University of Ottawa

Interdisciplinary School of Health Sciences

Doctoral Program in Population Health

Thesis

Tuberculosis in the Qu'Appelle Agency: 1885-1926

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Student Name: Igor Zverev

Thesis Supervisor: Professor Daniel E. Lane

Telfer School of Management

University of Ottawa

November 2017

1. Abstract

Introduction: Tuberculosis (TB) is an infectious disease that causes significant morbidity and mortality. Despite the fact that the total burden of TB has decreased dramatically, the distribution of that burden across the Canadian population has not changed. A century ago, the Indigenous population of Canada had a significantly higher TB mortality than the non-Indigenous population. This gap still exists today. TB is a disease of poverty, and understanding the role of the social determinants of health (SDH) may provide insights into the causes of persistence of TB in the Indigenous population.

Research questions: This thesis tackles three questions: 1) Can a TB outbreak that took place over a century ago be reconstructed? 2) What can we learn about the relationship between the disease, the population it afflicted, and the environment in which the outbreak took place? 3) How can reconstruction of a TB outbreak be used to evaluate policy interventions?

Area studied: Analyses were limited to the Qu'Appelle Agency, located in Southeastern Saskatchewan.

Methodology: An agent-based model of socioeconomic environment of the Qu'Appelle Agency was developed to study the relationship between TB and SDH. Data on TB mortality, demographics, agricultural production, material circumstances, and economic factors of production were used to study the relationship between TB and SDH at the aggregate level.

Results: 1) Extensive aggregate data analyses were carried out and an agent-based model of TB transmission and of the socioeconomic environment of the Qu'Appelle Agency was developed. 2) Results of these analyses identify a number of important parameters responsible for the high TB mortality in the Agency. These parameters include biological factors, housing, social characteristics, agricultural output, and policies of the Department of Indian Affairs.

Conclusions: This research demonstrates that reconstruction of an outbreak of an infectious disease that took place over a century ago is a complex undertaking that hinges on availability of data and significant expertise in a variety of fields, such as health sciences, economics,

mathematics, and modelling approaches. The further one goes into the past, the more one is forced to rely on assumptions, which make the reconstructed web of relationships between agent, host, and environment that caused the outbreak less certain. Despite the inherent uncertainty, the process of outbreak reconstruction provides a deep and multi-faceted understanding of the interactions among the agent, the host, and the environment. The resulting model is a useful way of studying policy interventions that could be applied in other contexts as well – to other infectious diseases or TB outbreaks on other reserves.

Keywords: [population health, epidemiology, tuberculosis, Indigenous peoples, agent-based modelling, social determinants of health]

Glossary

Agency	A Canadian territorial administrative unit created by the Department of Indian Affairs. An Agency included several reserves
Agent-based model	A mathematical model that uses artificial computer-generated beings to study a variety of social phenomena. Typically, in these models heterogeneous agents follow decision rules as they interact in an explicitly-defined environment.
Band	A body of Indians for whose collective use and benefit lands have been set apart or money is held by the Crown, or declared to be a band for the purposes of the *Indian Act*. A band may also be known as a First Nation. Each band has its own governing band council, usually consisting of one chief and several councillors. Community members choose the chief and councillors by election. The members of a band generally share common values, traditions and practices rooted in their ancestral heritage. (Aboriginal Affairs and Northern Development Canada, 2010)
Indian Agent	An employee of the Department of Indian Affairs in charge of an Agency
Miliary TB	"Description of the clinical disease caused by the widespread hematogenous dissemination of bacteria to most organs of the body." (Long R. E., 2007, p. 100)
Prevalence	A number of existing cases of a given disease (Greenberg, Daniels, Flanders, Eley, & Boring, 2005, p. 21)
Pulmonary TB	Contagious bacterial disease of the lungs that can be spread by coughing
Reserve	A tract of land, the legal title to which is held by the Crown, set apart for the use and benefit of an Indian band. Some bands have more than one reserve. (Aboriginal Affairs and Northern Development Canada, 2010)
Scrofula	Tuberculous disease of a cervical lymph node with sinus tract formation or ulceration of the overlying skin. (Public Health Agency of Canada, 2014, p. 158)

List of abbreviations

ABM	Agent-based model
AUC	Area under the curve
CSDH	Committee on Social Determinants of Health
DIA	Department of Indian Affairs
OPA	Ordinal pattern analysis
PCA	Principal component analysis
PHAC	Public Health Agency of Canada
SDH	Social determinants of health
SES	Socio-economic status
TB	Tuberculosis
WHO	World Health Organization

Table of contents

List of Tables

List of Figures

List of Equations

1. Introduction

Imagine yourself in the Prairies in the 1870s. The world you would see around you would be very different from the Prairies as we know them today. The farms that came to define the landscape of the Prairies were not yet there. The European settlers were a minority, and the great Indigenous and Metis societies defined the Prairies.

But that world was beginning to change. The Province of Manitoba was created in 1870. The railroad linking Eastern Canada to British Columbia was slowly snaking to the west. The last of the once huge herds of buffalo that once roamed the Prairies were about to vanish, taking with them the way of life that has existed in the Prairies for centuries.

Today, when climate change has become a reality, the disappearance of a species may not be something unusual. However, the impact of the disappearance of the buffalo on the Indigenous communities of the Prairies cannot be overestimated.

> "There is one question at least which for some years past has sorely disquieted the mind of the Indian of the Saskatchewan, and which causes him to look forward with increasing, anxiety to the future. The question is this: How shall we find subsistence when the buffalo is destroyed? Until within the last two years the buffalo, which roamed over the prairie in apparently, exhaustless herds, furnished the Indians with a supply of food practically unlimited. For the last ten years the numbers of the buffalo have greatly diminished, and in another decade of years, unless prompt measures be taken in the meantime to prevent the catastrophe, the buffalo as a source of supply of food will be extinct.
>
> It must be remembered that this animal, besides furnishing the Indians with their principal means of subsistence as food, is otherwise of great importance to them. The traffic of the buffalo peltries is very considerable, and has hitherto enabled the Indian to supply his family with many of the necessaries of life. The Indian feels, therefore, that on the existence of the buffalo his own existence really depends. He cannot view without dismay the wanton and indiscriminate slaughter of these animals, mainly, be it observed by the Whites and Half-breeds, who have intruded into the domain of the red man, and who wage war upon the buffalo as an enemy instead of protecting him as a friend"(DIA., 1877, p.15).

Independent Indigenous societies that had existed in the Prairies disappeared with the signing of treaties between the Canadian government and the Aboriginal peoples, giving one side access to land and resources, and the other – reserves and annuities.

With the buffalo gone by 1879, the Indigenous communities had to undergo drastic changes to survive in the new reality. Largely semi-nomadic societies were transformed into sedentary communities in a span of a couple of years. Hunters of buffalo now had to learn how to grow potatoes and oats. The nomads had to exchange the wide expanse of the Prairies where they used to pitch their tepees for the confines of a reserve and a small wooden hut with an earthen floor and no windows or chimney. The proud nations that used to have the political authority over their land, used to forge alliances, make war and peace, now had to submit to a bureaucrat who considered their past and their culture nothing short of barbaric, and the people themselves – barbarians that had to be civilized.

The sheer magnitude of transformation that they had to go through is hard for us to imagine. The biggest recent shock many Canadians experienced may have been the 2008 economic crisis. Many people lost their jobs; some may have lost their houses. While individual people and companies no doubt suffered, institutions remained. The Canadian government was not replaced by a bureaucrat appointed from overseas who regards hockey as a barbaric game that had to be banned and Canadian children his best hope, with the adults too old to be civilized. The cities did not disappear, and while some people may have had to return to school to update their skills, the whole way of life that most of us are familiar with did not change.

And yet, the economic crisis of 2008 had a significant impact on our health(Karanikolos et al., 2013). As the Indigenous societies of the Prairies went through a complete transformation of their society, it would be reasonable to expect that the impact of these changes on their health status was very significant.

As the environment changed, so did the diseases that afflicted the Indigenous societies. Tuberculosis (TB) was not seen on the Prairies before the 1860s, with sporadic cases not contributing significantly to mortality appearing between 1865 and 1879. But it soon became a scourge of the Indigenous societies, resulting in a mortality rate of 93 deaths per 1000 population in 1886. Some 40 years later, in the late 1920s, the TB mortality rate of 8 per 1000 was still 20 times higher than that of the non-Indigenous population at that time (Ferguson, 1928, p.12).

Why did TB mortality increase sharply after the Indigenous societies settled on reserves? They were exposed to TB before, through their contacts with European settlers and traders, but there was no sign of TB on the Prairies before the 1860s.

Over one hundred years ago, a recorded TB outbreak in the Qu'Appelle Agency begun. The Qu'Appelle Agency, an administrative unit of the Department of Indian Affairs (DIA) that included several reserves, was named after the valley where it was located. The Agency was created in 1885. TB mortality peaked early on at nearly 100 per 1,000 population in 1886. By the 1920s it declined to between 8 and 16 deaths per 1,000 population. Compared with non-Indigenous population in the surrounding areas, TB mortality rate in the Qu'Appelle Agency was still approximately 20 times higher (Ferguson, 1928, p.12) (Figure 1-1).

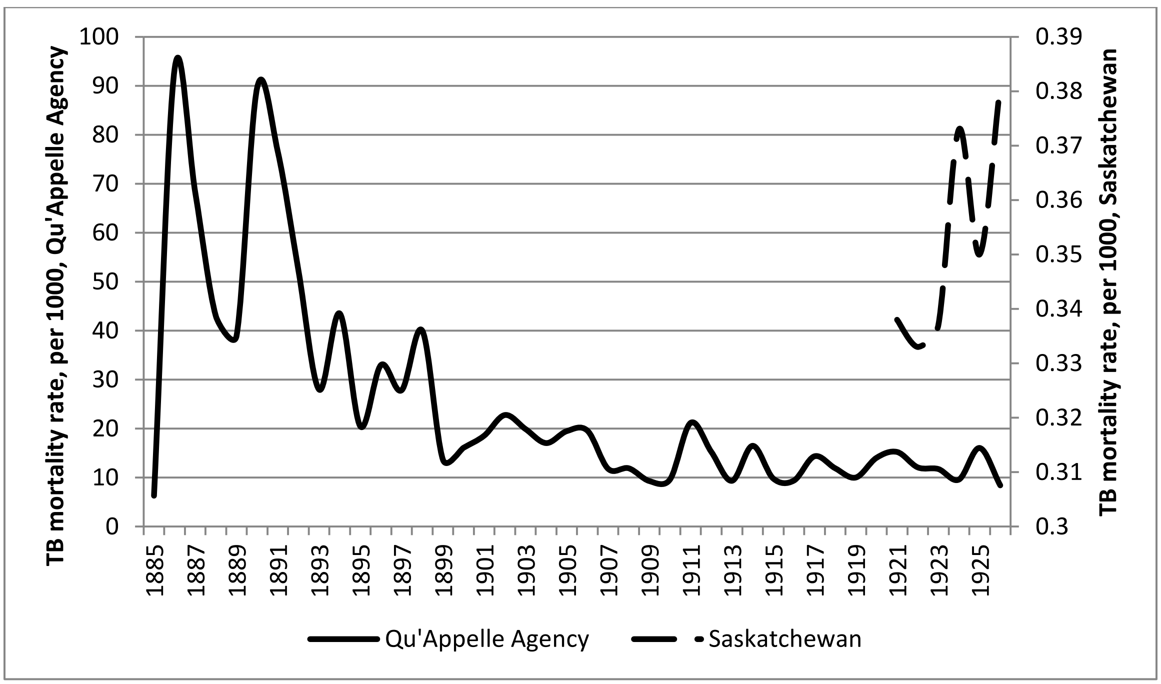

Figure 1-1: TB mortality rate per 1,000 population, Qu'Appelle valley, 1885-1926. Source: Ferguson (1928), Dominion Bureau of Statistics (1939)

In 2013, TB incidence rate in the Indigenous population was still over 33 times higher than that in the non-Indigenous population (PHAC., 2015).The reasons for the elevated incidence of TB in contemporary Indigenous and immigrant populations are cited in the literature as: (i) significant reservoir of latent disease; (ii) disproportionate rates of chronic diseases and HIV; and (iii) high housing density. As noted above, before 1860s, there were no signs of TB on the Prairies, of course there was no HIV, and housing density was low. Consequently, the real causes of the 1886 spike in mortality from TB have not been fully investigated or explained. The upstream determinants of this high mortality are still open questions.

The epidemiologic transition (Omran, 2005) that accompanied changes in social and economic conditions experienced by the Indigenous communities of the Prairies is not surprising given what we know about the relationship between socioeconomic conditions and health. Today, we are aware that health status is largely determined by socioeconomic conditions, and therefore, any change in these conditions is expected to be reflected in the population's health status.

TB has been called a "social disease with a medical aspect" and is related to every measure of poverty. As such, it is particularly important to understand the role of social determinants of health in the spread and persistence of TB.

Despite the fact that a change in health status could be anticipated, the nature of the change may be surprising. The fact that the Indigenous societies of the Prairies experienced such a sharp increase in TB mortality might suggest that their socioeconomic conditions deteriorated for TB is a disease of poor environmental and living conditions. Yet, when it comes to food security, exposure to intoxicants, external violence, and access to health care, their conditions were likely improved by the signing of the treaty in 1874 as treaty conditions included the payment of annuities, agricultural assistance to the bands, and a ban on alcohol (DIA., 1966). In addition to this, the Government presented alleviation of food insecurity as one of the benefits of signing the treaty (Taylor, 1985). Thus, the increase in mortality has not been clearly linked to the social determinants of health in this case.

The issue of the role of social determinants of health in the spread of TB raises other questions. While TB has afflicted humans for thousands of years, our understanding of the role of social determinants of health in the spread of TB is based on the last 200 years of data. Is the currently observed inverse relationship between social determinants of health and illness universal, or is it specific to certain populations? Is there some threshold level below which improvements in social determinants of health do not result in any declines in TB incidence? What is the relationship between social determinants of health and susceptibility to TB? While current discourse on the social determinants of health examines the impact of certain individual factors, such as income and housing on health, it does not discuss the complete fracturing of a whole society, like the one that took place on the Prairies in the late 19th century.

When studying health in general, and TB in particular, one must not forget the human side of the story. When we conduct epidemiological studies, we often focus on quantitative things like incidence and prevalence rates, confidence intervals, odds ratios, heteroskedasticity and multicollinearity. These are important. But it is also important to remember that behind those numbers are stories of people. People who, in the case of the Qu'Appelle Agency outbreak, saw their world change completely in a matter of years, and who witnessed their fathers and mothers, brothers, sisters, and children die a slow death from a disease they had not seen before. Understanding and incorporating the context in which the outbreak took place may improve our understanding of relationship between TB and social determinants of health as well as our understanding of the impact of TB.

Palaeontologists work with data from fossilized bones to reconstruct not only the way the dinosaurs looked, but also how they hunted, what they ate, and how they moved. Similarly, a complete reconstruction of a century-old TB outbreak requires the understanding of complex relationships between the disease, the population it affected, and the environment in which the outbreak took place. This requires work with fossilized evidence of the outbreak and the context in which it took place, as it was captured in the written documents.

Reconstructing micro-level relationships between TB, the host population, and the socio-economic environment may prove to be challenging. The initial reconstruction of an old outbreak, of all relationships between the agent – TB, the population, and the environment, may similarly be off the mark. And some of these fossils, being data collected and documents written by people, may not be as objective as bones.

1.1. Research problem

TB in the 21st century still causes significant morbidity and mortality. Despite relatively low numbers of cases and deaths in Canada, TB is an infectious disease that could spread, particularly in the era of drug-resistance and co-infection, currently TB-HIV, but in future perhaps with new viral diseases. TB also tends to affect those who are least able to protect themselves, or those around, from the disease.

Canada's population falls into three categories according to TB incidence: (1) Indigenous peoples: First Nations, Metis, and Inuit (FNI), (2) the foreign-born population, in particular immigrants from countries with a high incidence of TB, and (3) Canadian-born non-Aboriginals. Rates of TB disease among the Canadian-born non-Aboriginal population are the lowest and approach elimination, suggesting there is no inevitability about the disease. Despite availability of modern antibiotics, in 2013, the TB incidence rate in the Indigenous population was thirty three times higher than those among the Canadian-born non-Aboriginals (PHAC., 2015).

In the case of TB in the Indigenous population of the Prairies, the beginning of the TB outbreak has been relatively well-documented thanks to the DIA data collection efforts. Analysing the role of the social determinants of health in early days of the TB outbreak may allow us to understand how the TB outbreak started, how it progressed, and how it may have been eventually contained. It may also provide a different perspective, as our current understanding is based on the relationship between TB, its host, and the environment of the last 200 years of data, while the relationship itself started thousands of years ago.

This thesis will address the following research objectives:

1. Can a TB outbreak that took place over a century ago be reconstructed?

   The goal of this research objective is to explore the reconstruction of a TB outbreak of over a century ago. This objective raises several subsidiary research questions:

   a. How can the complex web of relationships between agent, host, and environment be untangled and examined?
   b. What are the specific tools for modelling these relationships?
   c. What data are required for modelling and what are the best available data sources?
   d. What can we learn from the process of reconstruction itself?
2. What can we learn about the relationships between the disease, the population it afflicted, and the environment in which the outbreak took place?

   The objective is to use the reconstruction of the outbreak to answer the questions:

e. What is the relative importance of social circumstances and biological factors in driving the outbreak?

f. How did the actual social conditions affect the outbreak?

3. How can reconstruction of a TB outbreak be used to evaluate policy interventions?

The goal of this research objective is to explore the application of this model to other settings, such as other disease outbreaks in the Canadian Indigenous population.

1.2. Plan of the thesis

The following chapters present a response to the three research questions. Chapter 2 presents the literature, the history of the outbreak and the socio-economic context within which it took place.

The Methodology (Chapter 3) presents a technique to assemble relevant information to reconstruct the outbreak.

Analyses and Results chapter (Chapter 4) presents results of the reconstruction of the TB outbreak in the Qu'Appelle Agency, based on the methodology presented in Chapter 4.

Finally, Chapter 5 presents conclusions regarding the possibility of reconstructing an outbreak that took place so far back in history, as well as conclusions regarding the relative importance of biological and socio-economic factors in the outbreak.

2. Literature review

In the late 19th century, the Qu'Appelle Valley was the "Wild West" of a Western Canada that was only beginning to be colonised and populated by people of European descent. Despite this fact, the data on TB are of good quality. The fact that the Indigenous population was a ward of the state entitled to treaty payments following the signing of the treaty in 1874 was largely responsible for the quality of TB data. The state was interested in keeping an accurate population count in order to minimise treaty payments, and therefore it needed regular censuses of the Indigenous population and accurate birth and death certificates.

TB data can be derived from two sources: (i) annual reports of the Indian Agents, and (ii) death certificates that state the cause of death. While reports of Indian Agents provide general contextual information on the course of the epidemic, death certificates are much more specific.

Annual reports of Indian Agents tend to provide qualitative data on the course of the epidemic, and describe it in terms of it being on the rise or decline. To some extent these reports may be a product of their time and reflect prevailing ideas as to what was considered normal when it comes to TB in an Indigenous population.

Unlike the reports of Indian Agents, death certificates provide the exact number of people who died from TB. These were the source of TB mortality data presented in Figure 1-1(Ferguson, 1928).

Mortality data alone are not sufficient for studying disease outbreaks. An outbreak is defined in terms of a change from the normal level of a disease in a certain time and place. Is a mortality rate of 100 per 1,000 high or low? One simply cannot answer unless one knows which disease, which population, and which period in history is being discussed.

