5%, 11.5-55.5%, and 35.0-62.3% of contacts were repeated throughout the study week for facilities 1-4 (Chapter 4). Knowledge about the facilities' standard operating procedures enabled the prediction of 24.2-40.0%, 8.7-62.4%, 13.2-51.9%, and 45.9-61.4% of the daily contact patterns for facilities 1-4, respectively. In light of the large prediction ranges, the gold standard would be the collection of a full set of detailed contact data. However, for logistical reasons, it is not always feasible to collect such data. These results indicate that in the

case of little to no available data, it is possible to infer contact patterns given knowledge of the types of contact that are likely to occur within the population.

The empirical (RFID collected data) facility-level contact patterns did not conform to the assumption of homogenous mixing for the equine facilities examined (Chapters 3 and 4). These results support the importance of utilizing empirical network data given the perceived impacts (i.e. the over- or-underestimation of the epidemiological outcomes) of homogenous mixing patterns on disease dynamics. Interestingly, when the empirical networks were used to inform the contact rate of a disease transmission model (Chapter 5), the resulting epidemic curves were similar for all of the four empirical networks. Furthermore, the similarity in trends with respect to the disease prevalence curves extended to a random network with similar characteristics (average degree and network size) to the empirical networks. When the epidemiological outcomes obtained with a minimal network (i.e. a network in which each horse could contact an average of 4 other horses) were compared to the outcomes from the empirical networks, the minimal network resulted in a smaller epidemic peak and a longer epidemic duration than the networks informed by the empirical data. Lastly, a comparison of the empirical networks to a complete network resulted in a higher epidemic peak, and a shorter epidemic duration for the complete network. Interestingly, the structure gained from the empirical networks resulted in a bimodal incidence curve compared to the unimodal incidence curve that resulted from the theoretical networks. The outcomes from this study (Chapter 4) suggested that networks with comparable characteristics will result in similar epidemic outcomes. Therefore, it may be possible to generalize the effect of

disease prevention and control strategies for facilities with similar characteristics and management styles. Lastly, due to the challenges associated with the collection of empirical contact data, theoretical networks are often used as a basis for network epidemic models. While the theoretical networks have known characteristics that aid in the selection of appropriate networks, the difference in the structure of the incidence curves generated with the empirical networks compared to the theoretical networks reinforced the importance of using empirical contact data for populations which are not likely to exhibit random mixing.

6.1.4 The effect and feasibility of different disease control intervention strategies

Disease transmission model outcomes can be used to make improved decisions regarding disease prevention and control recommendations (5). The usefulness of the model outcomes, however, rely on the feasibility of the proposed intervention strategies, the characteristics of the population, and financial constraints. Hence, to examine the epidemic potential of an equine infectious agent, equine influenza was modelled within different contact networks. Interventions such as the proportion of the horse population vaccinated and the proportion of the infectious population isolated were assessed (Chapter 5). In the absence of an isolation program, an increase in the proportion vaccinated resulted in a decreased cumulative incidence. The decrease in the cumulative incidence, however, was gradual and required at least 50-75% of the population to be vaccinated. In contrast, the implementation of an isolation-only program resulted in a distinct decrease in the cumulative incidence as the proportion of infectious horses isolated was increased. The cumulative incidence was minimized with

a combined vaccination and isolation program, specifically, when a 100% isolation rate and 25-50% vaccination rate had been achieved. Overall, the implementation of any intervention strategy resulted in a decreased cumulative incidence compared to the base case scenario in which no interventions were used.

With respect to intervention feasibility, the participating facilities reported a 46-100% vaccination rate for equine influenza (Chapter 5). Furthermore, the majority of the facilities did not have a designated isolation barn, but rather an isolation stall in the main barn which is unlikely to offer the same protection as a designated isolation barn. Additionally, the presence of a single isolation stall as opposed to a designated barn limits the number of horses that can be isolated at the same time. It is important to consider the availability of such resources when developing a biosecurity plan, especially considering the non-linear interaction identified between the vaccination and isolation interventions. Although some intervention strategies will produce more optimal results than other strategies, the implementation of any intervention strategy will aid in reducing the cumulative incidence in the population. Therefore, an appropriate biosecurity strategy should be selected based on the population characteristics given the range of possible outcomes.

6.2 Limitations

Although no data or data collection technique is free from error or bias, it is important to describe the limitations of this study. First, the participating facilities were selected using a convenience sampling technique. While this sampling technique enabled the selection of facilities with favorable characteristics such as the willingness to participate in the

study, the sampling technique resulted in a non-random selection of facilities. The non-random nature of the participating facilities has implications in the ability to generalize the study results due to the limited variability in the facility characteristics such as the facility type and size. While the inclusion of similar facilities enabled a direct comparison between the study outcomes, it limited the ability to generalize the results to other types of facilities within Ontario (Chapters 4 and 5). Since contact patterns are likely to differ based on the facility type and size, the collection of contact data from facilities with a range of demographics will aid in the ability to better generalize equine contact pattern characteristics within Ontario and the associated epidemiological outcomes.

Equine contact patterns are likely to change throughout the year. For example, the competition season in Ontario occurs predominantly during the summer months, during which time horses are more likely to leave the facility to compete. In contrast, horses are likely to remain within the facility during the winter months as the competition options are limited in the winter. The data collected for this thesis did not span multiple seasons, and therefore, our data may not permit for year-round inference.

The network sizes used also presented a limitation to the thesis. The facilities used in the comparison of the contact networks and network models boarded 20-28 horses (Chapters 4 and 5). This small network size, made it difficult to carry out certain statistical analyses such as exponential random graph models. However, the small population sizes are characteristic of many equine facilities, with an estimated average of 7.89 horses per equine facility in Ontario (6), and methods should be developed to analyze such small networks.

The use of the RFID technology in the data collection process has inherent limitations ranging from the functionality of the technology to participant compliance. The limitations impacted the study in the following ways:

- With respect to the tag performance, the tags exhibited a short battery life of 7 days (Chapters 2-4). Although this might have been exacerbated by the colder temperatures (Chapters 2 and 3), the problem continued in subsequent studies that took place during the summer (Chapter 4).
- The nature and functionality of the radio-wave signals meant that the tags could not read through a body mass (7). This limitation affected the sensing ability of the tags that were on horses that were not facing each other. Although this is a barrier to collecting accurate contact data, we felt this was a reasonable limitation for equine contact structures and the nature of equine respiratory diseases. Equine respiratory diseases such as equine influenza are often transmitted through nose-to-nose contact and aerosolized respiratory droplets (8) and therefore, the limitation of a recorded contact between horses that did not have direct contact was not considered a marked limitation for the work presented.
- Data collection only occurred between individuals (i.e. horse to horse, horse to people, people to people) wearing the tags (9). This limitation has the potential to cause information bias since human participants (trainers and owners) reported forgetting to keep the halter with the attached tag with the assigned horse at all times during the study. This bias could have

resulted in the collection of data representing contacts that did not actually occur (i.e. a trainer carrying halter around barn) and a failure to record contacts that did occur (i.e. horses that came in contact in the riding areas). Similarly, the placement of the tags on the pasture fence while the horses were turned-out resulted in the potential to over-represent the contact frequency that occurred during this period (Chapters 2-4). However, since horses are known to be sociable animals (4), they are likely to come in contact while in the pasture. Additionally, the overall purpose of the data collection portion of this thesis was to collect information on the contacts that could enable disease transmission between the horses. Therefore, it was not important to collect detailed contact data, but rather the general contacts that occurred within each day.

The applicability of disease transmission models often stems from the suitability of the parameters used in the model (10). However, there is a limited amount of recorded incidence data for equine influenza in Canada. This limitation largely affected the capacity for model validation (Chapter 5). Furthermore, the scarcity of detailed data regarding vaccine efficacy may have affected the model predictions. While this uncertainty can often be addressed using sensitivity and uncertainty analyses (11), the implementation of such analyses in the current framework was computationally demanding. Although there is reported variation in parameters such as the vaccine

efficacy, as well as biological parameters such as the recovery rate (12), the model was unable to quantify the effect of the parameter ranges on the model outcome.