In fact, these three elements, the disease, population, and time and location, are three elements of a standard epidemiologic model for studying outbreaks called the Epidemiologic Triad (R. Miller, 2002, p.64). This model has three basic elements: (1) agent, (2) host, and (3) environment. The agent is the cause of the disease; the host is the organism afflicted by the disease; and the environment hosts the external factors that bring agent and susceptible host together and contribute to disease transmission. This framework provides a useful structure for studying the TB outbreak in the Qu'Appelle Agency by identifying a set of elements that have to be considered. The following sections present information on the published literature for each element of the triad.

2.1. The Agent

TB is caused by mycobacteria that belong to the *Mycobacterium tuberculosis* complex – a genetically related group of organisms that include *M. tuberculosis, M.bovis, M.bovis BCG, M.africanum, M.caprae, M.microti* and *M.pinnipedii* (Long & Ellis, 2007) (p.19). TB is an airborne infectious disease spread by aerosol droplet nuclei (Menzies et al. 2014) (p.26). It is primarily a disease of the lungs, but it can affect any part of the body. Transmission of the disease occurs as a result of close contact with a person with active pulmonary disease who can transmit *M.tuberculosis* during activities such as coughing, sneezing, or singing (Heymann, 2015) (p.640).

The bacteria in the form of droplet nuclei are inhaled and infection occurs primarily in the middle and lower lung zones that receive the most ventilation (Menzies et al., 2014) (p.26). Then immune response occurs, as the bacteria begin to replicate exponentially, doubling every 24 hours (Menzies et al., 2014) (p.26). The host develops specific immunity 3 to 8 weeks after initial infection (Menzies et al., 2014) (p.26), at which point *M.tuberculosis*-specific lymphocytes begin migrating to the site of infection (Menzies et al., 2014) (p.26).

If this initial immune response cannot destroy the bacteria, they replicate until the macrophage bursts and releases bacteria into the body. These bacteria are in turn captured by new macrophages, continuing the cycle. Bacteria may spread to other parts of the body through the lymphatic or circulatory systems (Menzies et al., 2014) (p.26).

The outcome of this process is creation of caseous and necrotic granuloma (Menzies et al., 2014) (p.26). At this point, infected individuals usually do not show any symptoms (Menzies et al., 2014) (p.26).Primary TB infection develops in approximately 5% of individuals, primarily young children and immunocompromised who are unable to contain the infection. Primary TB may manifest itself as a progressive Ghon focus, disseminated or milliary TB, and central nervous system disease from 2 to 6 months after infection(Menzies et al., 2014) (p27).

Approximately 90% of individuals who do not develop primary TB are left with latent TB infection and never develop active disease (Menzies et al., 2014) (p.27). The remaining 5% of individuals with latent TB develop late disease progression (Menzies et al., 2014) (p.27). Latent TB is a condition during which TB bacteria that were not destroyed by the initial immune response lay dormant in small granulomas. *M.tuberculosis* can survive in that state for years, and may later lead to reactivation of TB and active pulmonary disease, if the immune competency of the host deteriorates (Menzies et al., 2014) (p.30). Symptoms of pulmonary TB include chronic cough, fever and night sweats, hemoptysis, anorexia, and chest pain (Menzies et al., 2014) (p.45).

2.2. The Host

The following sections introduce the host – the people of the Qu'Appelle Agency, and how they interacted with TB.

2.2.1. Demographics

Total population of the Agency at the time of its creation was 1492 people (DIA., 1886). Early population estimates may have been inaccurate because of population movements. The Population of the Agency began to decline immediately after the establishment of the Agency and continued to do so until 1915, at which point it was 757 people, or almost half of the initial population (DIA., 1915, p.46). A similar or greater decline was observed in the Blackfoot[1] nation (Figure 2-1).

[1] Blackfoot Nation, located in Southern Alberta, was also studied by Ferguson, and thus presents a good comparison to the Qu'Appelle Agency

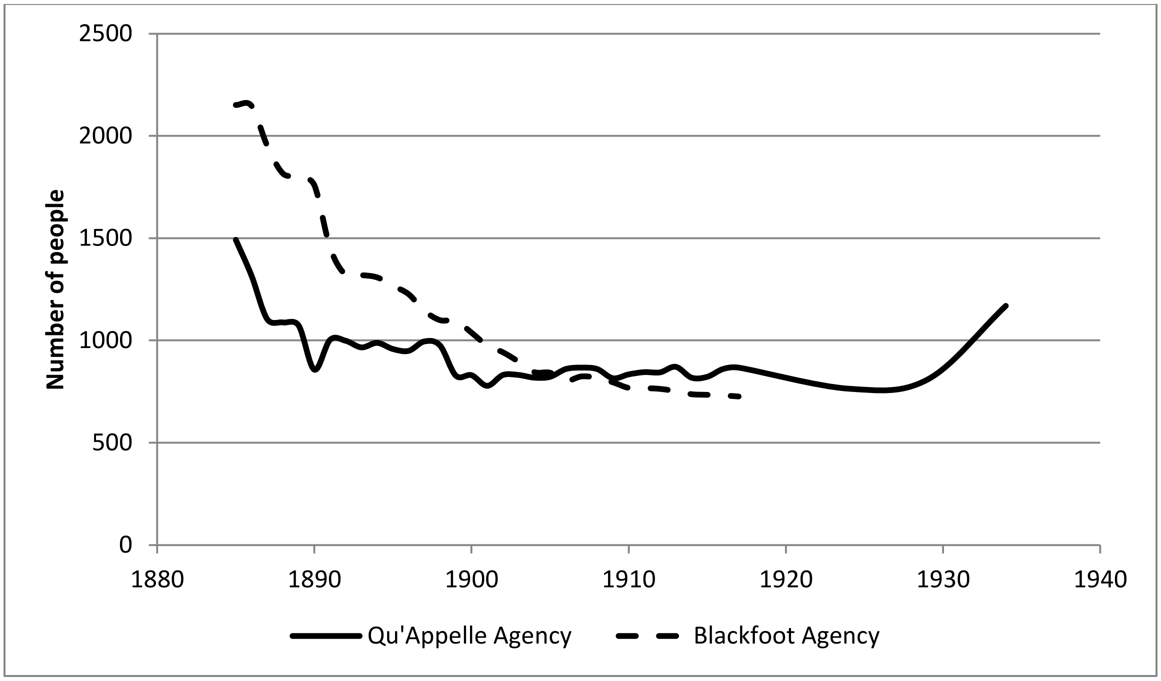

Figure 2-1: Population of Qu'Appelle and Blackfoot Agencies, 1885-1934. Source: Department of Indian Affairs (1885-1934).

Population age and sex distributions are presented in Figure 2-2 and Figure 2-3. The population of the Qu'Appelle Agency was relatively young, with children under 6 making up approximately 20% of the population, or almost 40% together with children 6 to 15 years old (Figure 2-4).

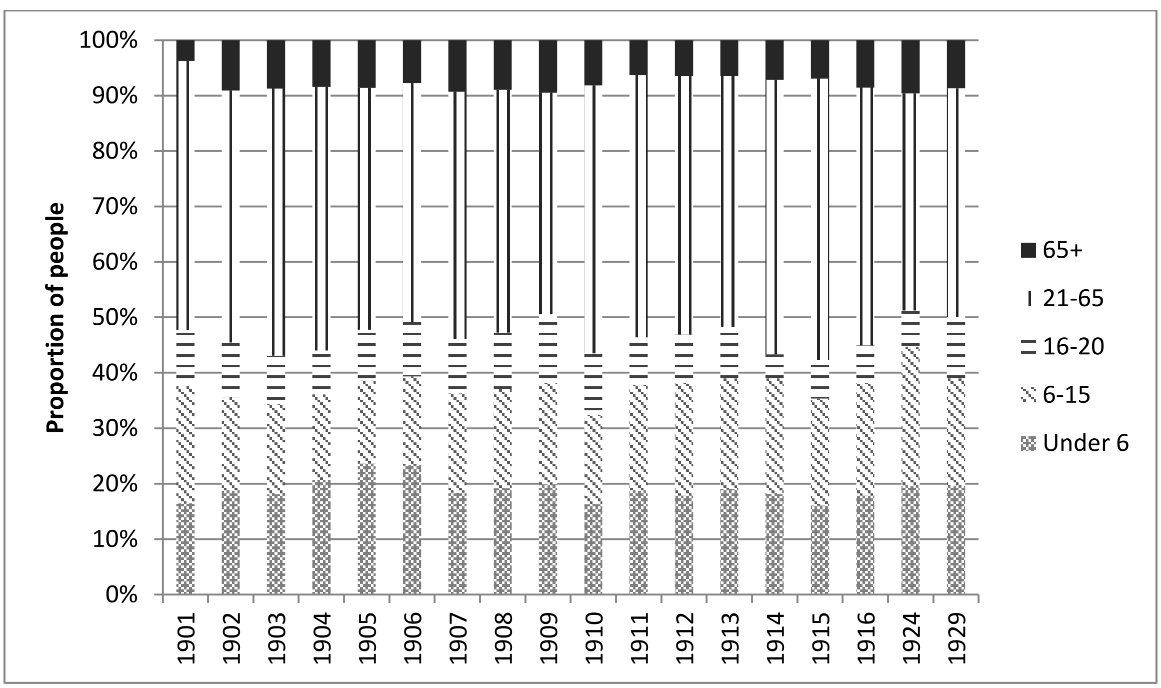

Figure 2-2: Population age and sex distribution, males, Qu'Appelle Agency. Source: Department of Indian Affairs (1901-1929)

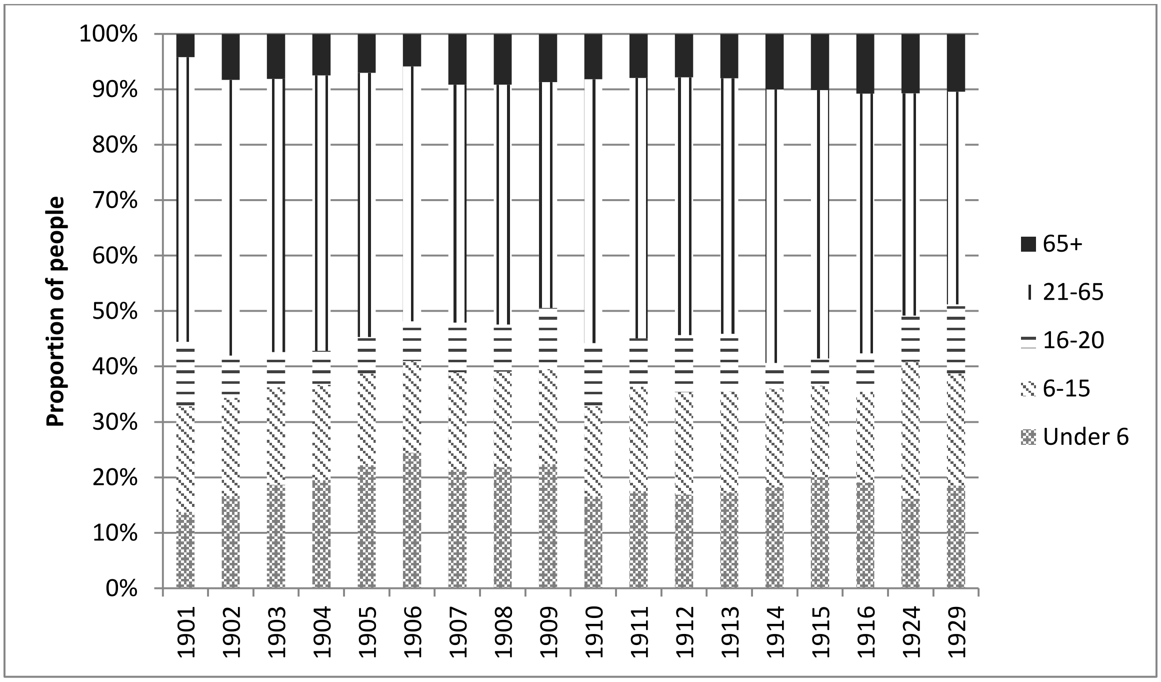

Figure 2-3: Population age and sex distribution, females, Qu'Appelle Agency. Source: DIA (1901-1929)

While the DIA did not provide data on the number of households, using data on the number of houses and assuming that there was one household per house, the number of nuclear households increased from 161 in 1885 to 243 in 1926 (Figure 2-5).

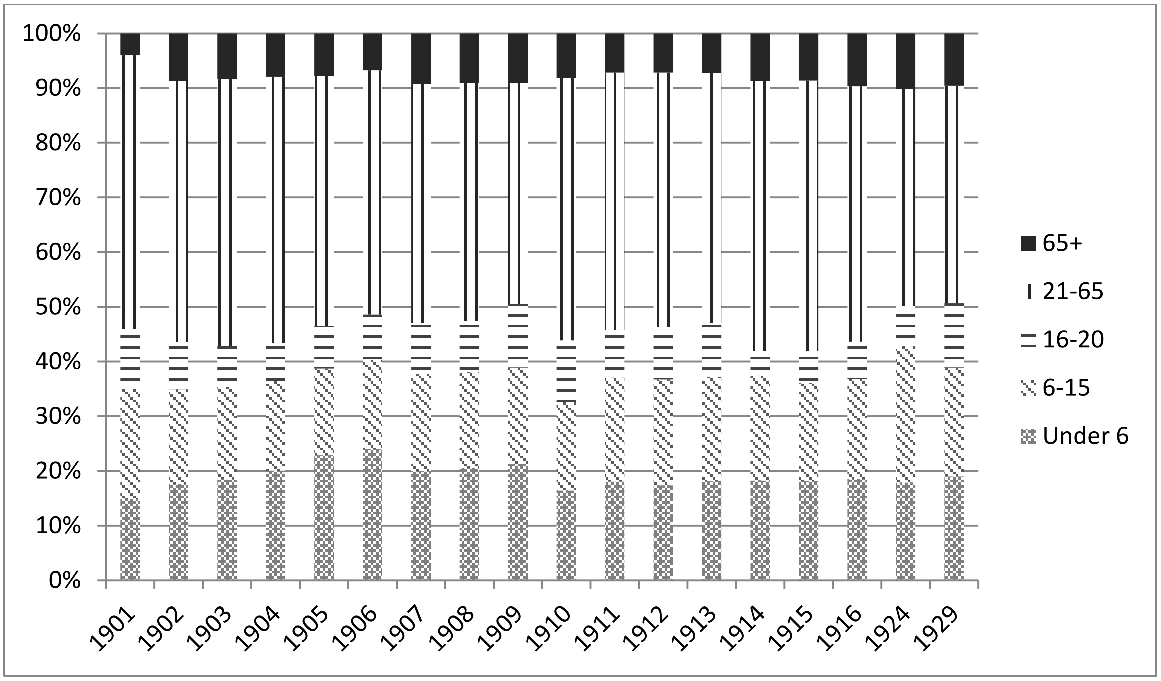

Figure 2-4: Proportions of population by age group, males and females, Qu'Appelle Agency. Source: Department of Indian Affairs (1901-1929)

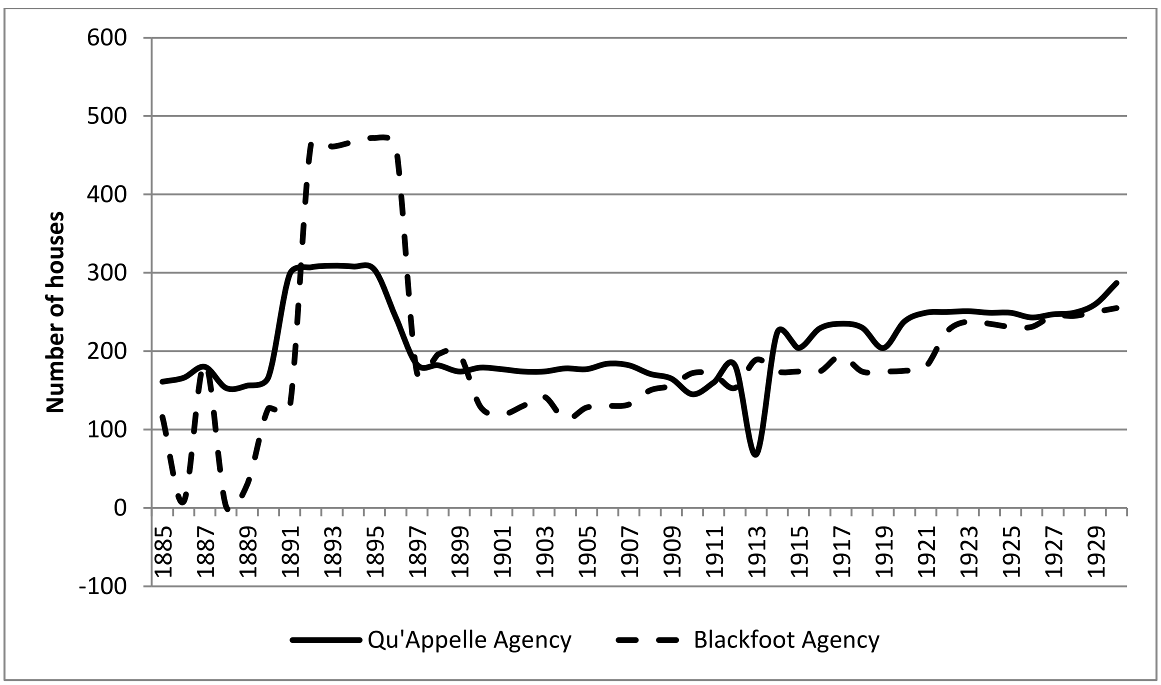

Figure 2-5: Number of houses in the Qu'Appelle and Blackfoot Agencies. Source: Department of Indian Affairs (1885-1926)

Housing is an important factor in the spread of TB, as will be discussed further below. Figure 2-5 presents an example of how different Indigenous nations were going through the same social processes. The number of houses and number of people per household are presented in Figure 2-6.

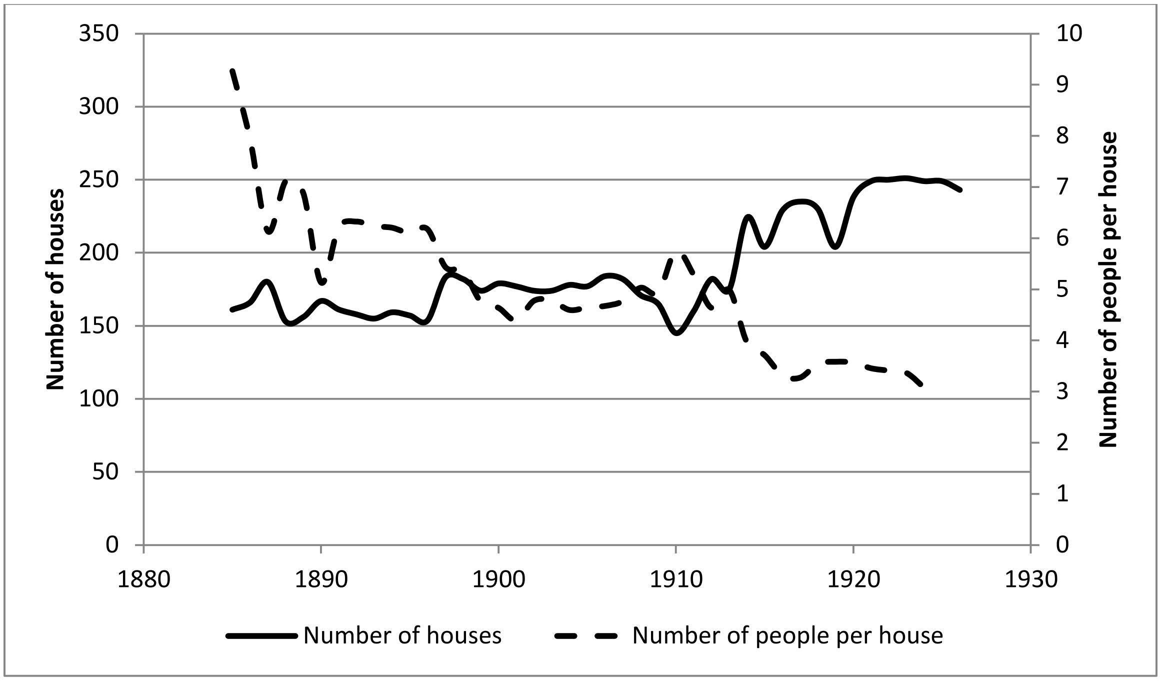

Figure 2-6: Estimated number of houses and population per house, Qu'Appelle Agency. Source: Department of Indian Affairs (1885-1926)

As Figure 2-6 demonstrates, the number of people per house fell from 9.3 in 1885, to just over 3 in 1924. While the population declined roughly by half from 1492 in 1885 to 764 in 1924, the three-fold decline in the number of people per household suggests that either the living arrangements changed significantly, or the data for the number of houses overestimate the actual number of houses in use for reasons mentioned above. Another possibility is that this decline reflects the removal of children to residential schools.