The model complexity also impacted the ability to calculate measures describing the disease influence such as the basic reproduction number (R_0). The calculation of R_0 can be accomplished with both mathematical and epidemiological methods (13). Traditional mathematical methods consider the mechanisms of disease transmission in the calculation of R_0 through methods such as the next generation matrix and the Jacobian matrix (13). In contrast, epidemiological methods often use incidence data to calculate the R_0 with methods such as the final size equation (13). However, the implementation of these methods such that they consider both the underlying contact network structure and the disease dynamics is complicated. Such a calculation has been documented for a susceptible-infectious (SI) model (14), however, given the complexity of the process for a simple model it is unrealistic to implement a similar analysis for the SEIR model used in this thesis.

Despite these limitations, the methodologies presented in this work provide a starting point for the characterization and implementation of contact networks to inform disease prevention and control strategies for equine populations. In comparison to other data collection methods, the RFID tags enabled the collection of large data sets that contained a high level of detail regarding the contact patterns of the entire participant pool. The collected data validated the hypothesis that equine contact patterns are not likely to be homogenous. Additionally, the data analysis enabled an in-depth

characterization and comparison of similar types of equine facilities. The consistency in the model results for each of the participating facilities suggests an ability to generalize the results to inform biosecurity strategies for similar types of facilities. To conclude, the modified RFID tags and the ensuing analysis contributed to the current knowledge of equine contact patterns and the disease dynamics within southwestern Ontario, Canada.

6.3 Future work

The work in this thesis provides a starting point for the characterization of equine contact patterns in Ontario, and the related disease spread dynamics. As such, it paves the way for future work in this domain in the following areas:

- **The need for the collection of Canadian equine disease outbreak data:** while equine disease surveillance efforts have been initiated (15), there is currently a deficit of recorded epidemic and endemic disease incidence data for Canadian equine populations. Disease incidence data is useful for many aspects of the modelling process including model parameterization through data fitting processes, and model validation (16). Therefore, the ongoing collection of Canadian specific incidence data will enable the formation and analysis of more useful and relevant models.
- **Knowledge about the efficacy of intervention strategies:** the continual evolution of the equine influenza virus due to antigenic drift implies a need for continuing surveillance efforts (17). The continual research into the vaccine

efficacy is important for both clinical and research purposes, as an up-to-date estimate of the vaccine efficacy can be used to accurately simulate the epidemiology of equine influenza given various intervention strategies.

- **The application of similar methods for other infectious equine diseases:**

disease transmission models can contribute to policy development (5). This is largely due to a model's ability to aid in the formulation of informed predictions in an ethical, and cost and time efficient manner. While disease transmission models have been used for many human and animal populations, few models have been used to study equine infectious diseases. This is likely due to limited recorded incidence data, vaccine efficacy data, and natural history parameters for equine diseases. Therefore, future work should concentrate on the collection and application of this type of data to other equine infectious diseases.

- **Applying network models to different types of facilities and environments:**

this thesis provided a starting point for the collection and analysis of empirical contact pattern data within equine facilities in Ontario, Canada. Future work should include the characterization of equine facilities varying in size and type, e.g. racing facilities and facilities that board a larger number of horses (e.g. > 50). The collection of this type of data would enable the formation of a representative database from which it would be possible to further the characterization and study of equine facilities. The analysis of a range of facility types would further enable the application of the results to create an effective biosecurity strategy for a broader range of facility types.

- **Combining the facility level models to form a metapopulation model:** this thesis collected equine contact patterns within equine facilities. While this is an important aspect of the disease transmission process, horses often leave the facility, increasing the potential for disease transmission and acquisition (18). Previous studies have focused on conducting a network analysis of equine movements between facilities (18). However, there are currently no studies that use empirical data to describe both the within-facility dynamics coupled with the between-facility dynamics. Therefore, future work should include a metapopulation model that describes the disease dynamics when both within-facility transmission and between-facility transmission are considered.
- **Advancing analytical methods for network models:** the network modelling framework used in this thesis provided the ability to generate increasingly realistic models. This increased realism stemmed from the ability to simulate both the underlying network, as well as the disease dynamics that occurred within the population. This complexity however, created a challenging environment for the computation of known mathematical constructs, such as the basic reproduction number, as well as standard sensitivity and specificity methods such as Latin Hypercube Sampling (LHS) and the computation of Partial Rank Correlation Coefficients (PRCC). Therefore, future work should concentrate on developing computationally accessible methods for these types of analyses.

6.4 Concluding remarks

This thesis provides a holistic start-to-finish approach for the analysis of the epidemic potential in equine populations using equine influenza as a case study. The consistency in the results for each participating facility implies a similarity in the underlying contact structure of similar facility types. These similarities lay the groundwork for the ability to generalize the effect of different biosecurity measures and the associated epidemic outcomes which can be used to inform both policy makers and individual facility owners. The methods and results discussed within this thesis provide the means to produce more realistic disease transmission models that can be used to aid in equine biosecurity recommendations and emergency preparedness.

6.5 References

1. Ruiz-Garcia L, Lundadei L. The role of RFID in agriculture-applications, limitations and challenges. *Comput Electron Agric.* 2011;79(1):42–50.
2. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound Emerg Dis.* 2009;56(4):109–20.
3. VanderWaal KL, Atwill ER, Isbell LA, McCowan B. Quantifying microbe transmission networks for wild and domestic ungulates in Kenya. *Biol Conserv.* 2014;169:136–46.
4. Lloyd AS, Martin JE, Bornett-Gauci HLI, Wilkinson RG. Horse personality: Variation between breeds. *Appl Anim Behav Sci.* 2008;112(3–4):369–83.
5. Garnett GP, Cousens S, Hallett TB, Steketee R, Walker N. Mathematical models in the evaluation of health programmes. *Lancet.* 2011;378(9790):515–25.
6. Equine Canada. Horses in Canada in 2010. 2011;(23):9–32.
7. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, van den Broeck W, et al. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS One.* 2011;6(2):e17144.
8. Myers C, Wilson WD. Equine influenza virus. *Clin Tech Equine Pract.* 2006;5(3):187–96.
9. Read JM, Edmunds WJ, Riley S, Lessler J, Cummings DAT. Close encounters of the infectious kind: Methods to measure social mixing behaviour. *Epidemiol Infect.* 2012;140(12):2117–30.
10. Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. *PLoS Med.* 2005;2(7):e174.
11. Wu J, Dhingra R, Gambhir M, Remais J V. Sensitivity analysis of infectious disease models: methods, advances and their application. *J R Soc Interface.* 2013;10(86):20121018–20121018.
12. Baguelin M, Newton JR, Demiris N, Daly JM, Mumford JA, Wood JLN. Control of equine influenza: scenario testing using a realistic metapopulation model of spread. *J R Soc Interface.* 2010;7(42):67–79.
13. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface.* 2005;2(4):281–93.

14. Leung KY, Kretzschmar M, Diekmann O. SI infection on a dynamic partnership network: characterization of R_0 . *J Math Biol.* 2015;71(1):1–56.
15. Ontario Animal Health Network. Staying ahead of the curve magazine: the disease surveillance plan 2013-2018 [Internet]. 2018 [cited 2018 Jun 6]. Available from: <http://oahn.ca>
16. Hyder A, Buckeridge DL, Leung B. Predictive validation of an influenza spread model. *PLoS One.* 2013;8(6):e65459.
17. Paillot R, Prowse L, Montesso F, Huang CM, Barnes H, Escala J. Whole inactivated equine influenza vaccine: Efficacy against a representative clade 2 equine influenza virus, IFN γ synthesis and duration of humoral immunity. *Vet Microbiol.* 2013;162(2–4):396–407.
18. Spence K, O’Sullivan T, Poljak Z, Greer AL. A longitudinal study describing horse characteristics and movements during a competition season in Ontario, Canada in 2015. *Can Vet J.* 2018;In Press.

APPENDICES

Appendix 1: Supplementary materials for Chapter 3

Table S3.1: Classification table output to assess the effectiveness of the survey collected data for predicting contacts. The networks formed with the tag-collected data represents the gold standard and the network formed with information from the survey represents the test. The test sensitivity always correctly predicted the existence of contacts, however, the specificity was relatively low, indicating an inability to correctly predict an absence of contacts. Both the positive and negative predictive values were relatively high (≥ 0.5).

	Day						
	1	2	3	4	5	6	7
Sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Specificity	0.33	0.24	0.24	0.24	0.22	0.23	0.20
Positive predictive value	0.75	0.61	0.61	0.61	0.56	0.58	0.50
Negative predictive value	1.00	1.00	1.00	1.00	1.00	1.00	1.00

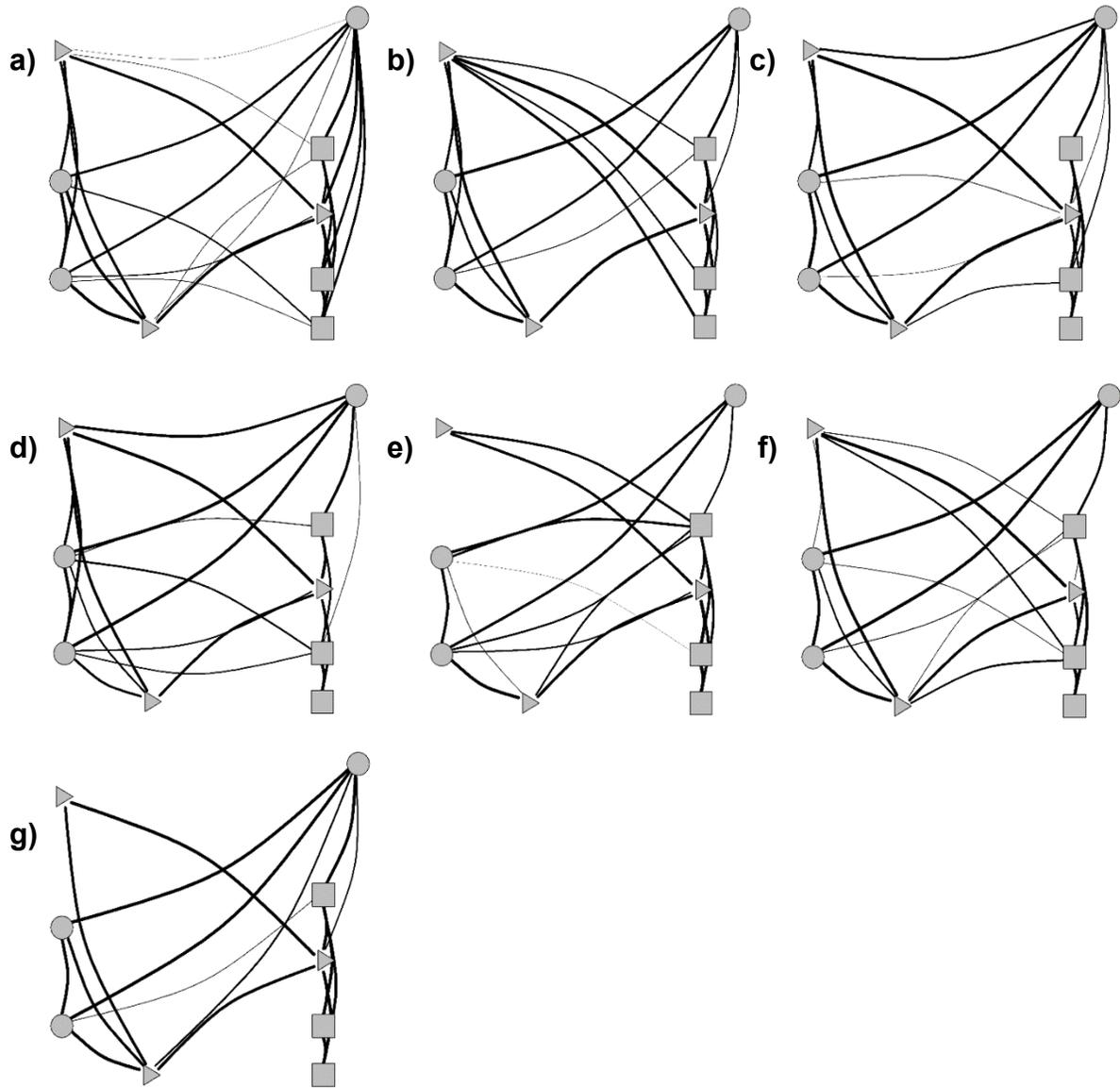


Figure S3.1: Complete set of network graphs for each day of the study period (panels a-g represent days 1-7 respectively). Nodes are laid out in a similar configuration to the horses' stall locations. The node shape is coded according to the horses' pasture locations.