The number of people per house appears to be low, particularly in 1924, but since census data published by the DIA do not provide the number of families, the number is difficult to verify. The number 3.3 people per house was reached at the end of a long period of population decline, and therefore does not represent a normal situation. Secondly, DIA was interested in demonstrating progress, so they may have counted uninhabited houses as well as inhabited ones.

In summary, at the time the Qu'Appelle Agency was established, the population of the Agency was 1492 people. The population was young, with people under the age of 16 making up almost

40% of the population. Despite the significant decline in the number of people, the make-up of the population did not change significantly. In addition to being relatively young, the population of the Agency may have had a relatively high population density, particularly in the early years.

While the preceding discussion focused on the population of the Agency, it is clear that it is difficult to separate it from its environment. Metrics such as population density describe the population in its environment, which is important not only for understanding demographic changes over time but, as the Epidemiologic Triad demonstrated, for disease dynamics as well.

2.2.2. The Agent and the Host

TB existed in the Americas prior to the arrival of Europeans. Over 133 pre-contact TB cases are currently documented (J. Waldram, Herring, & Young, 2006) (p. 31), with mycobacteria similar to modern one identified in remains going back to 160 CE (Gomez i Prat & Mendoca de Souza, 2003). TB was present in Arizona, Ontario, and the Prairies, and existed in societies at various stages of socioeconomic development (Gomez i Prat & Mendoca de Souza, 2003). While it is not known which TB strains circulated in the Americas prior to the arrival of Europeans, currently circulating strains can be traced back to Quebec (Pepperell et al. 2012a).

Currently, the TB incidence rate in the Indigenous population is 33 times higher than in the Canadian-born non-Aboriginal population (PHAC, 2015). TB incidence rates within the Indigenous population differ as well, leading some to hypothesize that they are related to time of contact (Table 2-1).

Table 2-1: Time of contact and TB incidence in Status Indians (S. A. Grzybowski & Allen, 1999)

Region	First contact	Incidence (1970-1974)	Incidence (1990-1994)
Atlantic	17th century	100	7
Pacific	18th century	125	52
Prairies	19th century	225	79
Territories	20th century	250	97

Despite the fact that TB was present in the Americas for centuries, some have suggested that genetic susceptibility to TB brought about by the lack of prior exposure to the disease could explain high TB mortality in Indigenous communities (S. Grzybowski & Allen, 1999; J. B. Waldram, Herring, & Young, 2006).

Recent studies evaluated both the prevalence of genes conferring susceptibility to TB, as well as their contribution to developing the disease. The first study to attribute TB susceptibility to specific genes identified four natural resistance associated macrophage protein 1 (NRAMP1) polymorphisms that were significantly associated with TB (Bellamy et al., 1998). A recent meta-analysis summarised evidence on the contribution of NRAMP1 to the risk of developing TB, estimating an odds ratio of 1.35 (Li et al., 2011).

The prevalence of NRAMP1 polymorphisms in Canadian Indigenous population was estimated at the nation (Larcombe et al., 2015) and family level (Greenwood et al., 2000). Of particular interest is a study that evaluated the frequency of NRAMP1 polymorphisms among the Saulteaux, Cree, and Dene. Unlike the Dene who have a significantly higher frequency, Cree and Saulteax of the Qu'Appelle Agency, have NRAMP1 polymorphisms frequency similar to that among European Canadians (Larcombe et al., 2015).

Since NRAMP1 increases the risk of developing TB, and therefore the risk of death, it is likely that the frequency of these polymorphisms in the Indigenous population was higher in the past. TB incidence rate might be related to the time of contact (Table 2-1), and the Dene who live further north may have come into contact with TB more recently than the Cree and Saulteaux of the Qu'Appelle Valley.

As there was TB in pre-contact Americas, why did the Indigenous population suffer from such high TB mortality rates after the contact? Some argue that high mortality from TB could be explained by the genetic factors and lack of exposure to European strains of the disease (S. Grzybowski & Allen, 1999). Others argue that TB outbreak in the Qu'Appelle Agency took place when the Indigenous population was under tremendous social stress given rapidly changing environment, stressing the importance of social determinants of health (G. A. Clark et al., 1987). The two explanations have something in common: given the geographic variation in frequency of NRAMP1 polymorphisms and TB incidence rates favoured by one side of the argument, and social determinants of health favoured by the other, it is clear that environment plays a significant role in the explanation of TB outbreak in the Qu'Appelle Agency.

2.3. The Environment

Environment is the third part of the Epidemiologic Triad that describes the place of the outbreak. The following sections describe the physical and socio-economic environment of the Agency.

2.3.1. Geography

The Qu'Appelle Agency was created July 1, 1885. Located in the Qu'Appelle valley of Southern Saskatchewan, the Agency occupied a total area of 301 square miles, stretching between Regina and the border with Manitoba (DIA., 1886, p.122; DIA., 1888, p.74).

In addition to Regina, the Agency is surrounded by several towns such as Fort Qu'Appelle, White City and Indian Head to the south, Southey to the north, and Melville to the east (Figure 2-7). The Agency is located in a densely populated part of Saskatchewan, and was well-connected to rail lines. While rail lines were just being built in the late 19th century, Indigenous trade networks that crossed the entire continent existed for centuries (Dickason, 2002). Access to the railroad may have been beneficial to the Agency from the economic point of view, but it may also have exposed its inhabitants to the infectious diseases brought by the immigrants.

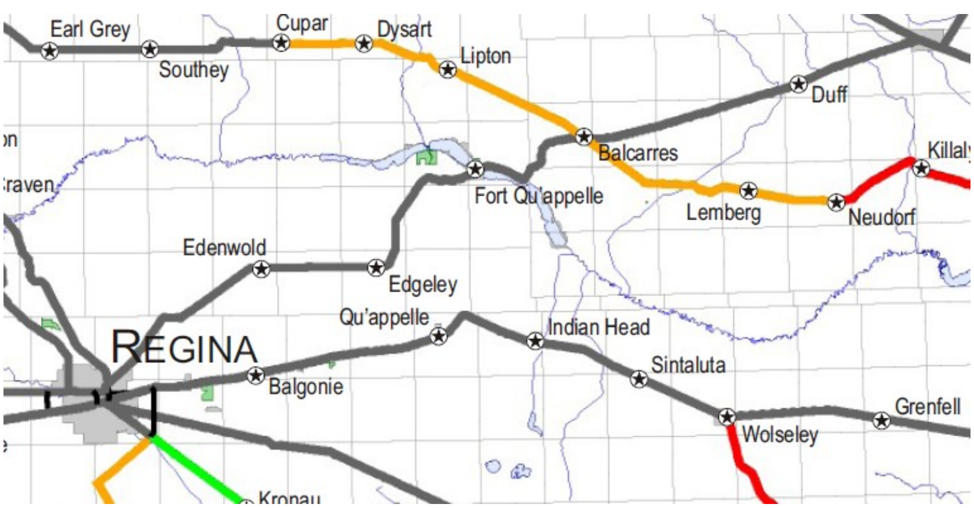

Figure 2-7: Rail lines, Southern Saskatchewan. Source: (Zimmer, 2008)

The role of trade networks in the spread of infectious diseases in general, and TB in particular, will be discussed in section 3.1 below.

2.3.2. Housing

Housing was a source of concern to the Indian Agents, as is made clear by the ubiquitous presence of the topic in their annual reports. More specifically, Indian Agents were concerned with two aspects of the housing: quality and location. These two aspects may also be important considerations when it comes to disease spread. The houses were only occupied during winter, (DIA., 1900) (p. 160) a season during which presumably people would spend much time indoors. In the summer, when members of the bands moved to the lodges, their health improved, according to the Indian Agent (DIA., 1900) (p. 162). This aspect may have made the quality of housing particularly important when one considers its impact on disease spread.

2.3.2.1. Housing quality

The concern with the quality of houses existed from the time the Agency was formed. It was poor in the beginning, but improved over the years, as indicated by Indian Agents: "the fittings of the dwelling houses are more complete, and in general the surroundings more comfortable" (DIA., 1894) (p. 55). Later the Agent noted the appearance of tables and chairs, (DIA., 1896) (p. 135) then ventilation and burning of refuse, (DIA., 1897.) (p. 188) and what may have been the first shingle roof, (DIA., 1897.) (p. 190) which became more general over time (DIA., 1899.) (p. 156). While the houses were getting more comfortable, building large houses proved a challenge because of the lack of suitable lumber (DIA., 1898) (p. 162).

Similar to all other aspects of life in the Agency, there may have been differences in housing quality across the bands. Writing about Pasquah band in 1897, Agent Lash remarked that "the Indians on this reserve are building a better class of houses, and some of them are very comfortable, with second stories and suitable furniture" (DIA., 1898) (p. 164).

There were also differences in the quality of the housing within the bands:

> “On this reserve there are two classes of Indians. One class consists of those who have been sufficiently progressive to remove from the original settlement in the valley and take up separate holdings in the farming lands of the reserve on the uplands. These have better houses than the second class, who are those who have not energy enough to make the change, and whose houses are of an inferior class. Among the former the sanitary conditions are better than among the latter. All, however, give some measure of attention to keeping their dwellings and premises clean and in proper condition, and each spring sees each place thoroughly cleaned and the refuse that has accumulated during the winter, is burned” (DIA., 1900) (p. 163).

Another aspect of the quality that Indian Agents remarked on constantly was the cleanliness.

> “There was a marked improvement in the sanitary condition of the houses. The rubbish in the spring was early raked together and burnt, so that the vicinity of the houses presented a very creditable appearance” (DIA., 1890) (p. 56).

Ferguson provided a timeline of changes in housing quality in the Agency. In 1875 the residents of the future Agency lived in tepees (Figure 7-1). The first log huts with fireplaces appeared around 1879, and were replaced by log houses with wooden floors by 1894. By the turn of the century residents of the Agency lived in larger log houses with wooden floors in the majority of the houses. Fireplaces which lacked chimneys remained in less than half of the houses. Such fireplaces had a negative impact on the air quality inside the house. By 1914, these houses were being replaced by larger log houses with good windows and no fireplaces (Ferguson, 1928).

2.3.2.2. Location

The issue of location of houses on reserves refers to the existence of the village system in the Qu’Appelle Agency and the attempts of the Indian Agents to destroy it. Initially, members of the bands were allowed to settle in villages, but soon Indian Agents discovered that the village system interfered with their ‘civilizing’ policies and attempted to break it down.

In words of the Indian Agent Mitchell,

> “On all the reserves in this Agency the Indians have in the beginning been allowed to settle in villages in the Qu'Appelle Valley. This militates very seriously against the advancement of the Indians, as they are too far from their farms, which are on the uplands, and the close proximity of the houses affords too much opportunity for visiting and gossiping and the promotion of gambling among the men during the winter nights. It

> also tends to strengthen the clannish feeling which renders it so difficult to deal with individual members of the bands instead of with the band as a whole. Efforts have been made for years past to break up the villages, but except in the case of Pasquah's Reserve not much success has been attained thus far" (DIA., 1900) (p. 166).

Bands differed in their attachment to the village system. By 1899 the Agents succeeded in breaking up the system on the Pasquah reserve. Despite that fact, there was "an increasing tendency on the part of the better Indians on both Piapot's and Muskowpetung's Reserves to abandon the village system and build on their farm holdings on the uplands" (DIA., 1900) (p. 166).

The preceding report may have been slightly optimistic, as the next year Agent Mitchell reported that on the Piapot reserve "one house has been this year erected on the 'bench' - the beginning, I trust, of the long planned for abandonment of the village system" (DIA., 1901) (p. 167).

The movement away from the village system proceeded at a different pace among different bands. The same year as the first member of the Piapot band built a house on the 'bench' – land in the hills, rather than the low land in the bottom of the valley, there were only three houses left in the Standing Buffalo village, the rest having moved to the 'bench' already (DIA., 1901) (p. 173).

In 1905, buildings on the Piapot reserve were "scattered along the valley, which is a great improvement on the village system" (DIA., 1906) (p. 145). On another reserve the process of break-up of the village system was supposed to be completed in 1907 (DIA., 1907) (p. 141).

The break-up of the village system may have contributed to the stratification of the bands between 'the better Indians' who moved to the bench, and those, who chose to stay behind. Describing the Pasquah band in 1901, Agent Graham wrote that there were two groups of people: those who lived in the village and those who lived on the bench. New houses located on the bench were of better quality, one and a half stories high with an upstairs, and shingled or thatched roofs (DIA., 1902a) (p. 156). Thus, the break-up of the village system contributed to the creation of classes in the Qu'Appelle agency. As will be seen later, social stratification and socioeconomic inequality that comes with it are important social determinants of health.

In addition to contributing to the process of social stratification, the break-up of the village system is important for other reasons. Some incentives such as free seeds were available only to those who agreed to relocate (DIA., 1888), thereby perpetuating existing inequality.

Break-up of the village system may have changed interaction patterns in the Agency. Close proximity of houses in a village allows one to interact with a large number of people outside of one's household. With the move to higher ground, interaction may have become limited to the members of one's own household. Lack of opportunities for spending time outside of one's own house, particularly in winter, may have contributed to the spread of TB within households.

2.3.3. Economy

DIA paid significant attention to the economy of the Agency, and its annual reports provide a wealth of information on the state of the economy and major occupations of the inhabitants of the Agency. The economy underwent significant changes between establishment of the Agency in 1885 and the 1920s.

Initially, the economy of the Agency was based to a large extent on collection and sale of dry wood, cutting of hay, and wage labour on farms surrounding the Agency (DIA., 1897., p.188). DIA was very involved in the development of the economy of the Agency, one of the main motivations for which was to civilize the population and to reduce the burden on the department (DIA., 1892, p.54).

At the urging of the Department and with the exhaustion of the supply of dry wood, the economy of the Agency began to change, and in 1899 for the first time farming and ranching were mentioned as primary occupations of one of the bands (DIA., 1900, p.163). Adoption of farming and ranching was neither quick, nor enthusiastic. It appears that initially few individuals were involved in each type of activity and that the number slowly increased over time.

Ferguson provided a timeline of changes in economic activity on reserve. In the 1870s, members of these bands were buffalo hunters, forced by the extermination of the buffalo to look for other sources of livelihood. By the late 1870s, the population of the future Agency began to raise cattle using animals loaned to them. By the early 1890s, they were engaged in ranching. Small farming followed in the first decade of the 20th century, with full-fledged farming activity being carried out in the Agency towards the end of the second decade of the 20th century(Ferguson, 1928).

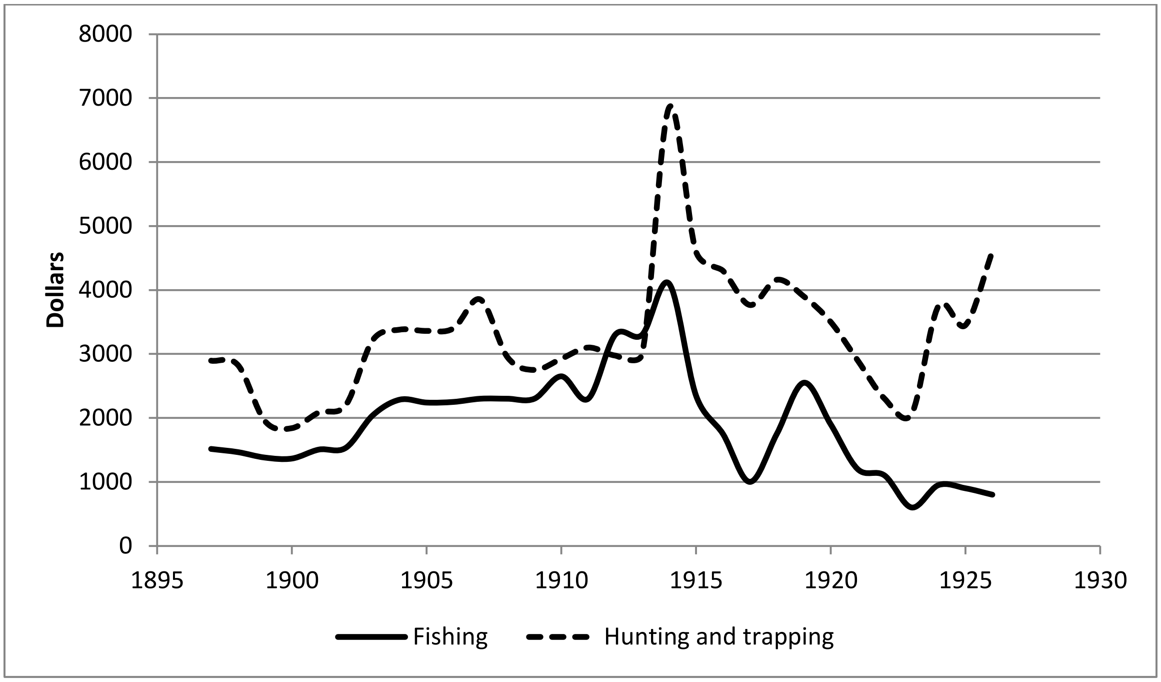

Figure 2-8: Income from fishing, hunting, and trapping, Qu'Appelle Agency. Source: (DIA 1885-1926)

2.3.3.1. Subsistence economy

When DIA started collecting data on income towards the end of the 19th century, the main sources of income for the band were farming, ranching, wage labour, and other activities (Figure 2-9). That was not always the case, and the transition to agriculture was slow (Figure 2-12). Since there are no data, it is not clear whether subsistence activities, such as fishing and hunting, were a significant source of income in the late 19th century, when agricultural output was low.

When DIA began collecting income data, hunting and fishing accounted for just over 13% of total income. With time, importance of these activities declined, with fishing and hunting accounting for 2.4% of total income in 1926 (Figure 2-9).

2.3.3.2. Income economy

Total annual income of the Agency increased over time from $33,173 in 1897, (DIA., 1898, p.450) when data began to be collected, to $109,544 in 1930 (DIA., 1930, p.75). The following sections provide a more detailed account of each type of economic activity.

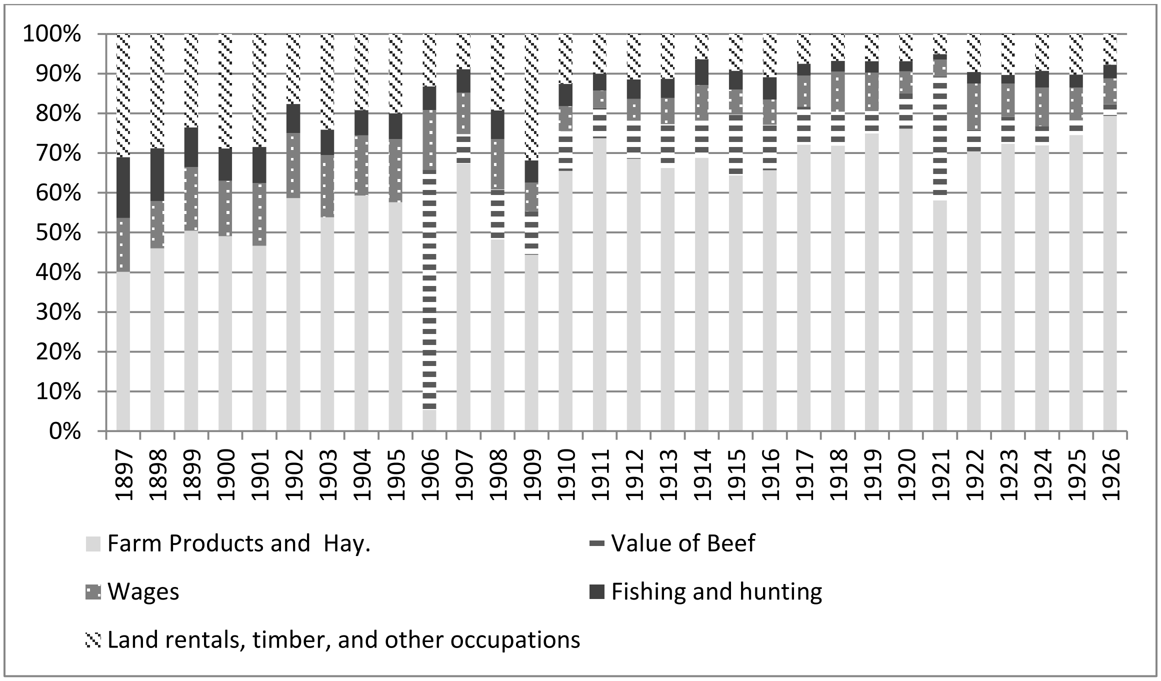

Figure 2-9: Proportion of total income, by type of activity. Source: Department of Indian Affairs (1897-1926)

2.3.3.2.1. Sale of dry wood

Sale of dry wood was one of the first major types of economic activity in the Agency after the Agency was created. Despite its importance as a source of income, there are little data to quantify the number of people involved, or the volume and price of wood sold. Income data that could be used to estimate the relative importance of various sources of income began to be published when the sale of dry wood was losing its importance. In reports for preceding years, earnings of the dry wood industry were described in a way that does not allow for disaggregation from other types of activities: “The individual earnings for the year of the three bands, viz., Pi-a-pot, Muscowpetung and Pasquah, which have come directly under my notice, from the sale of hay and wood, working, freighting, & c., & c., amount to $5,217” (DIA., 1891, p.41).

Toward the end of the century, as the supply of dry wood was becoming exhausted, “greater eagerness to become cattle-owners is being noticeable”(DIA., 1901, p.167).

2.3.3.2.2. Cutting of hay

Unlike sale of dry wood, the cutting and sale of hay was much better documented (Figure 2-10). The total amount of hay cut followed two major cycles, the first beginning in 1885 and ending in 1907, and the second beginning in 1907 and ending in 1930. During the first cycle, the amount of hay cut increased from a low of 1,077 tons in 1885 to an early peak of 4,278 tons in 1899, and then dropped to 1,100 in 1907. The second cycle saw many fluctuations, during which the amount of hay cut increased from a low of 1,100 in 1907 to a high of 6,955 in 1914.

The amount of hay cut probably depended on many things, one of them being technology. Investments in technology began fairly early on, with the purchase of a mower being mentioned for the first time in 1895 (DIA., 1896, p.138). The number of mowers steadily increased between 1894 and 1916, at which point it increased from 90 in 1916 to 240 in 1917 (Figure 2-11).

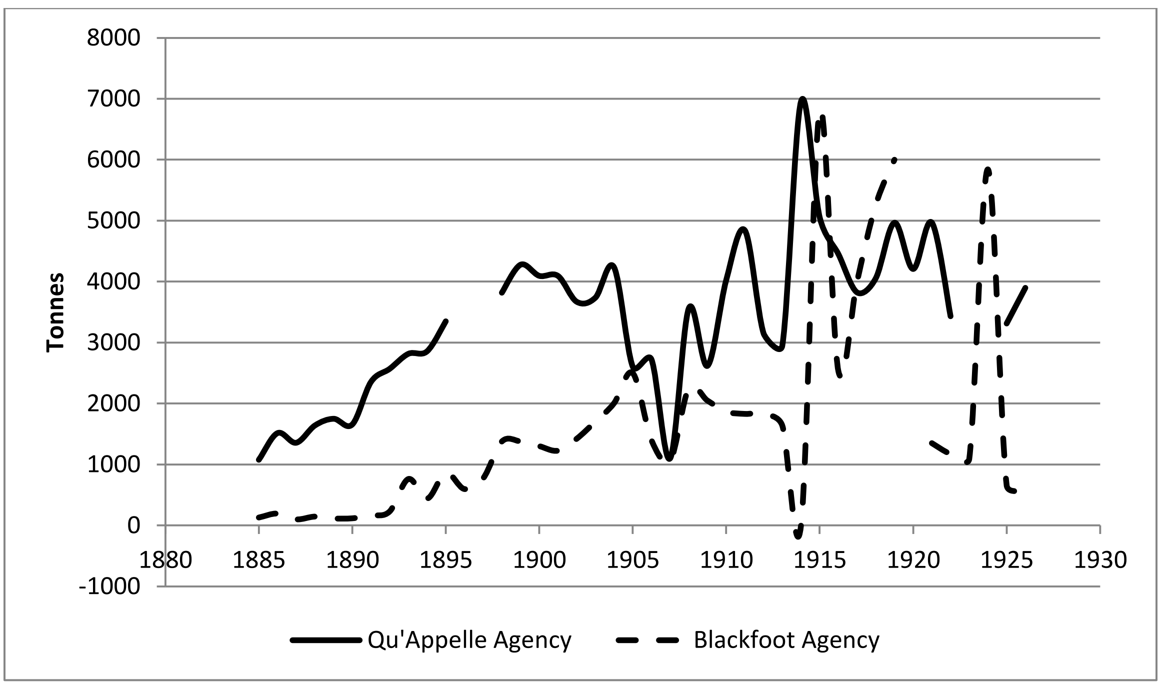

Figure 2-10: Hay cut, tons, Qu’Appelle Agency and Blackfoot Agencies. Source: Department of Indian Affairs (1885-1926)

The importance of hay as an income-generating resource may have varied across the bands of the Agency. In 1902, members of the Piapot band sold "a large quantity of hay and wood in the town of Regina, and a great deal of their time is taken up with this work" (DIA., 1902). Hay, however, was not mentioned as an occupation of the Muskowpetung's band (DIA., 1902, p.163).

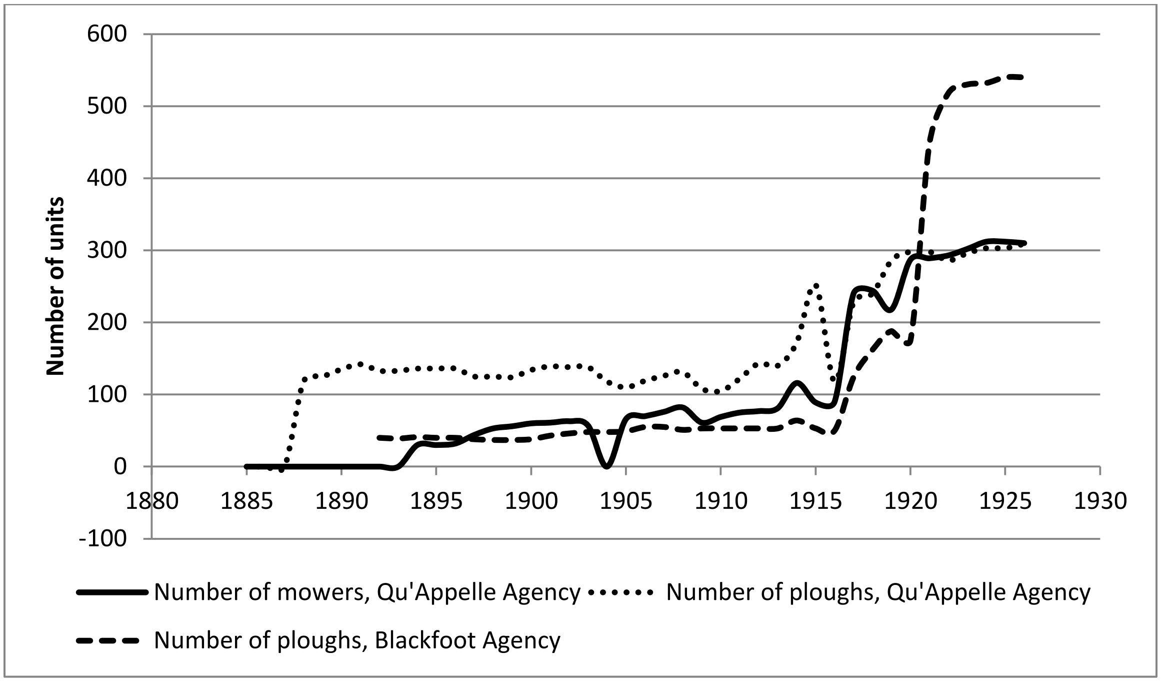

Figure 2-11: Number of mowers and ploughs, Qu'Appelle and Blackfoot Agencies. Source: DIA (1885-1926)

In addition to variation across bands, there may have been variation in use of hay across time. As the importance of raising livestock increased, production of hay may have been reoriented to supply internal needs, rather than to be sold to settlers.

2.3.3.2.3. Farming

The Agency produced a variety of root and grain vegetables, such as wheat, oats, barley, corn, potatoes, carrots, and turnips (DIA., 1913, p.115). Farming was the most important source of income (Figure 2-9).

Agricultural output was unevenly distributed across the various types of output, with wheat, oats, potatoes, turnips, and hay being the main agricultural products. While the prairies are famous for producing grains, potatoes were also a successful crop. As an Indian Agent wrote in 1899, "Roots are a successful crop on this reserve and large quantities of potatoes are sold by these Indians each year at remunerative prices"(DIA., 1900).

The process of adoption of agriculture in the Qu'Appelle Agency was slow, despite efforts of the DIA aimed at its promotion. Initially, abundance of dry wood and hay provided sufficient income on the one hand, and may have been easier to practice than agriculture, both of which may have contributed to the slow adoption of agriculture in the Agency (DIA., 1900, p.160).

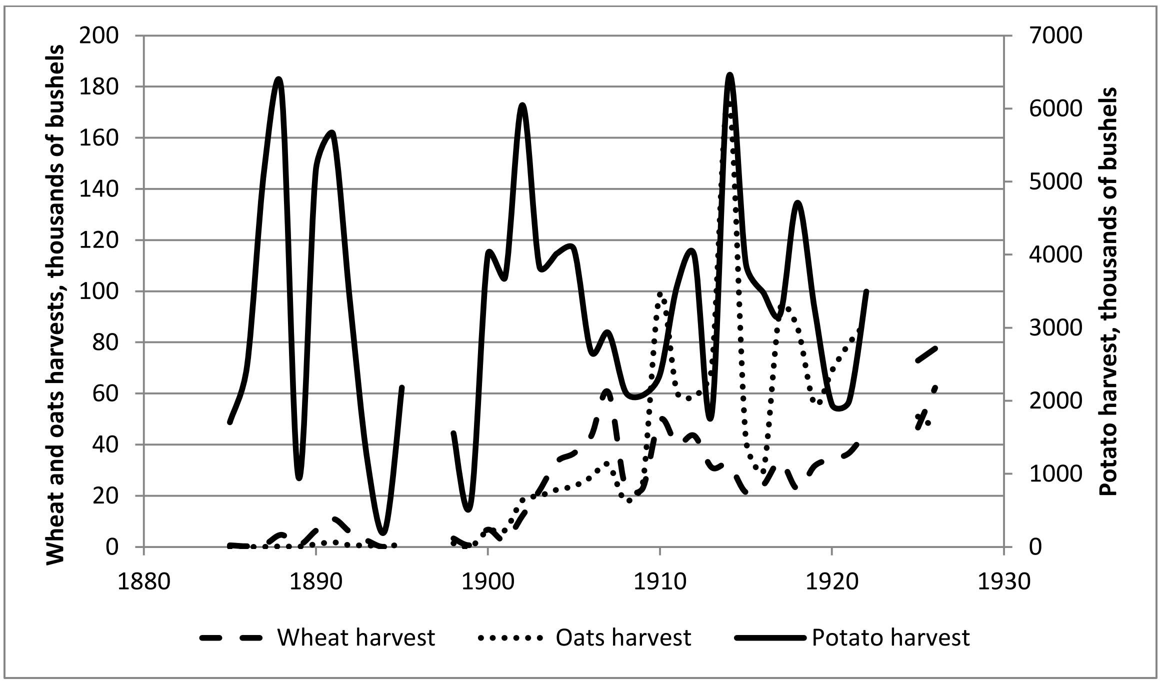

Figure 2-12: Wheat, oats, and potato harvest, thousands of bushels, Qu'Appelle Agency. Source: DIA (1885-1926)

As Figure 2-12 demonstrates, output of wheat and oats was low in the beginning. The causes for the low output are not entirely clear from the reports of the Indian Agents. Oftentimes in these early years, Indian Agents would cite poor weather as the cause for poor harvest. A typical statement from the annual reports would read like this: "The acreage, under grain this spring was increased fifty per cent over last year, and the prospects were most encouraging up to the early part of June, but the continuous drought from that date injured the crop and our returns this season will be comparatively small"(DIA., 1890, p.57).

Presumably the climate did not change significantly in the beginning of the 20th century when the output of wheat and oats began to increase rapidly. A likely explanation of the low initial agricultural output and a fast increase in the first decade of the 20th century is a slow adoption of agriculture. It is likely that only a few people adopted agriculture initially, but as the supply of dry wood dwindled, more and more people became actively involved in farming, possibly influenced by the policy change on the part of the DIA. This hypothesis is supported by the reports of the Indian Agents in parts that describe a high rate of increase in area under crop.

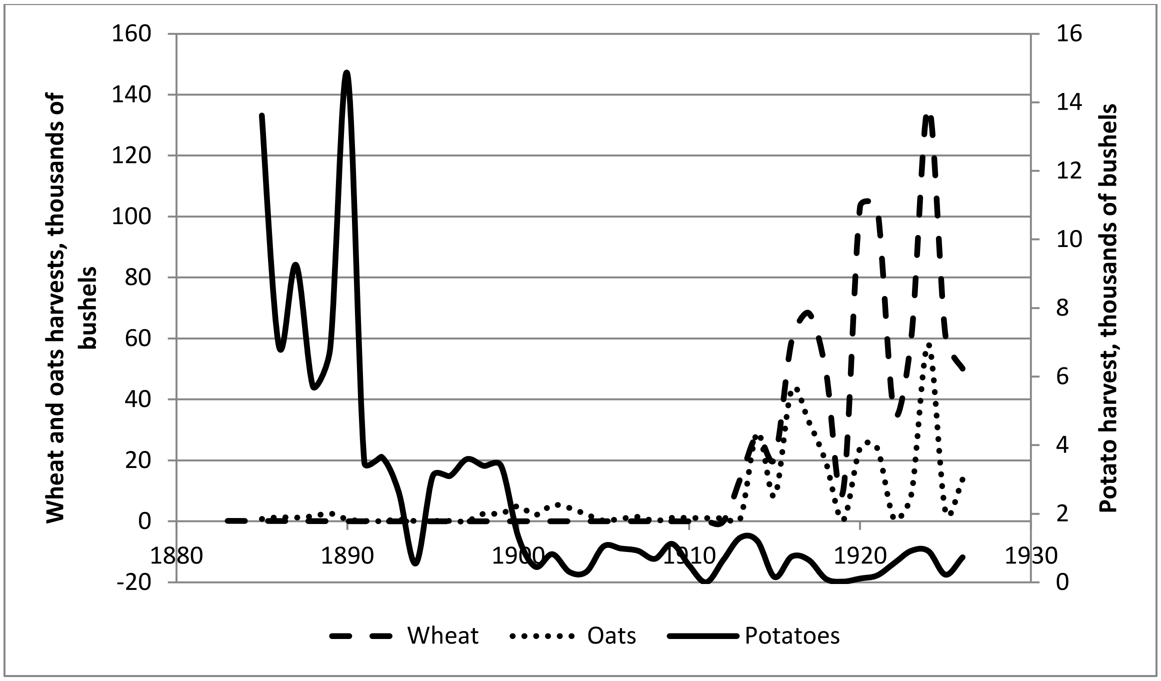

Figure 2-13: Wheat, oats, and potato harvest, Blackfoot Agency, Alberta. Source: Department of Indian Affairs (1883-1926)

Agricultural output in the Qu'Appelle agency demonstrated considerable volatility (Figure 2-12). A comparison with the Blackfoot Agency located in Alberta shows that agricultural output on the Blackfoot Agency reserve demonstrated a similar general profile as well as similar volatility. This similarity between the profiles of agricultural output in two geographically separate places suggests that they may have been driven by a similar set of factors.

2.3.3.2.4. Livestock

Like the agricultural output, residents of the Qu'Appelle Agency used to raise different types of livestock such as horses and oxen, chickens, pigs, and sheep (DIA., 1901, p.248). Similar to agricultural output, the number of cattle was fairly low until the first decade of the 20th century, when they jumped significantly (Figure 2-14).

As mentioned above, collection of dry wood and hay were providing enough income to prevent the residents of the reserve from engaging in other types of economic activity, including raising livestock. However, but by 1902 stock raising is mentioned as a principal occupation of at least one band – Muscowpetung's (DIA., 1902b, p.163). Despite the fact that stock-raising became a major industry in the first decade of the 20th century, its share in the total income was relatively small compared to agriculture and declining in importance (Figure 2-9).