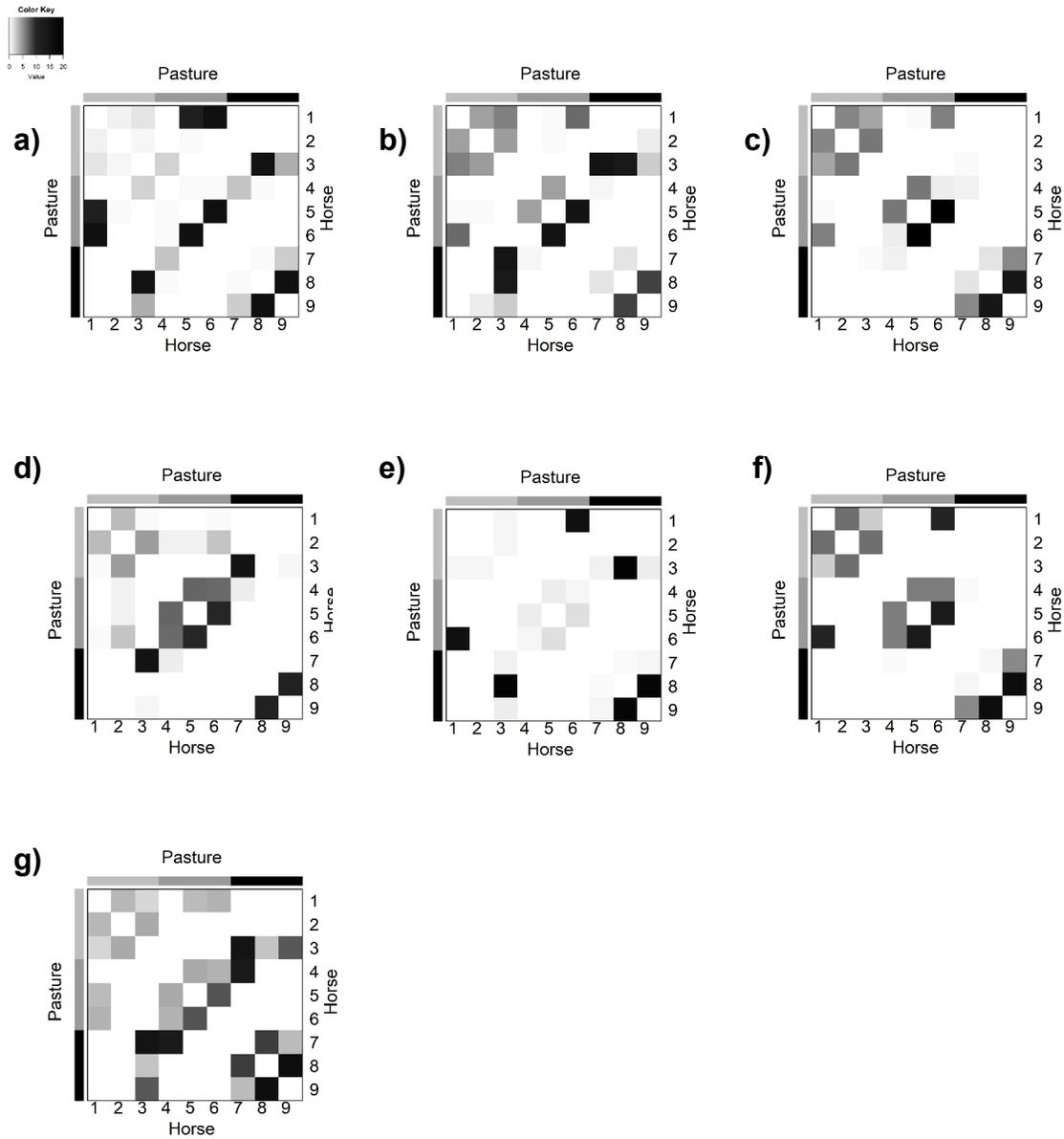


Figure S3.2: Complete set of heat maps for each day of the study period (panels a-g show the heat maps from days 1-7 respectively). Darker cells represent longer durations (hours) of contact between horses.

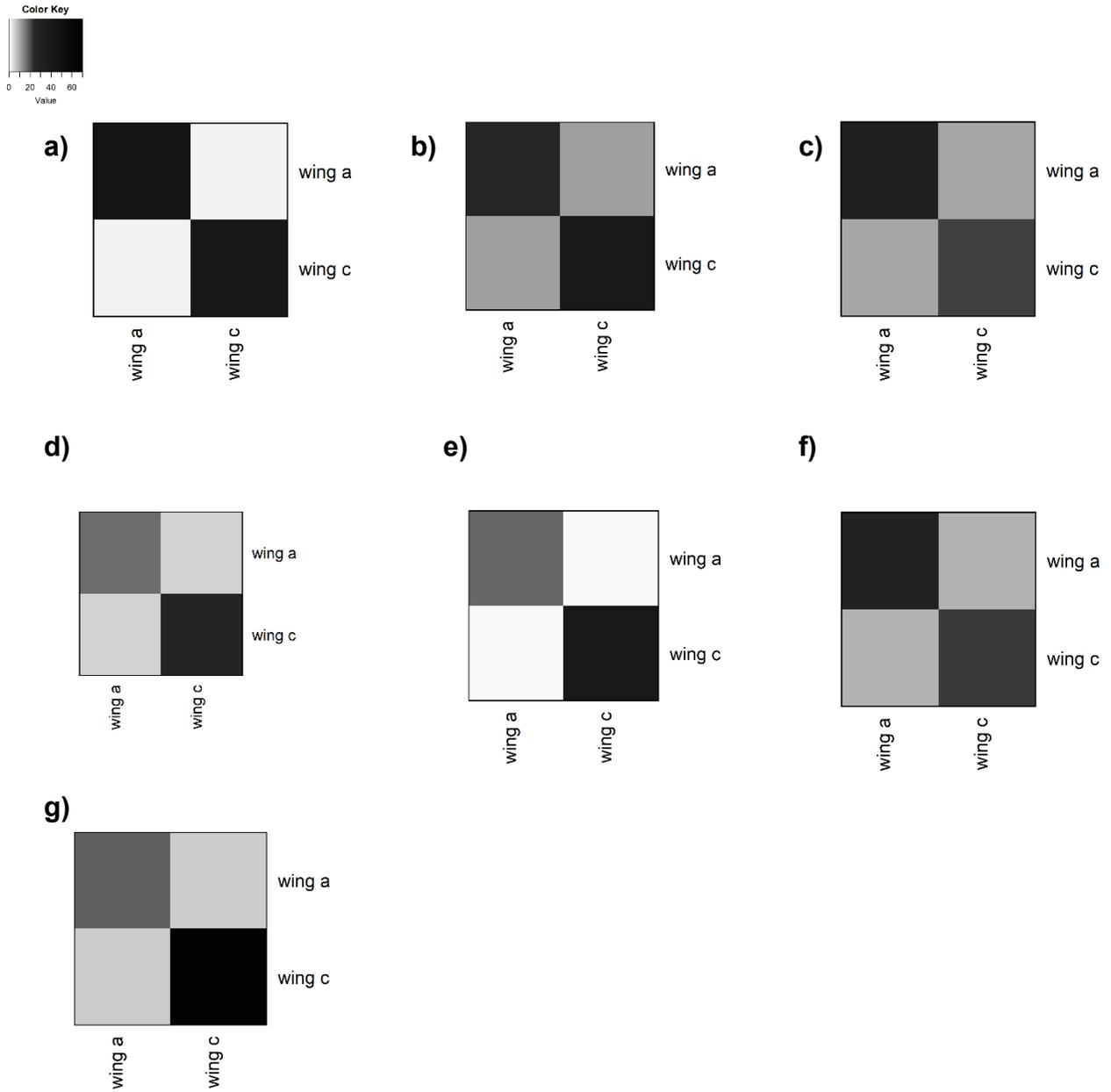


Figure S3.3: Complete set of heat maps aggregated by aisle for each day of the study period (panels a-g show the heat maps from days 1-7 respectively). Darker colors represent longer durations (hours) of contact between horses in the respective wings.

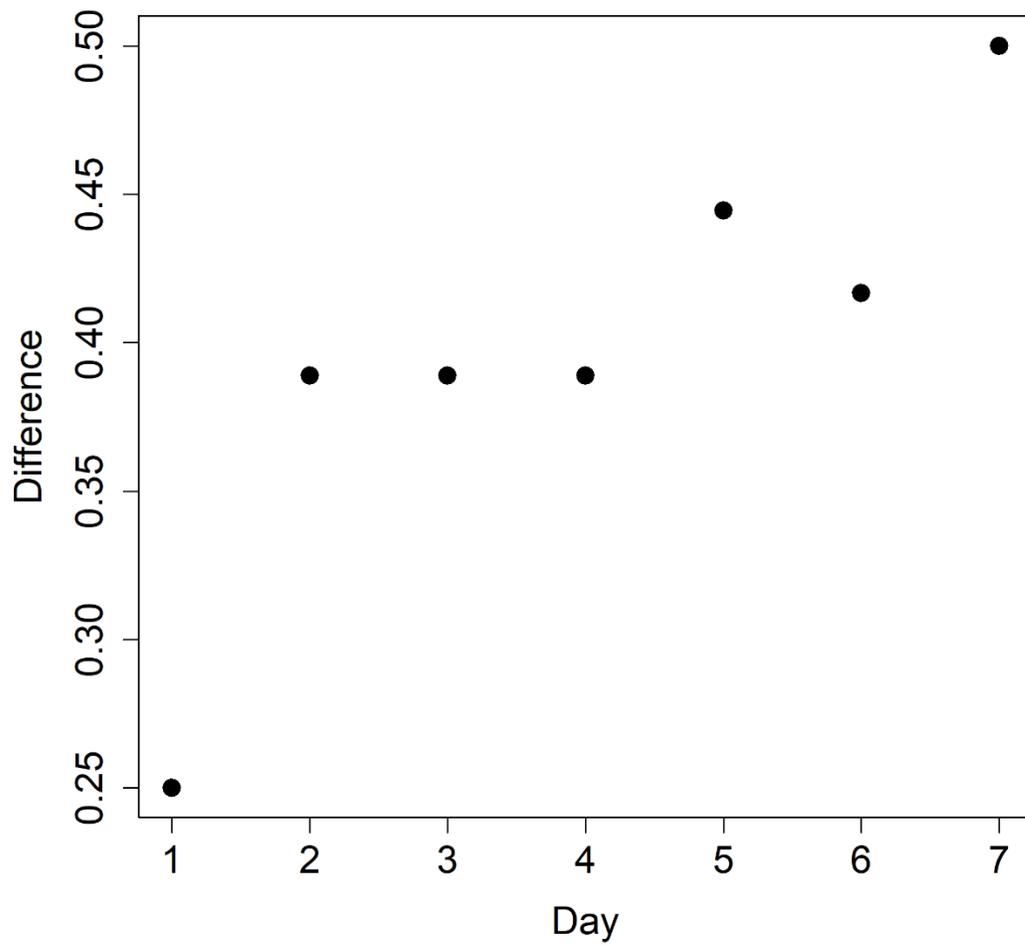


Figure S3.4: Plot of Jaccard distance index between the survey acquired network and the tag acquired networks. The static survey acquired network was compared to each day of the tag acquired network. All study networks shared at least 50% of contacts with the survey-based network.