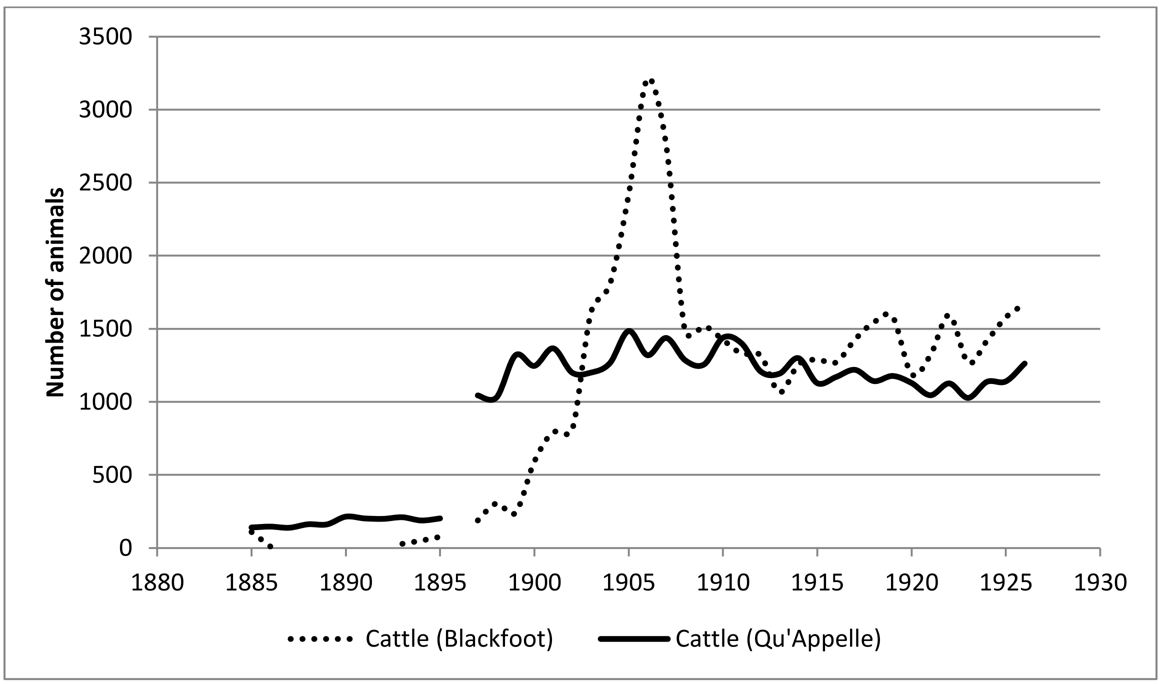

Figure 2-14: Number of cattle, Qu'Appelle and Blackfoot Agencies. Source: Department of Indian Affairs (1885-1926)

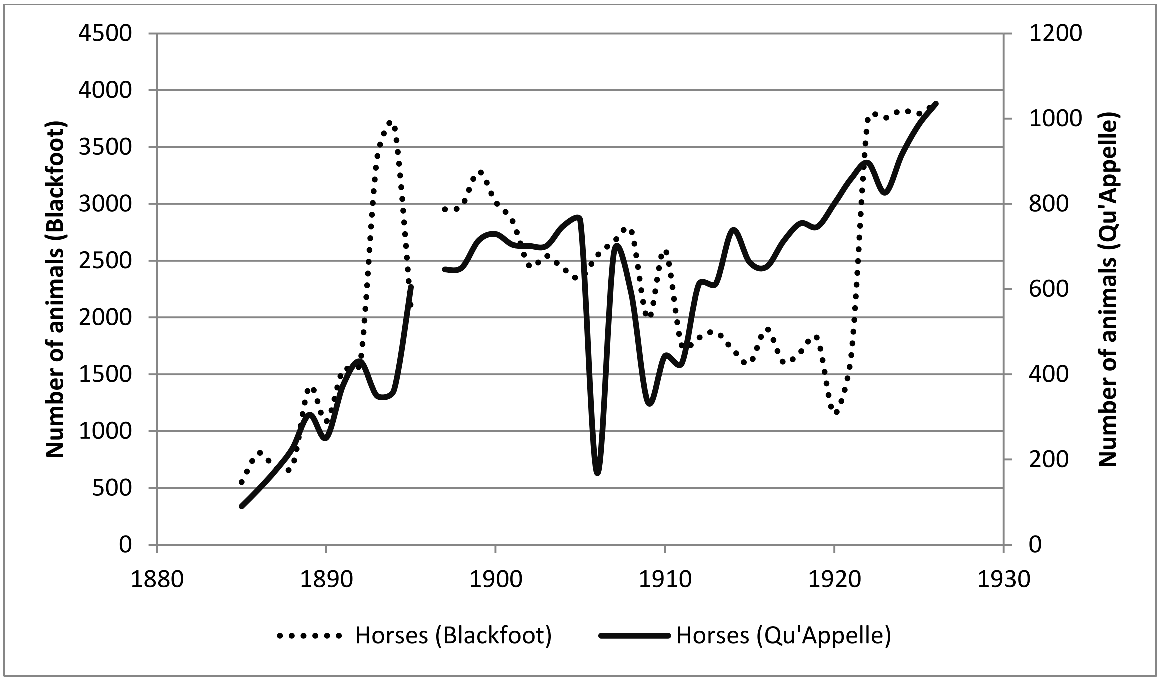

Figure 2-15: Number of horses, Qu'Appelle and Blackfoot Agencies. Source: Department of Indian Affairs (1885-1926)

2.3.3.3. Tuberculosis and the economy

Qu'Appelle Agency was located close to several towns where dry wood and hay from the Agency were sold, and where tools and other necessities could be bought (Figure 2-7). These towns were trade centres that were added to the traditional Indigenous trade network that criss-crossed the Americas. They provided opportunities for close interaction between the Indigenous population and European settlers and a source of economic benefit to the Agency. They may have had other impacts as well.

Historically, trade centres were centres of diffusion of infectious diseases through trade networks into the Indigenous societies of Canada (J. Waldram et al., 2006) (p. 62). Central nodes of these trading networks, such as the Norway House, were hard hit by infectious diseases, which were then spread by boats to other trading posts (J. Waldram et al., 2006) (p. 65).

In the early years of European settlement smallpox and flu wrecked havoc in the Indigenous societies. Sometimes the arrival of these diseases preceded the arrival of Europeans en masse.

"Decades before the arrival of the Europeans to the region, germs from distant Mesoamerica travelled along the essentially seamless route to the northwestern fringe of the Great Plains" (J. Daschuk, 2013) (p. 26)

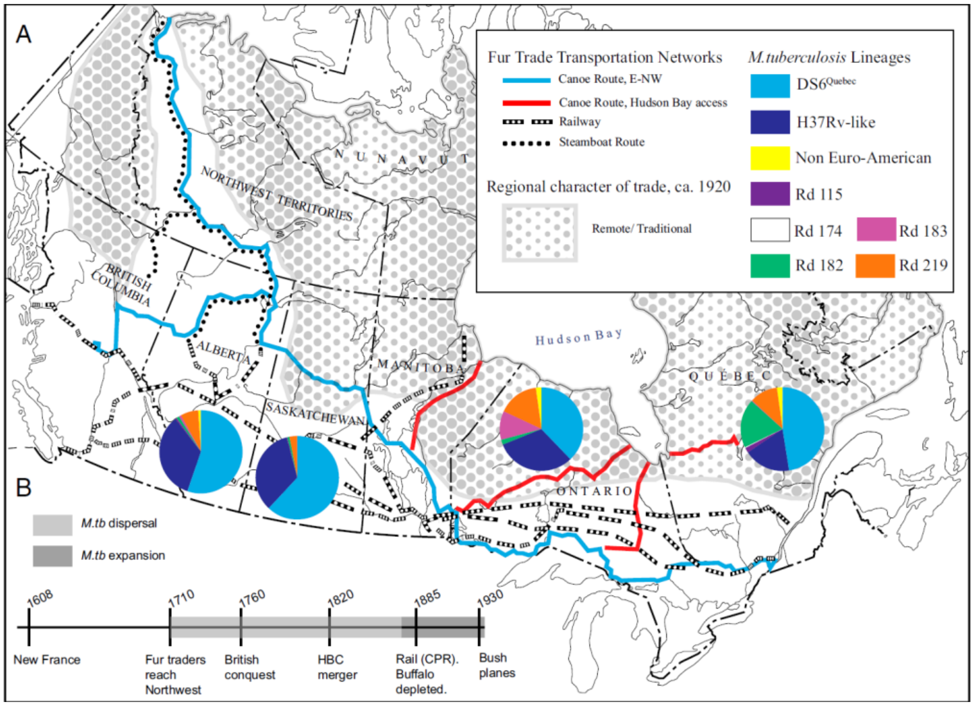

Figure 2-16: Geography of the fur trade and the spread of TB. Source: Pepperell et al., (2012)

Smallpox spread along Indigenous trade networks from Mexico City to the Great Plains (J. Daschuk, 2013) (p. 36). European colonization of Canada brought other diseases as well. A large proportion of TB strains circulating in the Plains can be traced back to Quebec from where it was brought by fur traders (Pepperell et al., 2012) (Figure 2-16).

As European settlements spread westward, they ensured continued presence of infectious diseases in close proximity to the Indigenous societies of the Prairies. While large-scale European settlement of the Prairies began only in late 19th century, recent results demonstrate that it does not take many people to spread TB: while a large proportion of TB strains circulating in the Prairies originated from Quebec, only approximately 5149 individuals migrated from Quebec to Western Canada in the 160 years of fur trade (Pepperell et al., 2012).

With time, epidemics of infectious diseases such as smallpox and whooping cough occurred with lower frequency and a transition to chronic diseases began to emerge (J. Waldram et al., 2006)

(p. 68). As smallpox was brought under control (J. Daschuk, 2013) (p. 69), TB continued to spread along the same transportation routes. Moving from east to west, TB became endemic in the east by the end of the 18th century, by the mid-1800s in Ontario, and became a serious problem in the West after the creation of reserves in the end of the 19th century (J. Waldram et al., 2006) (p. 69).

Traders were not the only people who spread the disease in the Prairies. Despite their frontier status in the 19th century, Canadian Prairies were not isolated from the rest of the world. Epidemiologic conditions in the Prairies were often affected by events that took place far away.

> "The opening of Oregon Territory to settlers in 1846 had serious medical consequences, even for population as distant as York Factory" (J. Daschuk, 2013) (p. 72)

Indigenous societies of the Prairies recognized the fact that the construction of the railway will bring a flood of European settlers which will bring diseases along (J. Daschuk, 2013) (p. 98). It is that recognition that led to the request of inclusion of the "medicine chest" into the terms of Treaty 6 (J. Daschuk, 2013) (p. 98).

Trade contributed to the spread of TB in other ways as well. Fur trade ultimately resulted in overtapping and reduction in the number of fur and game animals, leading to the dislocation of Indigenous societies that depended on these animals (J. Daschuk, 2013) (p. 61). Lack of fur animals meant that the foundations of Indigenous economy began to crumble and Indigenous societies were put under significant strain.

European settlement of the area increased the demand for meat, which put significant strain on the bison herds (J. Daschuk, 2013) (p. 61). The demand for meat had to be met, and domestic cattle were brought from the US to meet that demand. The cattle, not being used to the northern environment of the prairies, likely brought TB with them (J. Daschuk, 2013) (p. 102). Indigenous diet of sun-dried and organ meats, in this case infected with TB, may have contributed to the spread of TB in the Prairies (J. Daschuk, 2013) (p. 121).

Eventually, the bison herds disappeared altogether, leaving Indigenous societies without work or food. "As hunger spread across the plains, so did disease" (J. Daschuk, 2013) (p. 120). Faced with the famine in the Prairies the government, interested in bringing more settlers to the

Prairies, reluctantly began to provide food rations to keep peace in the Prairies. In his communication with the Hudson's Bay Company, Sir John MacDonald wrote:

> "It would be of advantage to us, & no doubt it would be of advantage to you, that we should be allowed to make use of your officers & your posts for the purpose of making those payments to the Indians which will have to be made annually by the Government of Canada in order to satisfy their claims & keep them in good humour." (Ray, 1990) (p. 5).

The government used all means to reduce expenditures on food rations. Often food rations rotted while the population starved to death (J. Daschuk, 2013) (p. 113). The cycle of famine and disease caused significant mortality in the Indigenous societies, leading *Saskatchewan Herald* to observe: "Thus the Red Man is surely and steadily passing away" (J. Daschuk, 2013) (p. 121)

2.3.4. Education

Since the focus of this research is on reconstructing TB outbreaks in the Qu'Appelle Agency, it would be important to understand how the education system may have contributed to the spread of TB. This implies the focus on immediate impacts, while understanding that the residential school system had significant lasting effects on the children. The focus of this research is on the everyday environment of the school.

Children were sent to the industrial school to learn trades, so they spent a large portion of their time there working. For example, in his 1887 annual report, principal Huggonard wrote that

> "At the present time there are four boys in the blacksmith's and four in the carpenter's shop, they work half a day in the shop and half a day in school; besides these, six boys are working with the farmer, although all the boys often work on the farm, weeding, hoeing, thinning vegetables, and doing other farm work. We also have a boy with the baker, but we cannot leave the same boy for any length of time at the trade, as the Indians seem to have some prejudice against it"(DIA., 1888, p.129).

Girls spent their time learning housekeeping. In addition to leaning trades, students were learning English and math, as well as religion (DIA., 1888, p.128). The division of time between learning trades and academic learning seemed to have favoured trades and work over academic learning, to the apparent dissatisfaction of students and parents (DIA., 1890, p.131).

Industrial schools were successful in teaching kids both trades and academic knowledge, at least to the extent that it was considered necessary at the time. Writing in 1899, Agent Mitch commented on the successes of students who returned from industrial schools (DIA., 1900, p.164).

Despite the success claimed by the Indian Agents,

> "residential schools played a significant role in engendering irreversible damages within Aboriginal society, and many of these changes [...] have a direct impact on the mental and physical health and wellbeing of Aboriginal people today" (J. Waldram et al., 2006, p.15)

The disease-spreading capacity of the boarding schools was first mentioned in 1889, in relation to an epidemic of sore eyes, which "interfered materially with the attendance at the boarding schools, as in many cases the parents kept the children at home for treatment" (DIA., 1890, p.56). In 1891, there was an outbreak of smallpox. The outbreak required quarantining and vaccinating the children (DIA., 1892, p.58). Boarding schools in the Qu'Appelle Agency also had a significant TB problem and may have amplified TB spread in the Agency. For example, among 392 school children examined by Ferguson, 38, or 9.7% had active TB (Ferguson, 1928, p.27). Incidence of new active cases among school children was 2.5% per year (Ferguson, 1928, p.28). In 1907, Peter Bryce released his report on the Indian schools of Manitoba and the North-West Territories in which he revealed that 24% of all residential school students died of TB (Sproule-Jones, 1996). Residential schools represented an important source of TB infection.

2.3.5. Healthcare

Health care interventions in the Qu'Appelle valley included doctor visits and public health campaigns. Two doctors were assigned to the Agency. Agent Lash's report mentions that

> "the doctor attended the reserve during the payment when the Indians are all on hand and vaccinated the majority in each band. The doctor again made a tour of the district in March, to see the result of his prior visit, and reported all successfully vaccinated with the exception of a few old and others who through prejudice would not be operated upon." (DIA., 1886) (p. 119)

Another doctor was reported making monthly visits to Pasquah and Standing Buffalo reserves, (DIA., 1886) (p. 119) as well as the File Hills section of the Agency, (DIA., 1888) (p. 78) and filing monthly reports on the health of the residents (DIA., 1891) (p. 41). The doctors were stationed in towns around the Agency, and visited it monthly in the beginning. By 1899, the frequency of visits was reduced to every second month, (DIA., 1900) (p. 166) and then slightly increased to once every 6 weeks in 1905 (DIA., 1906) (p. 145).

The tasks of the doctor may have included making monthly visits, providing vaccinations, and dealing with infectious disease outbreaks. This often involved quarantining the reserve, as was done in case of diphtheria outbreak on the File Hills section of the Agency (DIA., 1897.) (p. 179).

While vaccination campaigns were mentioned in the annual reports, the type of the vaccine used was not usually mentioned. Annual report for year 1899 mentioned vaccination campaign in relation to a smallpox outbreak in the US, which suggests that vaccination campaigns in the Agency may have targeted smallpox (DIA., 1900) (p. 166). Regardless of the actual type of vaccine, it could not have been Bacille Calmette-Guerin (BCG), vaccine for TB, as it was developed in 1921 (Long & Ellis, 2007, p.349).

In addition to TB, the population of the Agency suffered from other infectious diseases. Recently, a term syndemic was proposed to describe interactions among several infectious diseases, leading to compound health impacts (Singer & Clair, 2003). An example of a syndemic is a relationship between HIV and TB that accelerates the damaging effects of HIV (Ho, 1996).

Reports of Indian Agents mention interactions of several diseases, such as influenza and TB. Writing in 1890, Indian Agent stated: "Later on the influenza, so prevalent in the east, reached the reserves and proved fatal to a number suffering from lung complaints" (DIA., 1891).

Ferguson presented a timeline of epidemics in the Qu'Appelle Agency. In addition to the outbreaks mentioned, there was at least one case of smallpox in 1887 (DIA., 1888).

The nature of the health care system in the Agency continued to change over time, and a first 'cottage hospital' was established on the File Hills section of the Agency in 1915 (DIA., 1888) (p. 64). At this time, residents of the Agency were responsible for their own medical bills as well as for calling in the doctor (DIA., 1915a) (p. 64).

Table 2-2: History of infectious disease outbreaks in the Qu'Appelle Agency. Source: Ferguson (1928)

1890	Measles, Influenza
1891	Varicella
1896	Pertussis
1907	Influenza
1909	Influenza
1910	Measles
1911	Influenza
1914	Smallpox
1919	Influenza

Reports of Indian Agents do not mention TB sanatoria – one of the few TB interventions that existed at the time. A TB sanatorium in Fort Qu'Appelle was established in 1917 at which point reports of Indian Agents lacked the detail of the early years. Given the fact that the Fort Qu'Appelle sanatorium was established relatively late, as well as the absence of any reference to other sanatoria in the reports of Indian Agents, it is likely that people from the Qu'Appelle Agency were not sent outside of the reserve for treatment.

It is not clear if there were any health interventions targeting people with active pulmonary TB. There were, however, attempts to treat extrapulmonary TB. Ferguson mentioned that:

> "these physicians state that at any time during the first two decades fully one out of three of the Indians had visibly enlarged or discharging glands. In 1906 at one clinic held at the Qu'Appelle Industrial School, 40 children out of 205, or 20%, were operated on for enlarged tuberculous glands" (Ferguson, 1928) (p.15).

2.3.6. The Agent, the Host, and the Environment

TB was a serious problem in the Qu'Appelle Agency in the late 19th century (Figure 1-1). But was that always the case? The Great Plains were one of the last large geographic areas reached by the Europeans, and it occurred fairly recently. The first missionaries came to the Plains only in the 1860s, and the first settlers followed around 1878. Because there was so little contact until then, the only information on the presence of TB in the Prairies came from travellers and other occasional contacts. This evidence was summarized by Ferguson (1928):

> "From the accumulated evidence of those who knew the Plains Indians prior to the disappearance of the buffalo from their territory in 1879, it is certain that there were

> sporadic cases of tuberculosis, and equally certain that it was uncommon and unimportant as a cause of death" (Ferguson, 1928, p.8).

With the establishment of the Qu'Appelle Agency, Indian Agents began collecting TB data. As Figure 1-1demonstrates, TB mortality peaked early, reaching nearly 100 deaths per 1000 people in 1886, and then began a decline. Decline in mortality rates occurred amid growing concerns over TB as a cause of death as is evidenced by reports of the Indian Agents.

Annual reports in the early years of the Agency tended to be very positive, with statements such as: "The general health of the Indians has been very good; they are visited monthly by Dr. Seymour, the medical officer"(DIA., 1892, p.54). Then the tone began to change. First, TB was mentioned as something normal and expected. Writing in 1901, Agent Graham says:

> "The general health of the Indians of this band for the year has been fairly good, considering that quite a few of them are afflicted with pulmonary and scrofulous diseases"(DIA., 1902b, p.153).

Later on, the Agents treated it as something more serious, as is evidenced in this report:

> "These Indians are not nearly so healthy as the Piapot Indians. Many of the older members are affected with consumption, scrofula and sore eyes" (DIA., 1906b).

The first phase of the epidemic (1885 – 1900) (Figure 1-1) was characterised by a very high prevalence of glandular TB and high speed of disease progression. Physicians who worked on reserve estimated the prevalence of glandular TB in the population to be about 30% (Ferguson, 1928, p.15). The speed of disease progression was so high that infected individuals could die within several weeks or several months of disease onset, before emaciation typical of TB could develop (Ferguson, 1928, p.15).

2.3.6.1. Morbidity

In his study of the Qu'Appelle outbreak, Ferguson performed chest x-rays on people belonging to the first three generations initially affected by the outbreak and established that 22% of people aged over 50 had healed lesions. Among people aged 25 to 50 years 15% had healed lesions, and among school children aged 6 – 18 years, 9.4% had active, and 9.9% had latent TB (Ferguson, 1928, p.24).

Incidence of active TB in school-age children increased significantly after the age of 9 and reached its peak at the age of 12, dropping significantly after that (Figure 2-17). Ferguson conducted tuberculin sensitivity testing among school children and found that tuberculin sensitivity was about 50% at the age of 6, increasing to 100% by the age of 15 (Figure 2-18).

In addition to school children, Ferguson studied other subsets of population, such as differences in mortality between the "full-blood Indians [...] compared with the mortality of those who have more or less white blood on the same reserve"(Ferguson, 1928, p.24). Mortality of children with mixed ancestry was 12.5%, compared to 19.9% among those with no mixed ancestry.

When it comes to the sources of the disease, Ferguson tested 27 mycobacterial samples and concluded that a human strain of TB was responsible for the outbreak (Ferguson, 1928, p.30).

Data on TB morbidity off-reserve began to be collected later. Earliest TB admission to hospital data go back to late 1930s (Dominion Bureau of Statistics, 1939).

TB mortality did not differ across age groups enough to suggest that younger generations were developing resistance. Given that the three generations were alive at the beginning of the outbreak in 1885, it was not likely that the youngest generation could develop resistance.

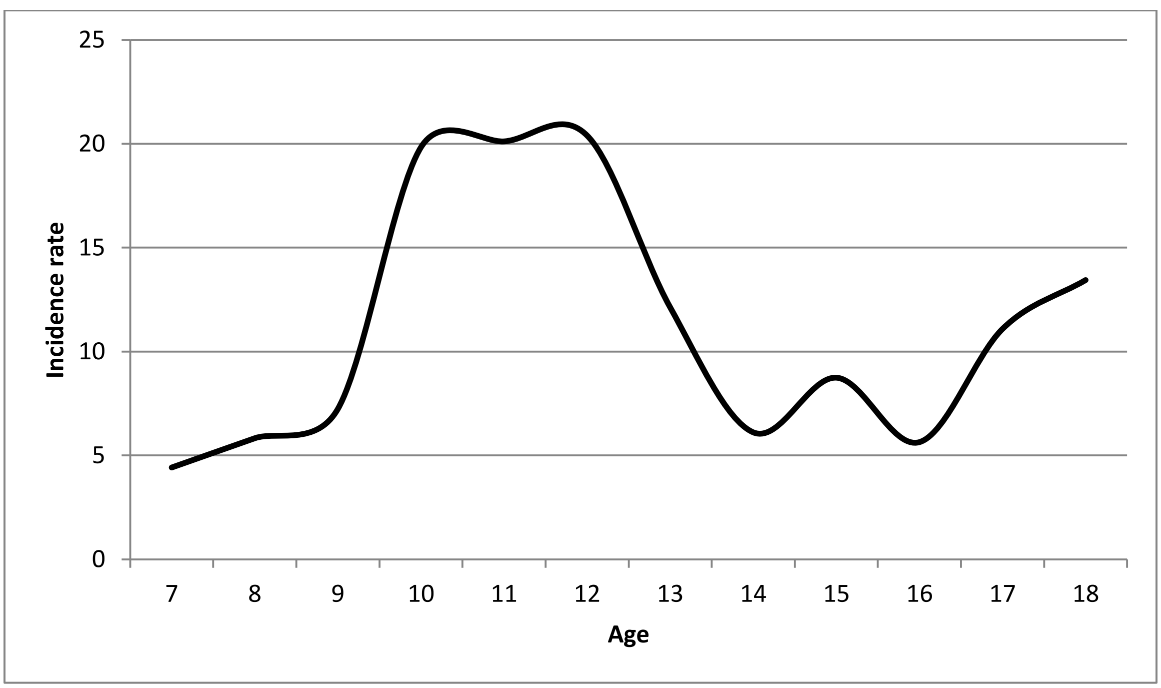

Figure 2-17: Incidence of active TB among schoolchildren. Source: Ferguson(1928)

Figure 2-18: Tuberculin sensitivity by age, percent. Source: Ferguson(1928)

2.3.6.2. Mortality

TB mortality varied by age: at almost 17 deaths per 1000 among 10-14 year olds, TB mortality rate was highest in children. It fluctuated around 16 deaths per 1000 and then dropped quickly after the age of 30 (Ferguson, 1928, p.13). Age distribution of TB mortality also varied by sex: TB mortality was lower in male children, but higher in adults (Figure 2-19).

Ferguson studied mortality across three consecutive generations defined as those who were adults, children, and children of those who were children in 1886. Mortality rate in the first, second, and third generation was 21%, 28.2%, and 19.2%, respectively (Ferguson, 1928, p.18). Since generation-specific mortality did not decline significantly, reduction in TB mortality evident from Figure 1-1 could be due to the fact that in the initial years of the outbreak, susceptible individuals of different ages were affected at the same time. As the number of susceptible individuals was reduced, subsequent generations accounted for the largest part of deaths from TB (Ferguson, 1928, p.18).

Given the high mortality in the early years of the outbreak, several genealogical families completely died out. Comparing families that died out because of TB as well as other causes with those that survived, Ferguson estimated TB mortality rate of 31% among those that died out, and 21% among those that survived (Ferguson, 1928, p.16). Only one family in the Qu'Appelle Agency did not have a single case of TB in the three generations examined.

Decline in TB mortality stopped around 1910 and then fluctuated between 9 and 21 deaths per 1000 per year (Figure 1-1) until the end of the study period in 1926. Comparable mortality data off-reserve may not be available. The oldest mortality data go back only to early 1920s. In 1921, at 0.34 per thousand, Saskatchewan had the lowest TB mortality rate in the country (Dominion Bureau of Statistics, 1939, p. 111).

The TB outbreak in Qu'Appelle Agency was associated with significant mortality, compared to TB mortality in the non-Indigenous population of Saskatchewan (Ferguson, 1928). Some argued it was because of their susceptibility, and lack of exposure to European strains of TB (S. Grzybowski & Allen, 1999). While currently some Indigenous populations have elevated susceptibility to TB (Larcombe et al., 2015) which must have been higher in the past, the Indigenous population of the Prairies was exposed to European strains of TB (Pepperell et al., 2012) without suffering significant morbidity or mortality from TB until the 1870s (Ferguson, 1928). This suggests that the relationship between the agent and the host alone cannot explain the outbreak, and that environment must have played a significant role.

Figure 2-19: TB mortality per 1000, by age and sex. Source: Ferguson (1928)

The hierarchical relationship between social determinants of health, risk factors, and TB infection was summarized by Lonnroth et al., (2009) (Figure 2-20). While this framework was not developed for the realities of the 19th century Canadian Prairies and therefore does not include complete destruction of social, political, and economic structures of the Indigenous societies in the Prairies, it provides some direction for understanding the reasons for the outbreak.

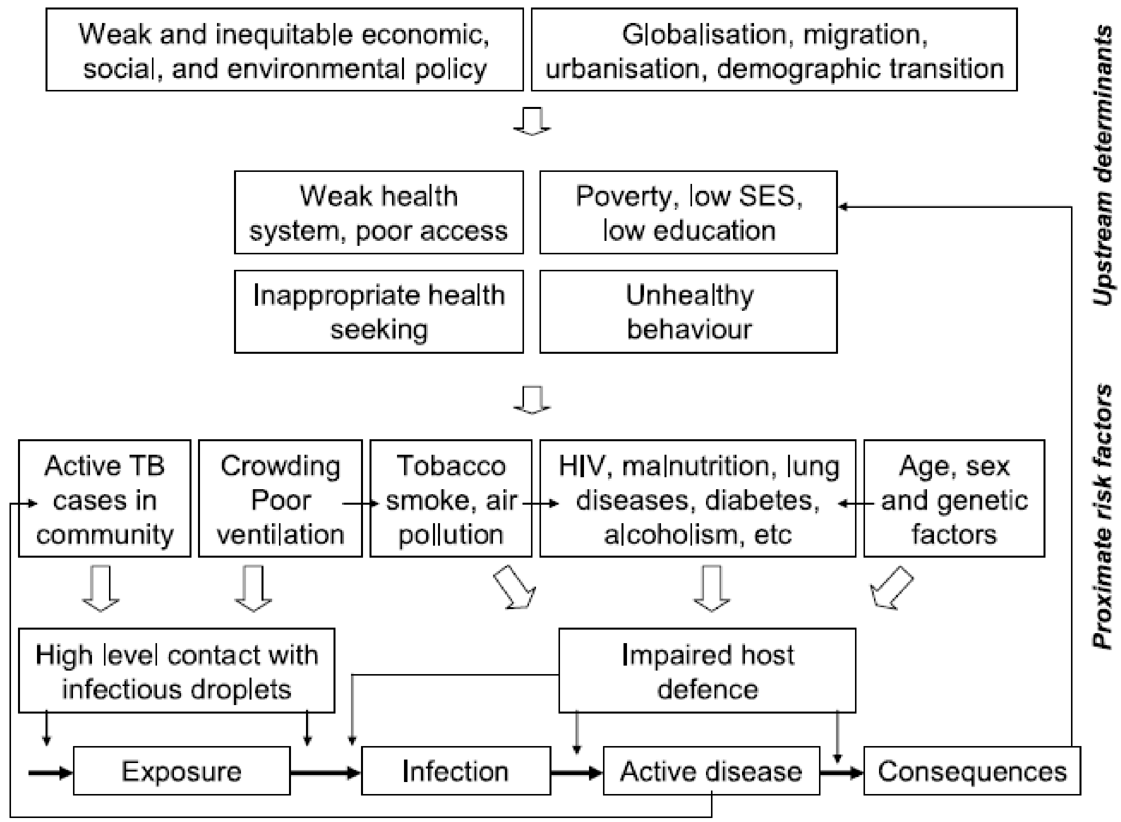

Figure 2-20: Risk factors and social determinants of TB. Source: Lonnroth et al., (2009)

Upstream determinants include "weak and inequitable social and environmental policy". Unfortunately, in the case of the Indigenous societies of the Prairies, social and environmental policies were not weak and inequitable, but outright destructive. Extermination of the buffalo and overtrapping of fur animals resulted in the destruction of the economic basis of Indigenous societies and widespread hunger on the Prairies. Hunger, a proximate risk factor (Figure 2-20), was not addressed effectively by the government interested in cutting costs, and as people starved to death, food supplies rotted away (J. Daschuk, 2013) (p. 113).

The second element of the framework – globalisation, migration, urbanisation, and demographic transition was also present in the Prairies. Globalisation was a result of colonization of Indigenous societies – a fundamental underlying determinant of health (WHO, 2007). Colonization was manifested in many ways, such as the loss of land and sovereignty, settlement on reserves, and migration of European settlers. Urbanisation directly involved Indigenous societies moving from semi-nomadic to a sedentary lifestyle, and the demographic transition was evident in the rapid change of the makeup of the population of the Prairies.

Of the next set of blocks of the framework, poverty, low SES, and low education is the one that was most obviously present in the Qu'Appelle Agency, particularly in the early years when agricultural output was low. These determinants of TB are also important determinants of Indigenous health, with SES being a major determinant (Gracey & King, 2009). As a direct consequence of colonization, Indigenous communities were denied opportunities to maximize their SES and have equitable access to educational opportunities (Loppie Reading, 2009). Food insecurity is also a continuing problem (Gracey & King, 2009).

The other three blocks on this level may not apply as well. The Agency had a doctor who was doing monthly visits. That, in itself, does not mean that people of the Agency had good access or good quality care. Since there was neither treatment nor vaccination available for TB at that time, healthcare quality should be evaluated from the point of view of addressing diseases other than TB. As was mentioned above, doctors performed vaccinations and diseases such as smallpox and influenza that used to wreck havoc in Indigenous societies were not as big a problem in the Qu'Appelle Agency. When it comes to health in general, doctors could do little as most of the problems lay in the poor state of SDH – malnutrition and poor housing.

Inappropriate health seeking and unhealthy behaviours may have been a factor to some extent, but very minor factors at best. Inappropriate health seeking is not very relevant to the consideration of TB at that particular time, as there was no appropriate treatment. Unhealthy behaviours such as drinking and lack of sanitation were being addressed by the Indian Agents. At a time when societal destruction, malnutrition, colonization, and poor housing were a reality, these two blocks may have been secondary at best.

The third level of the framework includes a number of proximate risk factors (Figure 2-20). Most of these factors were discussed in the sections above. People of the Agency likely had genetic susceptibility to TB (Section 2.22). Malnutrition was a problem, particularly after the extermination of the buffalo (Section 2.3.3). Indoor air pollution was a factor in the early houses built in the Agency. These houses lacked both chimneys and windows, with food cooked on the open fire (Section 2.3.2). Available data make it difficult to estimate the prevalence of crowding.

The number of people per house was fairly low (Section 2.3.2), but that number alone does not give a clear idea of crowding given that the size of houses is unknown. Finally, TB appeared in the Prairies after 1860 and became more of a presence in the 1870s. Some of these factors, such as indoor air pollution and malnutrition are determinants of Indigenous health that were a factor in the Qu'Appelle Agency and still remain a problem today (Loppie Reading, 2009).

The presence of these factors impaired host defences of people that may have been genetically predisposed to TB (Figure 2-20) and created conditions in which contact with people with active cases of TB was possible, leading to TB transmission in the Agency.

2.4. Mathematical modelling

One of the research questions raised in this thesis is the reconstruction of TB outbreaks in the Qu'Appelle Agency. The most logical way to attempt such a reconstruction is by using mathematical modelling.

Models of infectious diseases have a long history. One of the early attempts to develop a mathematical model of an infectious disease outbreak was undertaken by Kermack and McKendrick almost a century ago. Kermack and McKendrick developed a model of an epidemic "with a view to obtaining more insight regarding the effects of the various factors which govern the spread of contagious epidemics" (Kermack & McKendrick, 1927). Kermack and McKendrick published two follow-up papers developing their model further. The Kermack and McKendrick model is a compartmental differential-equation model that accounts for infection and removal of individuals through recovery or death. In its original form, the Kermack-McKendrick model assumed fixed homogeneous population without individual characteristics such as age, location, or social structure. Despite its simplicity, the Kermack-McKendrick model is important because it laid the foundations for the basic models of epidemiology, such as the Susceptible-Infected-Recovered (SIR), or Susceptible-Infected-Susceptible (SIS) model.

These models have a number of assumptions similar to those of the Kermack-McKendrick model, such as fixed population, random mixing, latent period of zero, and constant disease transmission probability (Giesecke, 2002), (Hethcote, 1989). These assumptions simplify the models and make them mathematically tractable. At the same time, they assume away factors important for the spread of infectious diseases, such as contact networks and human behaviour.

The desire to release some of these assumptions led to the development of more complex models, such as microsimulation models.

> "Microsimulation models are computer models that operate at the level of the individual behavioural entity, such as a person, family, or firm. Such models simulate large representative populations of these low-level entities in order to draw conclusions that apply to higher levels of aggregation such as an entire country" (Statistics Canada, 2015)

By using micro-level data, microsimulation models add agent heterogeneity to the toolbox of a modeller of infectious diseases. Microsimulation models have been widely used in policy analysis because of their ability to generate micro-level data (Zaidi,A., Rake, K., 2001). A number of microsimulation models have been developed over the years (Zaidi,A., Rake, K., 2001).

Microsimulation models fall into two categories: static and dynamic (Li et al., 2011). Static models are used to evaluate immediate distributional impacts of policy changes, while dynamic models allow individuals to change their characteristics in response to changes that take place within the model (Li et al., 2011). In other words, dynamic microsimulation models add a behavioural dimension to the model (Zaidi,A., Rake, K., 2001).

Addition of a behavioural dimension is achieved using two methods: static and dynamic ageing (Zaidi,A., Rake, K., 2001). Static ageing is a parameter updating process that involves re-weighting model micro data to align it with an external source (Zaidi,A., Rake, K., 2001). Dynamic ageing involves modelling changes in personal characteristics of interest, such as labour participation rate, as a stochastic process (Zaidi,A., Rake, K., 2001). While the stochastic process of updating individual characteristics is called behavioural, it is not behavioural in a sense of game theory with rules of behaviour such as: "If you do X, I do Y." Further refinement of behavioural aspects of the models brought in a new class of models – agent-based models (ABM).

ABMs are a tool of choice of generative social science (GSS), a field of social science that explains social phenomena by generating them. GSS uses ABMs to answer the central question tackled by the science: “How could the decentralized local interactions of heterogeneous autonomous agents generate the given regularity?” (J. Epstein, 2006) (p. 5)

An ABM

> “consists of a set of agents that encapsulate the behaviours of the various individuals that make up the system, and execution consists of emulating these behaviours” (Parunak, Savit, & Riolo, 1998).

The concept of an agent is key to defining the ABM. An agent is an entity that possesses several characteristics, such as (Bandini, Manzoni, & Vizzari, 2009):

- Ability to operate without human interventions
- Ability to interact
- Ability to react and respond to changes in environment
- Ability to act according to internal rules of behaviour

An agent can be any entity that has these characteristics, whether it is defined as an individual (J. A. Epstein & Axtell, 1996), a group of individuals in a family unit (Gumerman, Swedlund, Dean, & Epstein, 2003), or something different, such as a molecule (Troisi, Wong, & Ratner, 2005).

The GSS approach to the study of social phenomena can be summarized as follows: with a macro-level phenomenon of interest identified, an ABM with heterogeneous agents, rules of behaviour, and explicitly-defined environment is developed and its aggregate output evaluated and compared to the macro-level phenomenon of interest. In case of a fit between the simulated output of the model and actual historical data, the model is said to provide a candidate explanation of the macro-level phenomenon of interest.

The GSS approach of generating a macro structure in order to explain it raises the question of whether generation of the social phenomenon of interest suffices as an explanation. Generative sufficiency is a necessary, but not sufficient condition for explanation of the social phenomenon in question (J. Epstein, 2006) (p. 3). Intuitively, this point is clear: a micro-structure simulated by an ABM may generate the social phenomenon of interest becoming a candidate explanation, but it may be a weak candidate (J. Epstein, 2006) (p. 9). An example can illustrate this point. The documented TB outbreak in the Qu’Appelle Agency began in 1885. An ABM of government agents infecting people of the Agency with TB can have a perfect fit to the actual historically observed TB mortality in the Agency. Yet, that explanation will be a weak one as it is not supported by historical evidence.

ABMs have been criticised for lacking explicit equations, not being derived from first principles such as utility maximisation in microeconomics, and lacking generality (J. Epstein, 2006) (p. 4). At the same time, they are dynamic, better suited to representing social interaction than traditional mathematical models, and they produce large amounts of data that can be analysed to describe performance of the model over ranges of parameter values. More importantly, they allow for a link between micro-level behaviour and macro-level outcomes that allow for a reconstruction of TB outbreaks in the Qu'Appelle Agency.

2.4.1. Characteristics of agent-based models

Compared to other models, ABMs have a number of unique characteristics. As noted by (J. Epstein, 2006) (p. 6), these characteristics include

- Heterogeneity: agents used in ABMs may be heterogeneous in a multitude of ways, depending on the question at hand. Agents may differ genetically, culturally, by age and sex, location, and network. All these characteristics can change endogenously with time.
- Autonomy: ABMs are characterised by the absence of top-down control of agent behaviour. Instead, there are feedbacks between macro and micro levels of the model that affect behaviour. Prevalence of a disease is a good example – it affects agents that are born into the model, and it is created endogenously over previous iterations of the model.
- Explicit space: GSS focuses on local interactions between agents, which requires specification of the space in which interaction takes place. In their seminal 1996 work Axtel and Epstein (1996) specified a landscape of renewable resources where agents interacted to generate various social and disease patterns (J. A. Epstein & Axtell, 1996).
- Local interactions: unlike economic models that use uniform mixing such as molecules of air in a room, ABMs rely on local interactions between the agents and environment.
- Bounded rationality: a key feature of GSS models, requirement of bounded rationality imposes limitation on the amount of information and computing power agents have. Agents have only local information, and interact on the basis of simple behavioural rules.

These characteristics of ABMs set them apart from other approaches to modelling infectious diseases and fit the requirements set out in the introduction. ABMs bring heterogeneity – a key factor to understanding TB dynamics in the Agency. Agent heterogeneity is necessary to incorporate such parameters as susceptibility to TB, age and sex distributions, differences in occupations, and differences in decision-making styles. ABMs emphasise the importance of explicitly defining space. In case of the Qu'Appelle Agency, that means defining productive capacity of the environment in order to model agricultural output, as well as defining geographic location of homes and fields in order to model interactions. Interactions in the ABMs are local and explicitly defined, which is important when modelling spread of infectious diseases across networks of contacts and evaluating the impact of settlement patterns on disease spread.