Appendix 2: Supplementary materials for Chapter 4

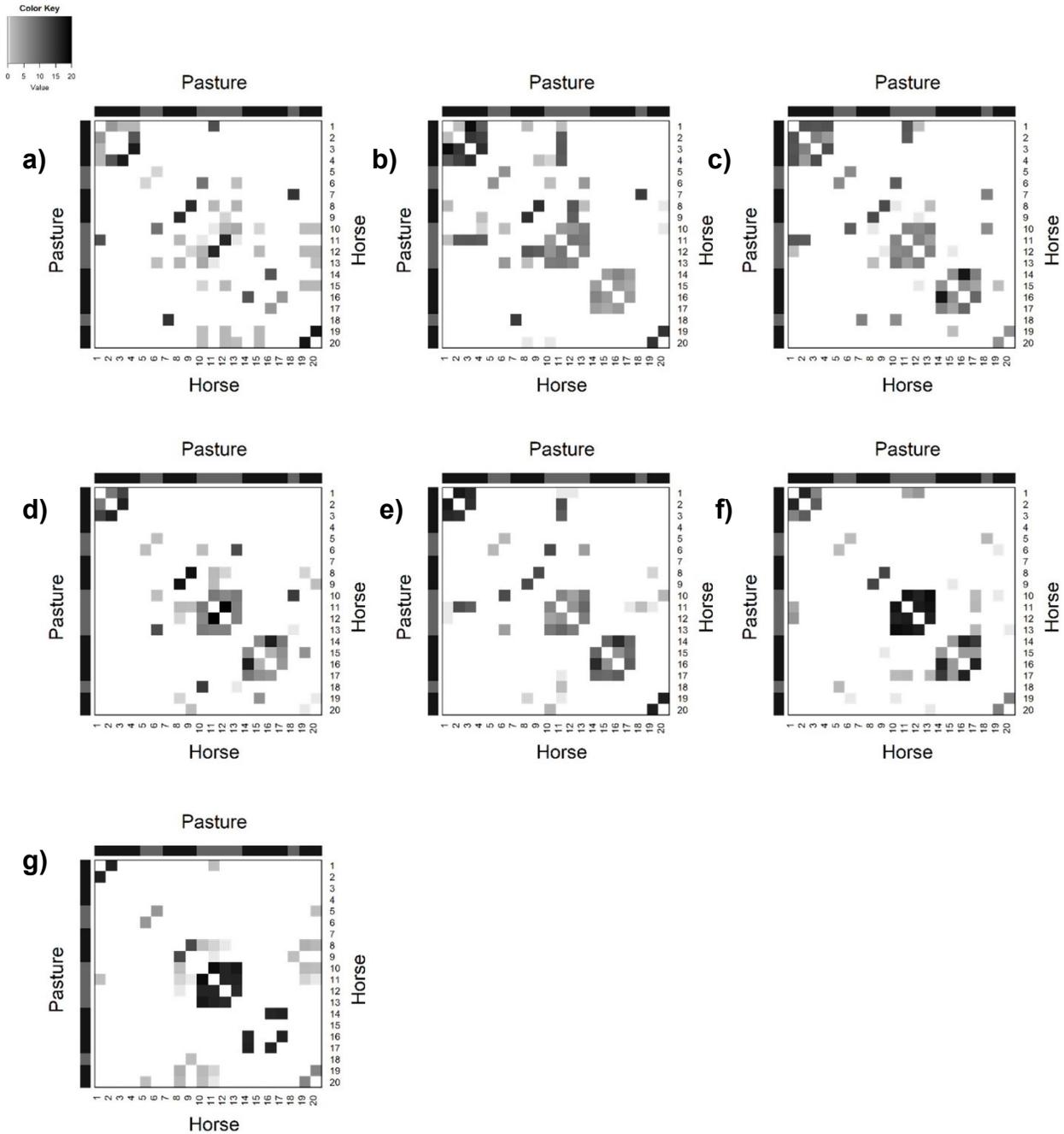


Figure S4.1: A complete set of heat maps showing the contact duration (hours per day) for each pair of horses on facility 1. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. The panels progress from day 1 (panel a) to day 7 (panel g).