ABMs make it possible to use available data to relive the transition of the bands of the Agency from buffalo hunters to farmers and ranchers, from semi-nomadic to settled life style, and from the state when TB was unknown among them, to the one in which it was highly prevalent. In fact, ABMs have been used to study past societies and their economies, as well as infectious diseases – two of the many uses of ABMs that are of direct relevance to the current research. The following section reviews some of the relevant applications of ABMs.

2.4.2. Application of agent-based models

ABMs have been applied to study various aspects of TB at the micro and meso levels. The following sections review some of these studies.

2.4.2.1. Micro-level studies

A number of recent papers used ABMs to study various aspects of TB at the cellular level. Segovia-Juarez, Ganguli, and Kirschner (2004) studied granuloma formation and function in the lung of a person infected with TB (Segovia-Juarez, Ganguli, & Kirschner, 2004). In this paper, granuloma formation was represented as a complex system involving interactions of bacteria, immune cells and immune effectors. The ABM had three major components: (1) environment representing a section of lung tissue, (2) entities or agents (macrophage and T cell agents, chemokine and bacteria variables), (3) rules of behaviour, and (4) time. The authors concluded that an ABM was an appropriate tool for exploring such a complex spatio-temporal system.

In a recent paper, Marino, El-Kebir, and Kirschner extended the model of granuloma formation by developing a hybrid model that used an ABM to describe granuloma formation in the lung, and ordinary differential equations to model lymph node function (Marino et al., 2011).

El-Kebir et. al, (2013) used an ABM to study the role of tumor necrosis factor-α in granuloma formation in the brain of a person with TB meningitis (El-Kebir et al., 2013). The ABM had four basic components: (1) environment (a cross-section of cerebral cortex), (2) model agents (*M.tuberculosis* and phagocytes), (3) rules of behaviour, and (4) time scales.

Cilfone et. al, (2013) studied role of tumour necrosis factor-α and interleukin-10 in the immune response to TB infection (Cilfone et al., 2013). The model incorporated three scales: tissue, cell, and molecule, where tissue and cell dynamics are described using an ABM, while molecular dynamics were described by differential equations. The ABM included several agents (immune cells and bacteria), rules of iteration, as well as the explicitly-defined environment.

2.4.2.2. Meso-level studies

While the micro-level studied TB at the tissue, cellular, and molecular level, the following two studies defined ABMs at the individual and household level.

Bishai et.al, (2010) studied spread of multi-drug resistant TB (Bishai, Bishai, & Bishai, 2010). The model used individually-defined agents "interacting with their neighbours in space and time following simple rules that the individual agents follow probabilistically" (Bishai et al., 2010). Characteristics of the agents were not defined, and may be limited to TB status only. They have no age or sex, which, according to the authors, made the model impossible to validate (Bishai et al., 2010). The model included neither a description of the environment, nor rules of behaviour – key elements of an ABM (J. Epstein, 2006) (p. 6).

Tian et.al, (2013) evaluated effectiveness of contact tracing in improving TB outcomes in Saskatchewan (Tian, Osgood, Al-Azem, & Hoeppner, 2013). The model had three explicitly defined elements: (1) population, (2) environment, and (3) rules of behaviour, where population represented a population of a hypothetical Indigenous community, environment was a simulated contact network, and rules of behaviour appear to refer to transitions between different disease states. While the authors referred to the model as an ABM, it is not clear that the agents have all the required attributes mentioned above, such as ability to act according to rules, ability to react, interact, and operate without human intervention (Bandini et al., 2009). In addition to this limitation, while environment is defined, it lacks data support.

Kasaie et.al, (2014) studied the impact of household contact tracing on the timing of TB transmission using an ABM with agents defined at the level of a household (Kasaie, Andrews, Kelton, & Dowdy, 2014). Agents had some explicit characteristics, such as household size, contact and interaction probabilities, but neither environment, nor rules of behaviour were explicitly defined. While the authors refer to their model as an ABM, like the model of Tian et.al, (2013) it lacks several key characteristics of an ABM.

While previous papers applied ABMs to the study of TB, the following paper applied an ABM to other factors relevant to the study of TB in the Qu'Appelle Agency, such as cultural change, social behaviour, and population dynamics of the Anasazi, a civilization that thrived in the present day Arizona between CE 800 and CE 1350, and then abruptly disappeared for a reason that is not entirely clear (J. Epstein, 2006) (p. 88).

The goal of the study was to recreate the actual spatiotemporal history of the settlement, development, and abandonment of the Long House Valley. The actual archaeological record of economic, settlement, social, and religious conditions provided an empirical data set which could be compared with simulated behaviour of the agents in the model.

The Anasazi model had some of the key features of the ABM, such as:

- explicitly defined landscape of annual variations in potential maize production values across seven environmental zones (Dean, Gumerman, Epstein, Axtell, & Swedlund, 2006)
- heterogenous agents with characteristics such as age, location, and stock of grain, death age and nutritional need (Dean et al., 2006)
- explicit behavioural rules based on detailed ethnographies available for the area

Initial specification of the model diverged quantitatively from the actual archaeological record, but qualitatively reproduced many important dynamics, such as the distribution of the population across settlement clusters, and population dynamics and changes in the size of settlement clusters in response to changing environmental conditions.

While a number of studies attempted to apply an ABM to various aspects of TB, it appears that only the micro-level studies carried out at the level of tissues, cells, and molecules have all characteristics of an ABM, namely, heterogenous agents, explicitly-defined environment, and rules of behaviour. Studies carried out at the individual and household level may have some characteristics, such as heterogenous agents, but lack the other two characteristics.

Among the three meso level studies, no model had an explicitly-defined environment like the one used in the Anasazi study. In studies where environment was defined in some way, it lacked the detail and empirical basis of the Anasazi and micro-level TB studies. Meso-level studies lacked defined rules of behaviour. While authors mentioned the existence of such rules, these rules were not explicitly stated, which raised serious doubts regarding the ability of these models to meet agency requirements described by Bandini et.al, (2009).

Ideally, an ABM would have all of the characteristics mentioned above. Developing such a model is a serious challenge, as is evidenced by the fact that not many ABMs have them. In addition, there is an issue of how well it could be done given data and other limitations. In developing the ABM of TB outbreaks in the Qu'Appelle Agency, the objective was to include all of the characteristics of an ideal ABM mentioned above.

2.5. Summary

The preceding sections described the three elements of the Epidemiologic Triad – the agent - TB, the host - people of the Qu'Appelle Agency, and the environment - socio-economic conditions that characterized the Qu'Appelle Agency between the years 1885 and 1926.

TB is an infectious disease most often affecting the lungs, but other organs as well. It is caused by bacterium *Mycobacterium tuberculosis*. *M.tuberculosis* complex includes several human, and one bovine strain of mycobacteria. TB is transmitted by exposure to bacteria contained in airborne droplets produced during coughing by people with pulmonary or laryngeal disease (Connolly, 2005).

TB existed in North America prior to the arrival of Europeans. Documented TB outbreaks in Indigenous communities were characterised by high TB mortality, which could be explained by changes in the SDH associated with the arrival of Europeans, or by genetic predisposition to TB because of lack of prior exposure of the Indigenous population.

One of the best-documented TB outbreaks took place in the Qu'Appelle Agency. TB was most likely present in the population of the Agency before its creation, but did not cause significant morbidity or mortality. With the establishment of the reserve and the transformation of its socio-economic environment, TB mortality quickly increased and then began to decline. The beginning of the documented outbreak corresponded to the time when the main staple of diet of the Plains Indians, the buffalo, was already gone, but the new economy based on farming and ranching was not yet established. TB mortality began to decline without any TB-specific health interventions. At the same time, a number of indicators of social determinants of health improved.

Adoption of the new agriculture-based economy was slow. It appears that only a few individuals initially took up farming, while the majority preferred to earn livelihood by more familiar means of collecting dry wood and cutting hay. As the supply of dry wood became exhausted, farming and ranching became more widely adopted and, as a result, most indicators of economic activity on reserve began an exponential increase starting in early 20th century.

Residents of the Agency initially lived in poor-quality houses organized in villages. Under pressure from the Indian Agents the village system was abandoned for houses located on individual farm fields. Over time the quality of the houses improved from houses with earth floors, no windows and open fires, to houses with wood floors, windows, and fireplaces.

Social determinants of TB and risk factors associated with it include poverty, income inequality, unemployment, low educational attainment, social and economic disruptions, smoking, alcohol consumption, malnutrition, low BMI, and indoor air pollution, among others.

These examples demonstrate that the social determinants of health impact the host, the agent, and the environment of the Epidemiologic Triad. Given this nature of the SDH, approaches that attempt to divide them into neat boxes may not be as useful as dynamic approaches that stress complexity and feedback loops, such as the Generative Social Science.

GSS is a suitable methodology to study TB outbreak in the Qu'Appelle Agency because ABMs are dynamic, include heterogeneous agents and explicitly defined environment. These models have been successfully used to study complex spatio-temporal problems like the immune response to TB infection and granuloma formation in lungs and brain, and socioeconomic and demographic dynamics in the Long House Valley.

ABM would produce output that could be compared to the actual aggregate TB mortality in the Qu'Appelle Agency. By linking individual characteristics of the artificial population of the model with its environment and *M.tuberculosis*, such an ABM would help study the links between the three elements of the Epidemiologic Triad and take a step towards reconstructing the TB outbreak of the Qu'Appelle Agency.

3. Methodology

Generative Social Science (GSS) and its tool of choice – the agent-based model (ABM) are used to reconstruct and study the TB outbreak in the Qu'Appelle Agency. This methodology reconstructs relationships between the agent, the host, and the environment in a way that allows all elements of the Epidemiologic Triad to be explicitly defined and placed in a dynamic relationship with each other.

ABMs require micro-level data in order to describe heterogeneous agents and their interactions. However, micro-level data are not available for the Qu'Appelle Agency. One of the advantages of the ABM is that it can use macro-level data to estimate micro-level data. Given that useful feature, ABM is the main methodological approach to the study of TB in the Qu'Appelle Agency. Regression analyses were used to analyse aggregate data in order to evaluate the relationship between TB mortality and social determinants of health at the level of the Agency.

The following sections describe the formulation of the ABM, supporting regression analyses used to derive the ABM, the data used to develop the model, and the research design.

3.1. Aggregate Data

The DIA collected, compiled, and published a significant amount of data on socioeconomic characteristics of the Qu'Appelle Agency (Appendix B). These data were published in the department's Annual Report, which will be the main source of data on social determinants of health of the Indigenous population of the Qu'Appelle Agency.

Wheat and oats harvested in the Agency become a source of income as well as a direct source of food. Oxen and horses were used as work animals, sold, and in case of cows, used for food. Economic factors of production, such as mowers and ploughs, and other types of property, are indirect indicators of socioeconomic status.

3.2. Data quality

Quality of data collected by the DIA is an important issue, particularly in light of the significant variability of data presented in Chapter 1 in Figure 2-5 and Figure 2-14. There are at least three potential sources of problems with the DIA data, namely, the worldview of the Agents who collected the data, self-interest of the DIA in collecting data, and the lack of standardised definitions to guide data collection efforts.

Table 3-1: Name, definition, and sources of data

Source	Sub-Section	Variables	Years available
Annual Report of the Department of Indian Affairs	Census of Indians	Population size	1885-1917, 1924
		Population by age	1901-1916, 1924
		Population by sex	1901-1916, 1924
	Agricultural and industrial statistics	Total income from all sources	1897-1926
	Agriculture	Area farmed	1885-1895, 1899-1920, 1922-1926
		Hay harvest	1885, 1886, 1888-1895, 1898- 1922, 1925, 1926
		Wheat harvest in bushels	1885-1895, 1898-1922, 1925,1926
		Oats harvest in bushels	1887-1926
		Potato harvest in bushels	1885-1895, 1898-1926
	Live stock and poultry	Number of oxen	1885-1916
		Number of horses	1885-1895, 1897-1926
	Private fencing and buildings	Number of houses	1885-1895, 1897-1926
		Number of stables	1885-1915
	Personality of Indians	Number of ploughs	1888-1926
		Number of mowers	1894-1902, 1904-1926
		Total value of property	1897-1926
Ferguson (1928)		TB mortality rate	1885-1926
		Total mortality rate	1885-1926
		TB prevalence	1928
		Disease progression	
		General information on the TB outbreak in the Qu'Appelle Agency	

Environment Canada	Historical weather	Daily temperatures Precipitation Heat days	1885-1888 and 1902-1926

DIA data collection efforts were criticised as ethnocentric (Blaine, 1991) . Blaine argued that data collection efforts represented “Western valuation of Indigenous goods, services, and time, which is inherently inappropriate given cultural differences between the two groups” (Blaine, 1991). Tough (1992) argued that this approach was not helpful, as it dismissed out of hand a very rich source of quantitative data.

The second potential source of problems with the data collected by the DIA is that the department was not a disinterested party in this process. The government wanted to show it was succeeding in managing Indigenous affairs, and in at least one instance DIA falsified the data to demonstrate success (J. H. Daschuk, Hackett, & McNeil, 2006)(p. 136).

The third potential source of problems comes from the fact that DIA did not seem to have clear definitions for some types of data, such as income data, meaning that these data may be approximate at best (Burrows, 1986). Despite these problems, DIA data are almost an exclusive source of data on the Indigenous population (Burrows, 1986).

The quality and lengths of data sets collected by the DIA differ significantly. The data were probably collected by the department for internal reasons of accountability and effective management of the reserves. Because of that, they may reflect changes in priorities of the department over time from taking care of the most pressing issues, such as issuing rations, to more in-depth understanding of economic and social processes taking place on reserve. The end result is that the longest data series are available for such information as the population size and the religious makeup of the population. The shortness of some data series, such as the break-up of income by source of income and the total value of personal property, may also reflect the complexity of the accounting system necessary to collect such data (Table 3-1).

The three potential sources of problems with DIA data as well as the volatility of some of the data series present a challenge. Is some of the volatility real or could it be explained by the problems with the data? For this reason the data were evaluated for the presence of outliers.

Some data series have missing data points. In order to increase the number of available observations for regression analyses, in cases in which missing data points appeared between available data points, missing data were estimated by linear interpolation. This way data on population size (years 1918-1923), wheat harvest (years 1896, 1897, 1923, 1924), potato harvest (1896, 1897), number of horses (1896), number of houses (1896), hay cut (1896, 1897, 1923, 1924), and area farmed (1896-1898, 1921) were estimated.

Despite potential outliers, these data demonstrate general consistency across reserves and represent an adequate picture of socioeconomic conditions in the Qu'Appelle Agency, at least from the DIA point of view. Since income data were estimated from data on agricultural production and other sources of data, it will not be used in this thesis. Instead, income will be estimated based on simulated sources of income such as agricultural production and livestock.

3.2.1. Aggregate data analyses

Aggregate data analyses included preliminary analyses that were meant to ensure the quality of the data, and correlation and regression analyses that studied relationships between data series.

3.2.1.1. Preliminary analyses

Preliminary analyses were carried out to assess and address potential problems of outliers and stationarity of data. Presence of outliers was evaluated using the methodology presented in Grubbs (1969). According to this approach, an outlier is defined as a data point located further from the mean than 1.5 times the distance between first and third quartile.

Second, the data were examined to determine stationarity. Examination of Figure 1-1, Figure 2-5, Figure 2-12, and Figure 2-14 suggests that the data are trended. The presence of trend may lead to the outcome under which a relationship between time series may be due to the time trend alone. Also, statistical methods applied to time series data require that the data be covariance stationary, which is not the case with trending data (Watson, 1986). For these reasons, data were examined to determine stationarity. If data were found to be non-stationary, the data were transformed by fitting an autoregressive integrated moving average (ARIMA) model.

ARIMA models are a class of models used to study and forecast time-series data. ARIMA models represent a time-series as a combination of its lagged values and/or errors.(Makridakis & Hibon, 1995) Lagged values are represented by the autoregressive (AR) components of the model, while the error terms are represented by the moving average (MA) components. Moving average components are called that way because they represent the weighted sum of past error terms. Autoregressive components represent weighted lagged values of the time-series (Hamilton, 1994). ARIMA models are usually described by specifying the number of times it was differenced, as well as the number of AR and MA terms and are an effective way to ensure stationarity of the time-series model (Makridakis & Hibon, 1995)

3.2.2. Correlation and Regression analyses

Several methods were used to analyse the relationship between TB mortality and the SDH. These analyses were carried out with the objective of developing the ABM by studying the relationship between TB mortality and SDH over time, as well as by developing important components of the model, such as agricultural production. Analyzing relationship between TB mortality and SDH can help identify important determinants of TB mortality, relationships between them, and changes in these relationships that need to be incorporated into the model in order to recreate relationships between the agent, the host, and the environment in a historically-accurate way.

First, correlations with TB mortality were calculated using current as well as lagged values. Given a value of the data series in the current period, lags of a data series represent the value of the data series in the previous period. For example, in year 1885 one period lag of TB mortality data would be the value of TB mortality in year 1884. In analyses that follow, lags of one period are used. Lag of one year was used to minimize the loss of data points, and also because of the high speed of disease progression in the Qu'Appelle Agency (Ferguson, 1928).

Second, univariate regression models of TB mortality on all data series as well as first lags of these series were fit. Univariate regressions had the following form:

Equation 3-1: Univariate regressions

$$TBmort_t = \propto + \beta X_t + e_t$$

where $TBmort_t$ is TB mortality in year t, and X_t is each of the data series mentioned above.

Third, parameters that were statistically significant in univariate analyses were retained for multivariate analysis to evaluate the relationship between TB mortality and the SDH. Separate from this multivariate regression, multivariate regression analyses were carried out to estimate the rules that governed agricultural and cattle production. In all multivariate regressions, parameters that entered the multivariate regression were retained based on adjusted R-squared, Akaike information (AIC) and Bayesian Information criteria (BIC). All analyses presented in this section were carried out using SAS EG software, Version 5.1 of the SAS System for Windows[2].

3.2.2.1. Agricultural production

The following section presents equations used to calculate individual wheat, oats, and hay production. Wheat and oats production was estimated as a function of weather (growing degree days and precipitation), area farmed, number of ploughs, and amount of hay cut under the following assumptions: Wheat and oats production:

- Depends on the area farmed
- Depends on environmental conditions, such as temperature and precipitation
- Inversely depends on time spent on competing occupations
- Depends on the availability of factors of production, such as ploughs

Given the above assumptions, the following regressions were estimated, with model selection based on adjusted R-squared, AIC, and BIC.

Equation 3-2: Wheat output regression

$$\boldsymbol{Wheat_t = \propto_w + \beta_{w1} Oats_t + \beta_{w2} Hay_t + \beta_{w3} GDD_t + \beta_{w4} Precip_t + \beta_{w5} Heat_t + \beta_{w6} Area_t + \beta_{w7} Ploughs_t + \beta_{w8} Stables_t + \beta_{w9} Oxen_t + \beta_{w10} Horses_t + \varepsilon_t}$$

Equation 3-3: Oats output regression

$$Oats_t = \propto_o + \beta_{o1} Wheat_t + \beta_{o2} Hay_t + \beta_{o3} GDD_t + \beta_{o4} Precip_t + \beta_{o5} Heat_t + \beta_{o6} Area_t + \beta_{o7} Ploughs_t + \beta_{o8} Stables_t + \beta_{o9} Oxen_t + \beta_{o10} Horses_t + \varepsilon_t$$

[2] Copyright © 2012 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

Equation 3-4: Hay output regression

$$Hay_t = \propto_o + \beta_{h1} Wheat_t + \beta_{h2} Hay_t + \beta_{h3} GDD_t + \beta_{h4} Precip_t + \beta_{h5} Heat_t + \beta_{h6} Area_t + \beta_{h7} Ploughs_t + \beta_{h8} Stables_t + \beta_{h9} Oxen_t + \beta_{h10} Horses_t + \varepsilon_t$$

where Wheat and Oats are wheat and oats harvests in bushels, GDD are growing degree days, Hay is the amount of hay cut, Precip is total precipitation in a given year, Area is area farmed, and Ploughs is the number of ploughs in the Agency.