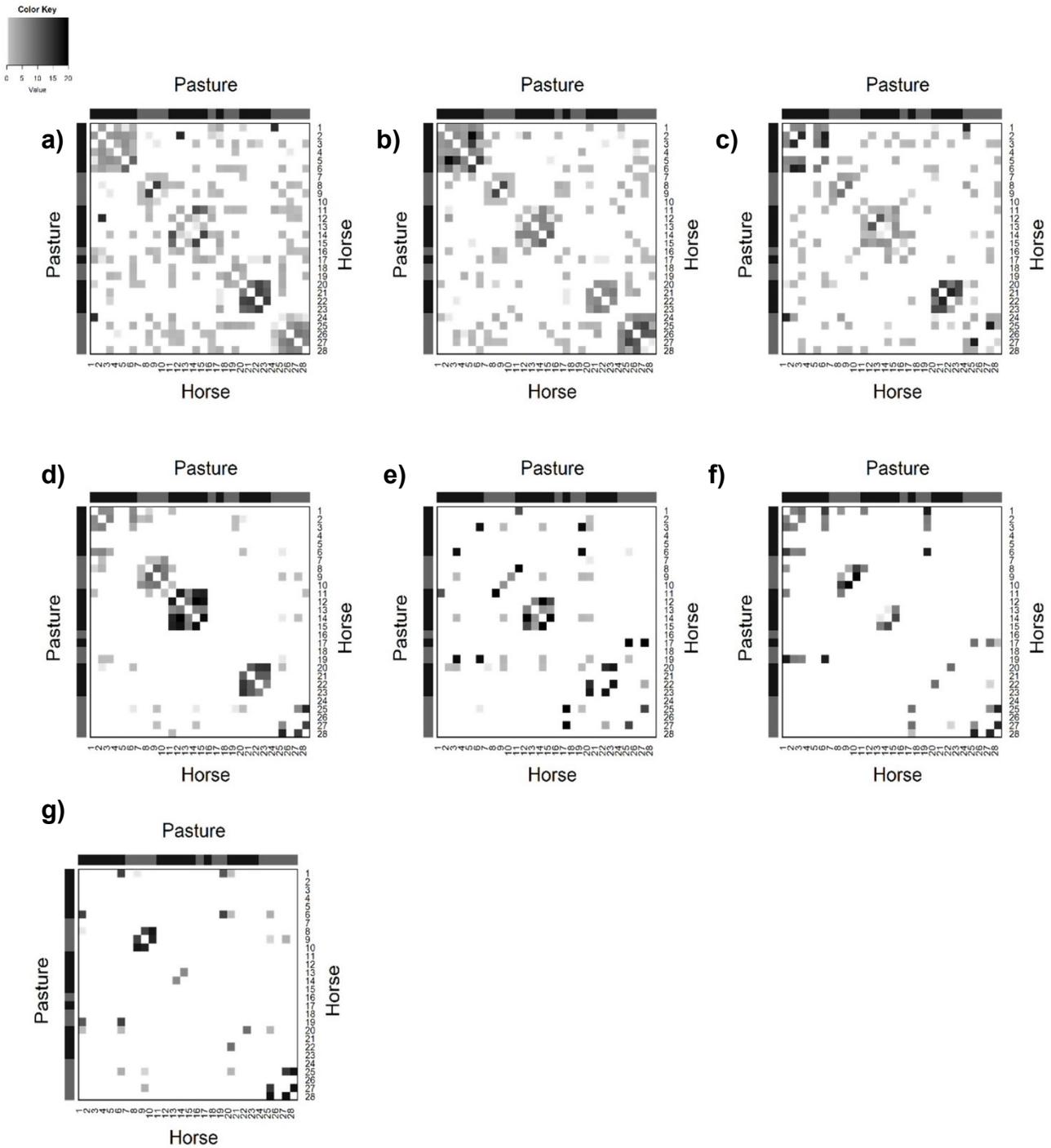


Figure S4.2: A complete set of heat maps showing the contact duration (hours per day) for each pair of horses on facility 2. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. The panels progress from day 1 (panel a) to day 7 (panel g).

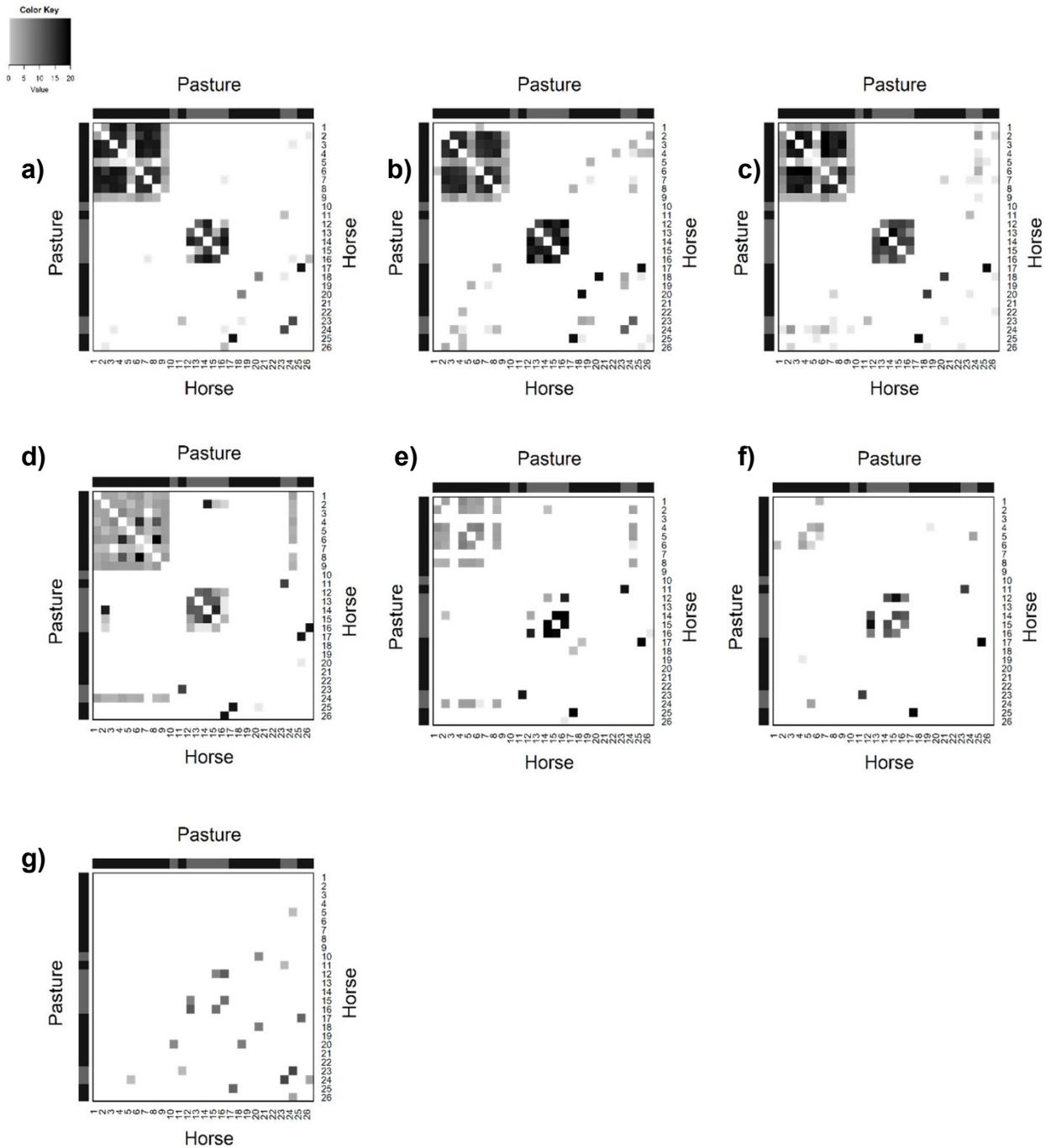


Figure S4.3: A complete set of heat maps showing the contact duration (hours per day) for each pair of horses on facility 3. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. The panels progress from day 1 (panel a) to day 7 (panel g).

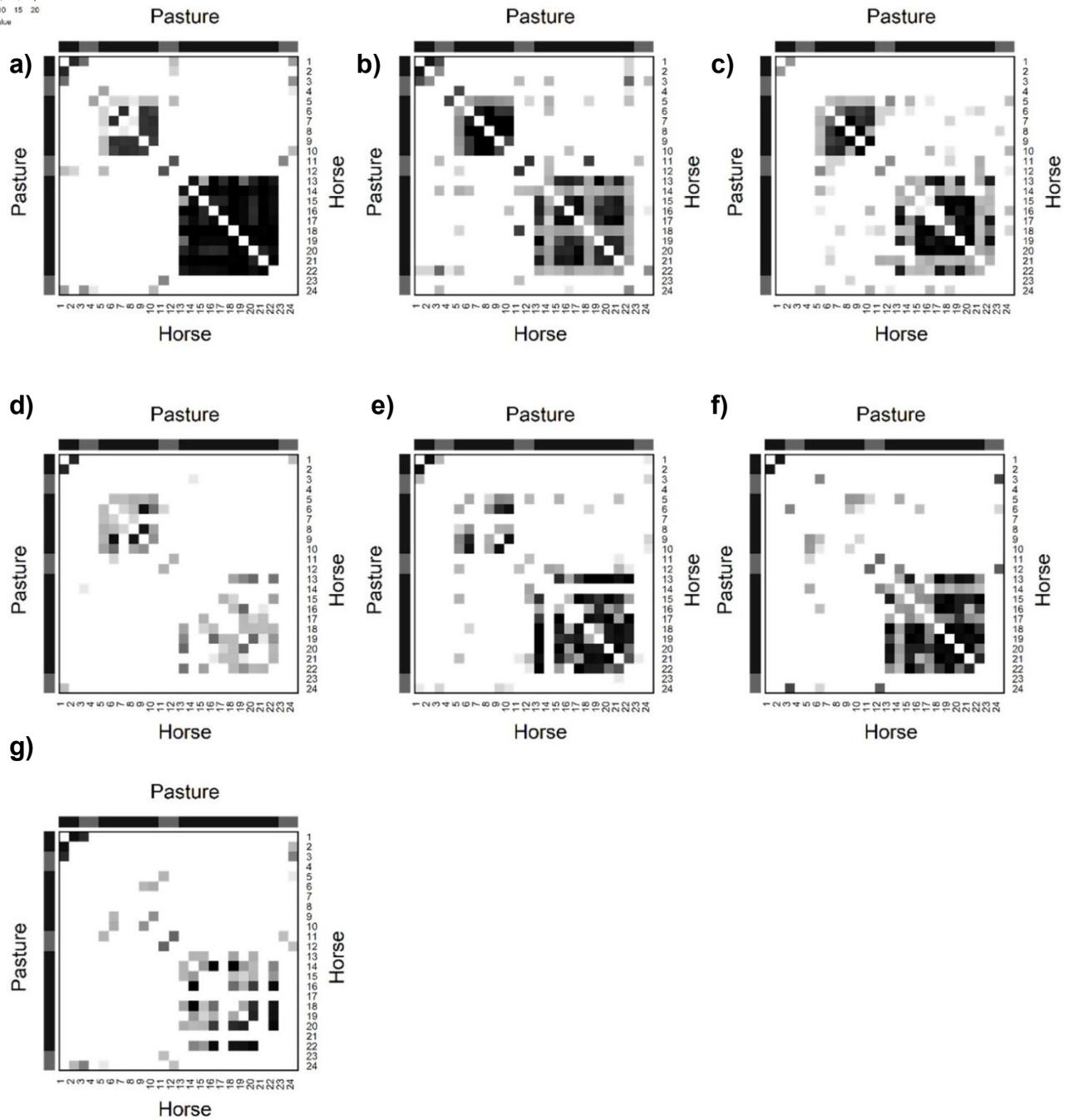


Figure S4.4: A complete set of heat maps showing the contact duration (hours per day) for each pair of horses on facility 4. Horses are grouped by pasture. Each section on the pasture axes, indicated with alternating colours, represents a group of horses that shared a pasture. The panels progress from day 1 (panel a) to day 7 (panel g).

Appendix 3: Supplementary materials for Chapter 5

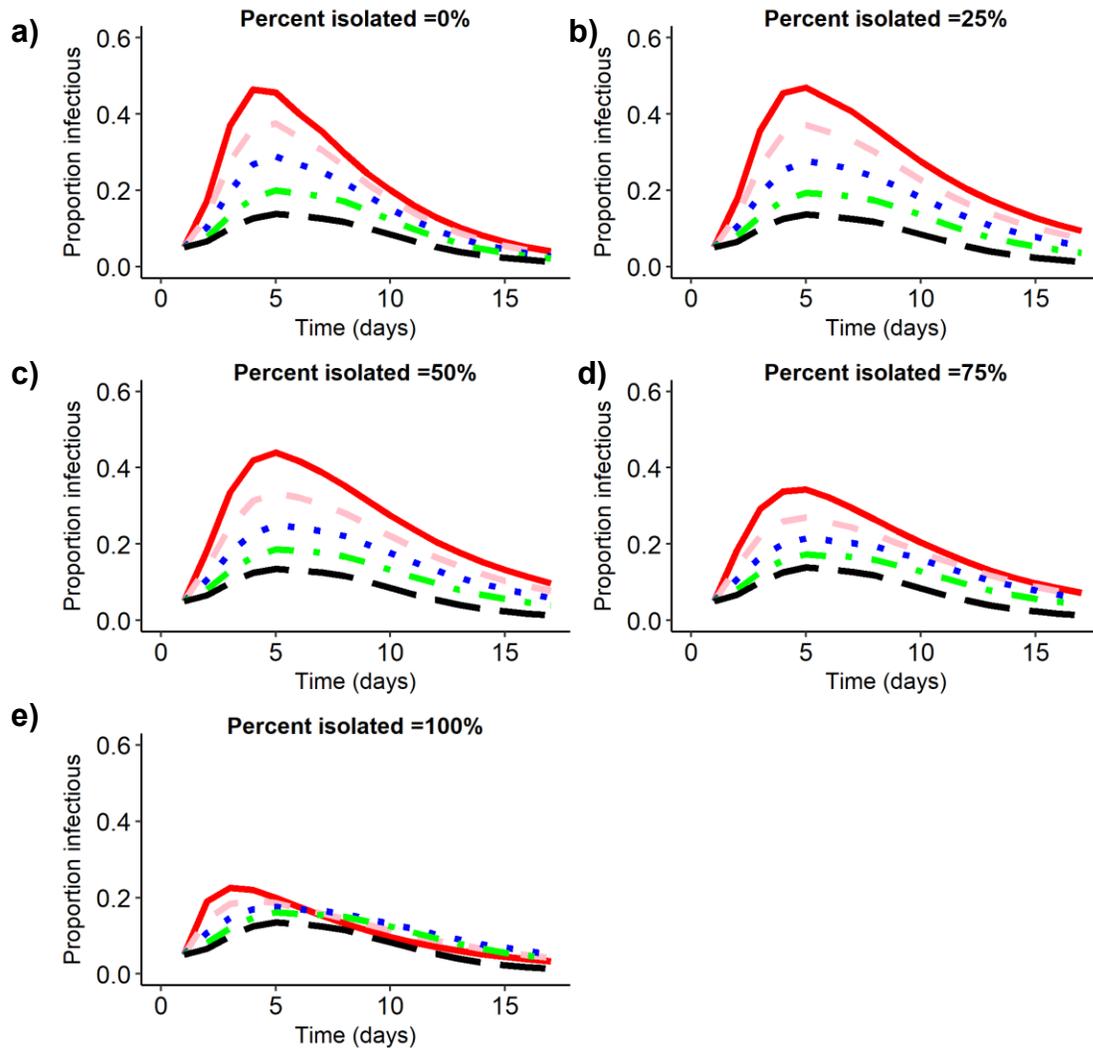


Figure S5.1: Prevalence plots for Empirical Network 1. Empirical Networks 2-4, the RMN, and the Complete and Minimal Networks resulted in similar trends. The plots show the effect of varying the proportion of the population that is vaccinated at the start of the model as the proportion of the infectious population that is isolated is kept constant. The proportion vaccinated is portrayed with different line colours: red solid line=0%, pink dotted line=25%, blue dotted line=50%, green dotted line=75%, and black solid line=100%. Panels a)-e) represent a population in which 0, 25, 50, 75, and 100% of horses are isolated. For vaccination rates of 0-75%, increasing the isolation rate caused a decrease in the disease prevalence. However, the difference in prevalence due to the increasing isolation rates decreased as the proportion vaccinated increased. By a 100% isolation rate, there was minimal difference in the prevalence with increasing isolation rates.

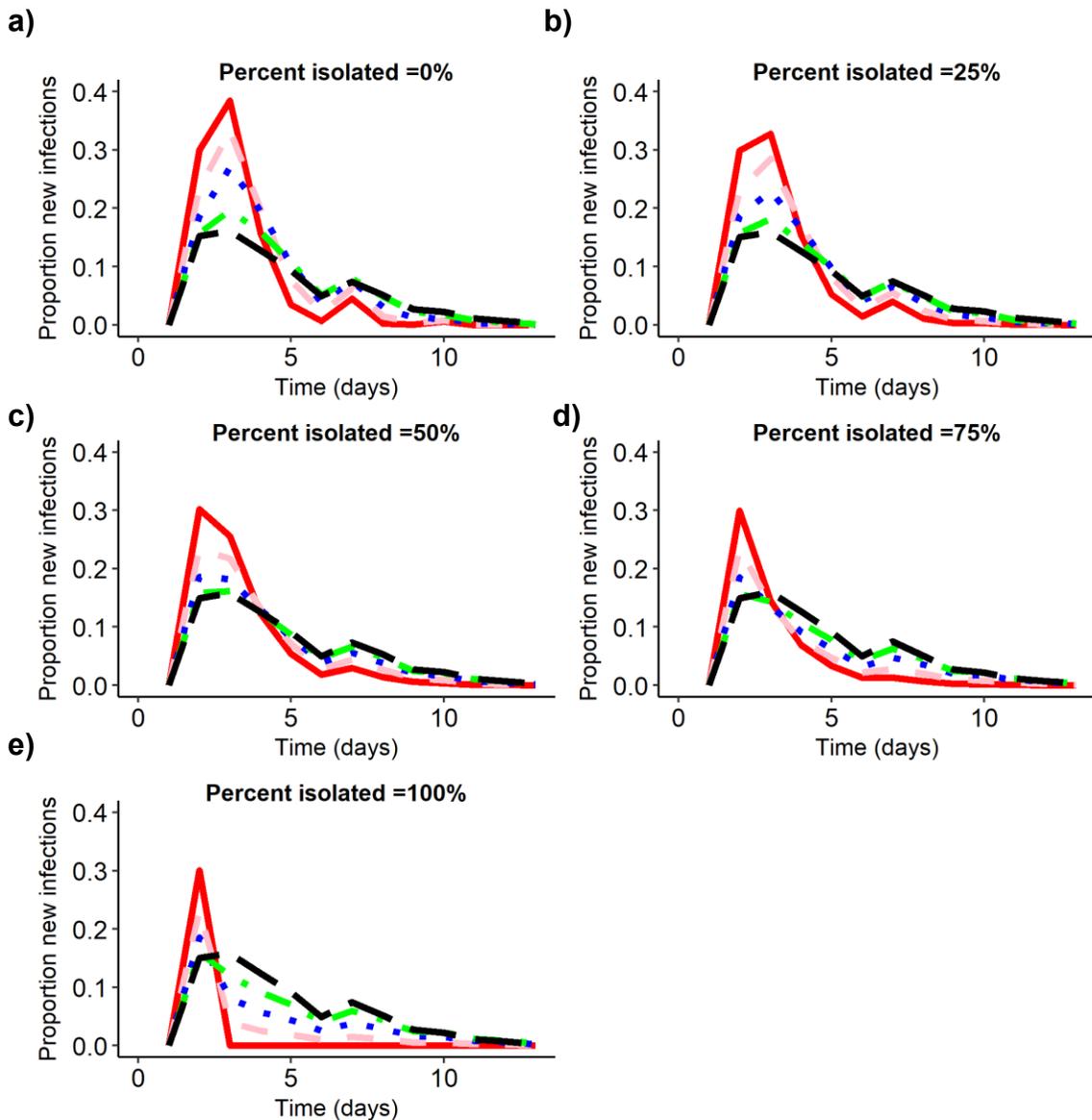


Figure S5.2: Incidence plots for Empirical Network 1. Empirical Networks 2-4, the RMN, and the Complete and Minimal Networks resulted in similar trends. The plots show the effect of varying the proportion of the infectious population that is vaccinated as the proportion of the population that is isolated is kept constant. The proportion vaccinated is portrayed with different line colours: red solid line=0%, pink dotted line=25%, blue dotted line= 50%, green dotted line=75%, and black solid line=100%. Panels a)-e) represent a population in which 0, 25,50, 75, and 100% of horses are isolated. A difference in incidence due to the proportion isolated was observed after the first peak in infections. By 75% vaccination coverage, there was minimal difference in in the number of new infections with an increase in the isolation rate, and by 100% vaccination, there was no difference in the disease incidence with an increase in the isolation rate.

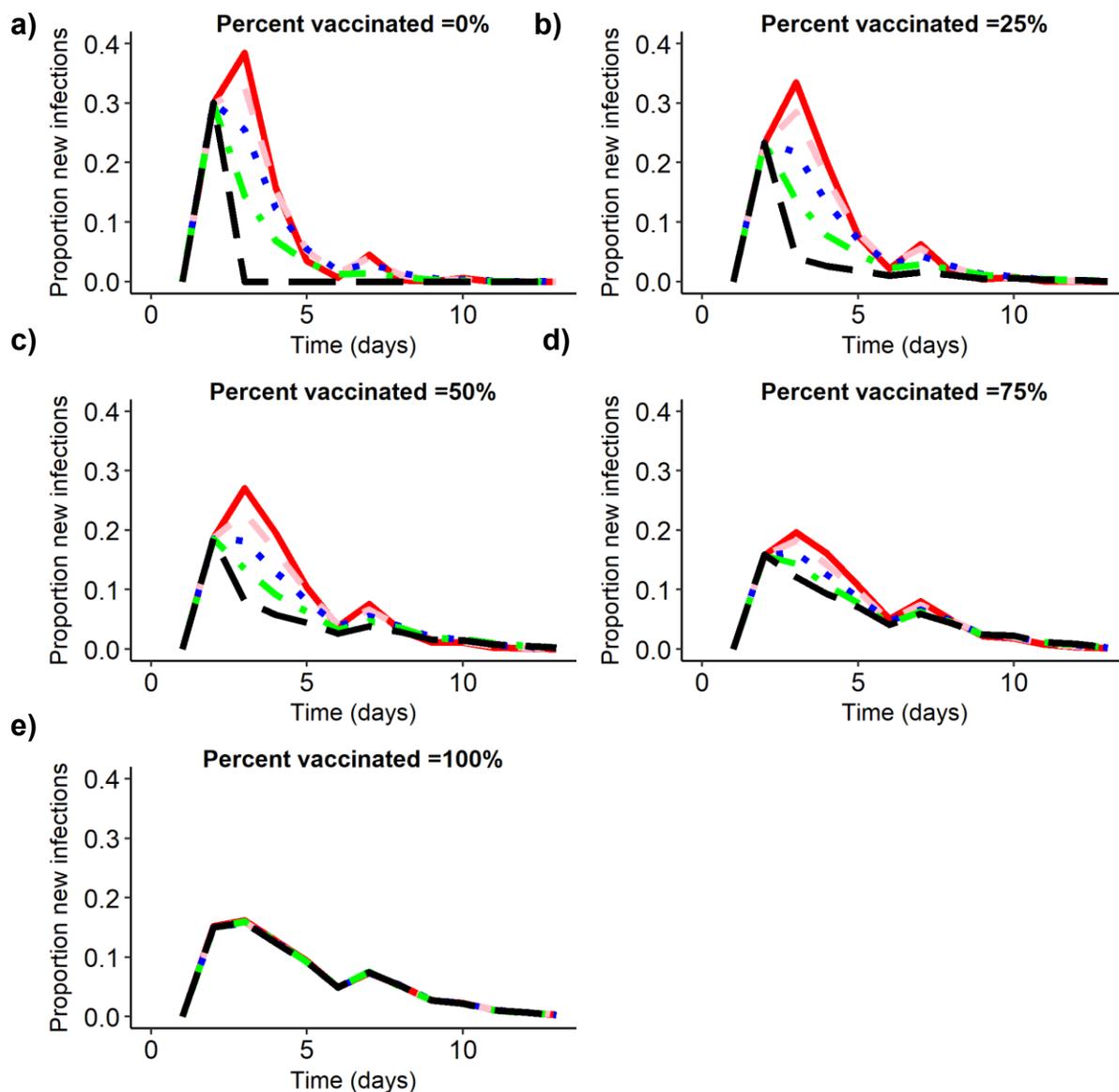


Figure S5.3: Incidence plots for Empirical Network 1. Empirical Networks 2-4, the RMN, and the Complete and Minimal Networks resulted in similar trends. The plots show the effect of varying the proportion of the infectious population that is isolated as the proportion of the population that is vaccinated is kept constant. The proportion isolated is portrayed with different line colours: red solid line=0%, pink dotted line=25%, blue dotted line= 50%, green dotted line=75%, and black solid line=100%. Panels a)-e) represent a population in which 0, 25,50, 75, and 100% of horses are vaccinated. The plots were bimodal, except for when high isolation rates (100%) were implemented with low vaccination rates (0%). An increase in the vaccination rate caused a decrease in the size of the first peak. With the exception of 0% vaccination rate, increasing the vaccination rate had no effect on the size of the subsequent peaks.