3.2.2.2. Cattle production

DIA collected a significant amount of data on economic conditions that can be used to estimate the number of horses and oxen (see Table 3-1) assuming that the number of horses and oxen:

- Depends on the number of horses and oxen in the previous year
- Inversely depends on the time spent on competing occupations
- Depends on the amount of hay produced in the Agency
- Depends on the amount of precipitation (Ndzinge, Marsh, & Greer, 1984)

Given the above assumptions, the following regressions are estimated:

Equation 3-5: Number of oxen equation

$$Oxen_t = \propto_{Ox} + \beta_{Ox1} Oxen_{t-1} + \beta_{Ox2} Hay_t + \beta_{Ox3} Horses_t + \beta_{Ox4} Precip_t + \beta_{Ox5} Stables_t + \beta_{Ox6} Ploughs_t + \beta_{Ox7} Area_t + \beta_{Ox8} Oats_t + \beta_{Ox9} Wheat_t + \varepsilon_t$$

Equation 3-6: Number of horses equation

$$Horses_t = \propto_h + \beta_{h1} Horses_{t-1} + \beta_{h2} Hay_t + \beta_{h3} Horses_t + \beta_{h4} Precip_t + \beta_{h5} Stables_t + \beta_{h6} Ploughs_t + \beta_{h7} Area_t + \beta_{h8} Oats_t + \beta_{h9} Wheat_t + \varepsilon_t$$

where Oxen is the number of oxen, Horses is the number of horses, OxenL1 and HorsesL1 are numbers of horses and oxen in the previous year, Hay is the amount of hay harvested, Stables is the number of stables, Ploughs is the number of ploughs, and Precip is the total amount of precipitation in a given year, oats, potatoes and wheat represent respective harvests.

3.2.2.3. Number of stables

Stables were used to house farm animals, both oxen and horses. These animals worked on the farm, doing ploughing and transportation. Because of that, the number of stables depends on the size of overall operations (i.e. size of area farmed), the number of animals, as well as factors that influence the number of animals (hay and growing degree days) as well as the number of ploughs that these animals were meant to use.

Equation 3-7: Stables

$$Stables_t = \propto + \beta_{s1} Area_t + \beta_{s2} Houses_t + \beta_{s3} Ploughs_t + \beta_{s4} Oxen_t + \beta_{s5} GDD_t + \varepsilon_t$$

3.3. The Qu'Appelle Agency agent-based model

Like previous chapters, the following sections use the three components of the Epidemiologic Triad as a structure to describe the application of GSS to the study of TB outbreak in the Qu'Appelle Agency. While in the terminology of the Epidemiologic Triad the agent refers to the disease, ABMs apply the term agent to the "subject" of the model, an entity that has a number of characteristics that give it agency (Bandini et al., 2009).

In the present case, the objective of developing the ABM is to study how the Indigenous people of the Agency adapted to the new environment that came into being after the signing of the treaty, and how that new reality affected their health status. From that point of view, the host of the Epidemiologic Triad, individual members of the Indigenous people of the Qu'Appelle Agency are the agents of the agent-based model.

Agents live and interact in a variety of environments. They are assumed to have bounded rationality, in line with evolutionary economics, and create a number of social and economic structures. Their bounded rationality means that they do not have perfect foresight and they work with limited information. They are also highly individualistic, or heterogeneous.

These agents, or artificial people, are born, grow up and go to school, work as farmers or ranchers, collectors of wood, or cutters of hay, or simply wage labourers, get married and have children, age and die from TB and other causes. They do all of that following certain behavioural rules. These rules are of the form "if X, then Y". Despite its simplicity, the fact that agents change the environment and then change their behaviour in response to these changes, creates significant complexity in the range of possible behaviours, as well as the outcomes of the model.

The following paragraphs describe the Qu'Appelle Agency ABM. Since time is an explicit feature of an ABM, the model iterates over certain explicitly-defined time periods. For this reason description of the model is separated into two parts that describe what happens before and after the first iteration of the model.

The Qu'Appelle Agency ABM starts with an initial population of 200 people and iterates 61,356 times using a 6-hour time step over 41 years starting on January 1st, 1885. The model stops running on January 1, 1927 at the end of the period of interest.

Prior to the first iteration of the model, the social, economic, and physical environment of the artificial Qu'Appelle Agency are explicitly defined. The following sections describe how these environments and agent characteristics are assigned prior to the first iteration of the model (section 3.3.1 Statics). As well, the changes that take place in the model after the time starts running are described below in section 3.3.2 Dynamics.

3.3.1. Statics

This section describes what happens in the model before time starts running. Since there are no interactions at this stage, this section is called Statics. It describes how the three parts of the artificial Qu'Appelle Agency, namely, agent, host, and environment, are generated.

3.3.1.1. Infectious Agent

The focus of this ABM is on pulmonary TB – infectious disease spread by coughing. TB is not modelled explicitly like a generalized non-specific infectious disease in (J. A. Epstein & Axtell, 1996). Rather, it is modelled by changing individual states of agents based on certain events, such as infection and progression from latent to active disease. Transitions from one state to another depend on individual characteristics and are described further below. In other words, TB is modelled through individual characteristics of the artificial agents. Figure 3-1 presents disease-specific states through which agents may progress. Upper arrows indicate that a person in that state can stay in that state for more than one 6-hour period over which the model iterates.

3.3.1.2. Host

The artificial people, the agents of the Qu'Appelle Agency are generated before the first iteration of the model by assigning to them a number of individual characteristics. These characteristics fall into several categories each of which is defined in more detail below:

a) Personal characteristics: preferences (preferences for distance as well as personal threshold for making decisions), agent type, farming, ranching status, and school status, construction and visitation probabilities
b) Demographic characteristics: age, sex, housing (address, location, and quality), initial income
c) Social characteristics: number of family members, number of friends, strength of family and social connections, contact networks
d) Economic characteristics: number of horses and oxen, land, number of ploughs
e) Disease characteristics: susceptibility to TB infection, probability of infection, progression to disease, and death

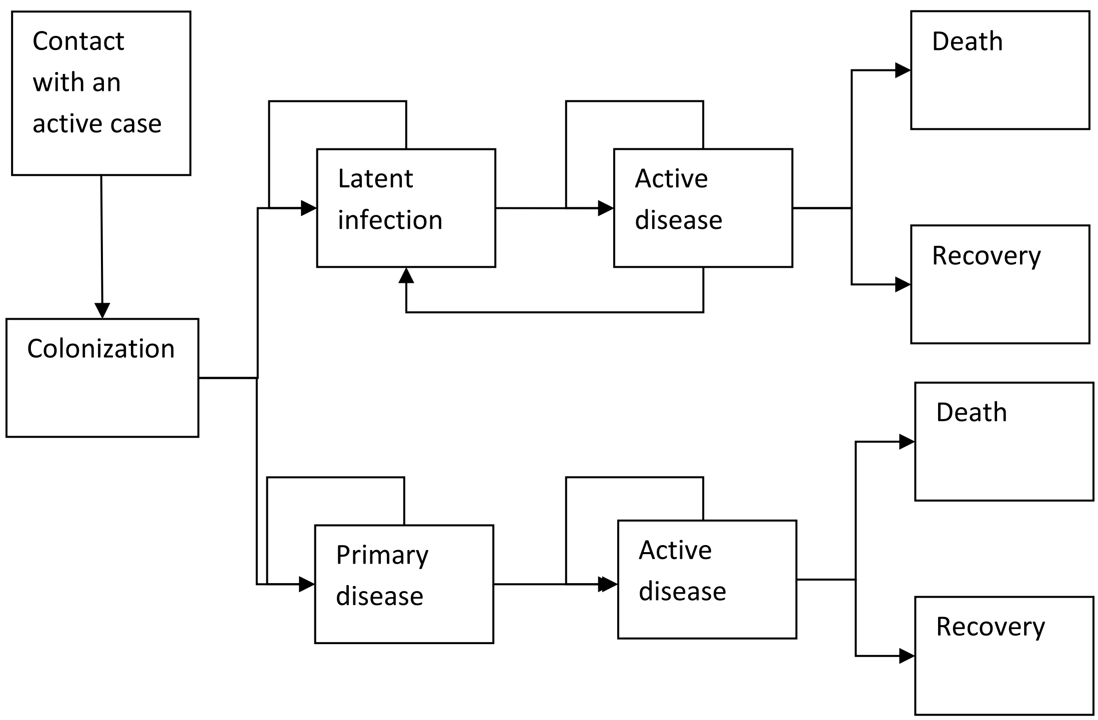

Figure 3-1: TB progression. Source: (Menzies et al., 2014)**(p. 25)**

These agent characteristics are described in more detail below.

a) <u>Personal characteristics</u>

Personal characteristics include (1) preferences, (2) agent type, (3) economic status, and (4) school status. These characteristics are defined below.

(1) <u>Preferences</u>

Every individual has preferences regarding how close they want to live to other people. Since the physical environment of the simulated Agency is a rectangular lattice with one square measuring one acre, the distance preference is expressed in acres. Acres are of identical shape.

In addition to preference for distance, they have an individual threshold that defines how much social support they need in order to take an action. This threshold is defined as a random number between 0 and 1, with 0 describing an agent that does not need any social support to make a decision, and 1 describing an agent who needs all of his contacts to take a certain action before he or she decides to do the same. This characteristic is related to the agent type presented below.

(2) <u>Agent-based model: Agent type</u>

Agent type is a binary number (0 or 1) that describes the agent as a leader (1), or a follower (0). Agent type has an impact on how agents make decisions once the model starts running. While all agents have a social support preference, leaders make decisions based on objective factors only, such as profitability of a certain type of activity, without taking into consideration how much social support there is for that activity. Followers use their social support preference and take social support into consideration in addition to objective factors such as profitability.

(3) <u>Economic status</u>

Annual Reports of the Indian Agent indicate that there were two non-Indigenous farmers living in the Agency at the time of the Agency's creation. These farmers were hired by the Indian Agent to teach farming to the people of the Agency. While the farmers were not Indigenous, in this model 2 adult Indigenous agents are assigned a status of a farmer. Non-Indigenous ranchers were not mentioned in the Reports. For simplicity 1 person is assumed to be a rancher.

Indian Agents had a significant impact on the Qu'Appelle Agency. Indian Agents are not modelled explicitly like other artificial people of the Agency, but they do have an impact on the amount of rations allocated according to the relationship between rations and agricultural output identified in Chapter 4 (Results).

(4) School status

Children between ages of 7 and 15 are in school. Children before the age of 7 are assumed to be at home. Like everything else in the Qu'Appelle Agency, the school system underwent transformation over time, with day schools being gradually replaced by boarding schools. While DIA published data on school attendance, these data were not published consistently. Given that the focus of this model is on the adults, all children are assumed to be in boarding schools.

b) Demographic characteristics

Demographic characteristics include (1) age and sex, (2) housing (address, location, and quality of dwelling), and (3) initial income as described in further detail below.

(1) Age and sex

Agents are assigned sex (male or female) and age. Age is assigned based on the simplifying assumption that proportion of population made up of children (0-15 years old), adults (16-64 years old), and elderly (65 years old and over) in 1885 when the Agency was created was 30%, 55%, and 15%, respectively.

(2) Housing

Individuals are initially assigned housing based on their family status so that members of one family share one house. Housing is assigned to families that are described in more detail below.

Two aspects of housing are relevant to the model: location of the house and quality of the house. The two aspects of the housing are introduced in order to explore their role in TB spread. Houses are distributed on the reserve based on individual preferences over distance to the nearest house.

The quality of the house, its ventilation, and size, may be related to the spread of TB. The impact of the quality of the housing is modelled through the reduction of TB transmission probabilities with the increase in the quality of the house. Quality of a house is expressed as a discrete positive number, such that house of quality 2 is better than house of quality 1.

Each house has an "address": a number and unique location on reserve, corresponding to a one acre square on the lattice that makes up the simulated Agency. Only one house can be located in a single square of the lattice. Both characteristics of housing change with time as agents make their decisions regarding where to live and whether to improve the quality of their housing.

(3) Initial income

Agents are assigned an initial income in order to enable them to take certain actions requiring income, such as getting married and having children. Since there are no data on actual income in the Agency until 1897, initial income assigned before the first iteration of the model is a random variable from 1 to 10 denoting $1 to $10.

c) Social characteristics

Social characteristics include (1) the number of family members and friends, (2) strength of family and social connections, and (3) networks of contacts as described and defined in further details below.

(1) Number of family members and friends

Based on a distribution described in section 3.3.3 each agent is assigned a random number that describes the size of his or her family as well as the number of friends. Based on these numbers, each member is assigned a corresponding number of family members of various ages.

(2) Strength of family and social connections

Along with family members and friends, each agent is assigned a parameter that governs strength of relationship between the agent and each person in the family and contact network. This parameter is used in modelling social contacts as described further below.

(3) Networks of contacts

Outcome of the above two steps is creation of families made up of three generations of artificial agents: children, adults and elderly. Since the number of family members is an individual characteristic, families tend to be different. Network of friends is assigned in a similar manner.

The only difference being that friends are assigned from the same age category, so a child has friends who are children, an adult has adult friends, and an elder has elderly friends.

These networks of contacts, both of family and friends, are closed, in a sense that if person A knows person B, then person B must know person A. Similarly, at the family level, if one spouse has children, they are by definition on the list of family members of the other spouse.

The requirement that contact networks be closed, along with the fact that allocation is probabilistic and initial number of agents is fixed, results in the outcome that initially some agents may have incomplete networks in that they have fewer contacts than they are supposed to have given their personal characteristics. Once the model starts running, incomplete networks are filled up by new agents born into the Qu'Appelle Agency.

Contact networks are not permanent: if someone loses a "friend", he or she may meet another agent and add that agent to the contact network.

d) <u>Economic characteristics</u>

Economic characteristics include the number of horses, oxen, and ploughs, and size of land. Since prior to the settlement on reserve bands of the future Qu'Appelle Agency were buffalo hunters who did not engage in farming or ranching, for simplicity before the first iteration of the model two randomly chosen adults become farmers, and one becomes a rancher. Since there are no data on size of land allocated, all individuals start with a one acre plot of land regardless of their farming status. Only the rancher gets horses and oxen, while only the two farmers are assigned one plough each. In the absence of data on the number of horses and oxen in possession of the first rancher, it is assumed that the first rancher starts with five horses and two oxen.

e) <u>Disease characteristics</u>

Disease characteristics include: (1) susceptibility to TB, and probabilities of (2) infection, (3) progression, and (4) death.

(1) Susceptibility to TB

Susceptibility to TB is assigned in the beginning, and is one of the factors that can modify probabilities of disease progression and mortality from TB. It is a genetic characteristic that is passed from parents to children. Inclusion of susceptibility to TB is justified for several reasons.

The role of genetic factors in TB progression has been debated for a long time. McKeown and Record (1962) supported the importance of TB susceptibility in TB progression, and Ferguson who observed the TB outbreak in the Qu'Appelle Agency argued that the course of the disease was different in the Indigenous population.

More recently, Gzybowski and Allen (1999) argued there was "growing support for the role of Darwinian genetics in shaping a TB epidemic" (S. A. Grzybowski & Allen, 1999), and that "recent incidence rates for TB among geographically defined aboriginal groups [...] have reflected the time of contact with the European settlers" (S. A. Grzybowski & Allen, 1999). Summarizing the debate around the issue of susceptibility, Waldram et al., (2006) state: "It is now generally accepted that ethnic differences in susceptibility to tuberculosis can be explained in terms of how long and under what circumstances the population has experienced the epidemic, and the extent to which herd immunity has been acquired through experience with it" (J. Waldram et al., 2006).

Recent genetic studies identified the mechanism through which genes affect susceptibility to TB (Bellamy et al., 1998; Li et al., 2011), and their prevalence in Canadian Indigenous population has been estimated (Greenwood et al., 2000; Larcombe et al., 2015). In addition to this, the concept of susceptibility to TB was included in the model in order to test the hypothesis that the severity of TB outbreak in the Qu'Appelle Agency could be explained solely by biological factors, as opposed to broader social ones.

Despite some support in the literature, the concept of susceptibility to TB has some drawbacks. First of all, the contribution of genetic factors to disease acquisition (Menzies et al., 2014) (p.25) and disease progression (Menzies et al., 2014) (p.26) currently is not well understood.

Susceptibility is modelled as a constant attribute that does not change. A person is either "susceptible" to TB, in which case he or she has a higher probability to transition from latent infection to active disease, or not susceptible. As evidence on the interaction of NRAMP1 polymporphisms with social and environmental factors becomes more substantial, this work could be extended to quantify this attribute of susceptibility, possibly as a residual once contribution of all other factors to the risk of developing TB have been taken into consideration.

Since the ABM iterates over a 6-hour period, probabilities of infection, disease progression, and mortality from TB are calculated for a 6-hour period of time. These probabilities are calculated from a triangular distribution based on the following parameters:

(2) Probability of infection

Individual probability of infection in a 6-hour period given contact with a person with an active form of TB was drawn from a triangular distribution with minimum of 0.0000001027, maximum of 0.0001027, and mode of 0.000001027 (Table 3-2).

(3) Probability of disease progression

Individual probability of disease progression from latent TB to active TB equals 0.00001712329 (Table 3-2).

(4) Probability of death

Probability of death from active TB in a 6-hour period of time equals 0.00009. This probability is calculated based on data from the Qu'Appelle Agency (Ferguson, 1928) (p. 28).

While these probabilities are assigned before the first iteration of the model, they are later modified by individual socioeconomic status.

3.3.1.3. Environment

Along with the social and demographic characteristics described above, the physical environment is also generated prior to the first iteration of the model. Physical environment includes the geography of the Agency, housing, and weather.

The Agency is simply a rectangular lattice where each square is assumed to be one acre in size. The size of the lattice is chosen so that it would correspond to the actual area of the Qu'Appelle Agency. Housing has two characteristics, as was described above, namely, location and quality.

Daily weather data for Regina were uploaded to simulate historic climate in the Agency. Weather data include daily temperature and precipitation for years 1885-1888 and 1901-1926

(Environment Canada., 2013). Data for years 1889-1900 were not available. With these characteristics assigned, the artificial world of the Qu'Appelle Agency is generated and the model is fully initialized.

3.3.2. Dynamics

The following sections describe what happens in the model once the time starts running. The model iterates 61,356 times, corresponding to 4x 6-hour time periods each day over 41 years from 1885 to 1926. At every iteration, agents are called up randomly, as opposed to being called up according to their number, so that their position does not play a role in the outcome of the model. With each iteration the model goes through several steps that, for simplicity, are presented by agent age and sex, namely, adult male, adult female, children, and elderly.

The diagrams presented below include only substantive activity blocks that describe what agents do. Administrative parts of the model responsible for things like keeping track of time and updating various matrices are omitted.

The following sections describe model flow chart components presented in Figure 3-2 below.

3.3.2.1. Adult men and unmarried adult women

Adult males are the principal agents of the model. They carry out most of the economic activity, generating the social determinants of health. It is assumed that adult unmarried women have the same occupations, i.e., they can be farmers and/or ranchers.

The ordered model cycle of events for this group of agents begins with the process of diffusion of preferences as described below.

Step 1: Diffusion of preferences

Diffusion of preferences is a two-step process that models decisions regarding: (1) the choice of occupation, with a particular attention to farming and ranching; as well as (2) the conditional choice of what activity to carry out on any particular day.

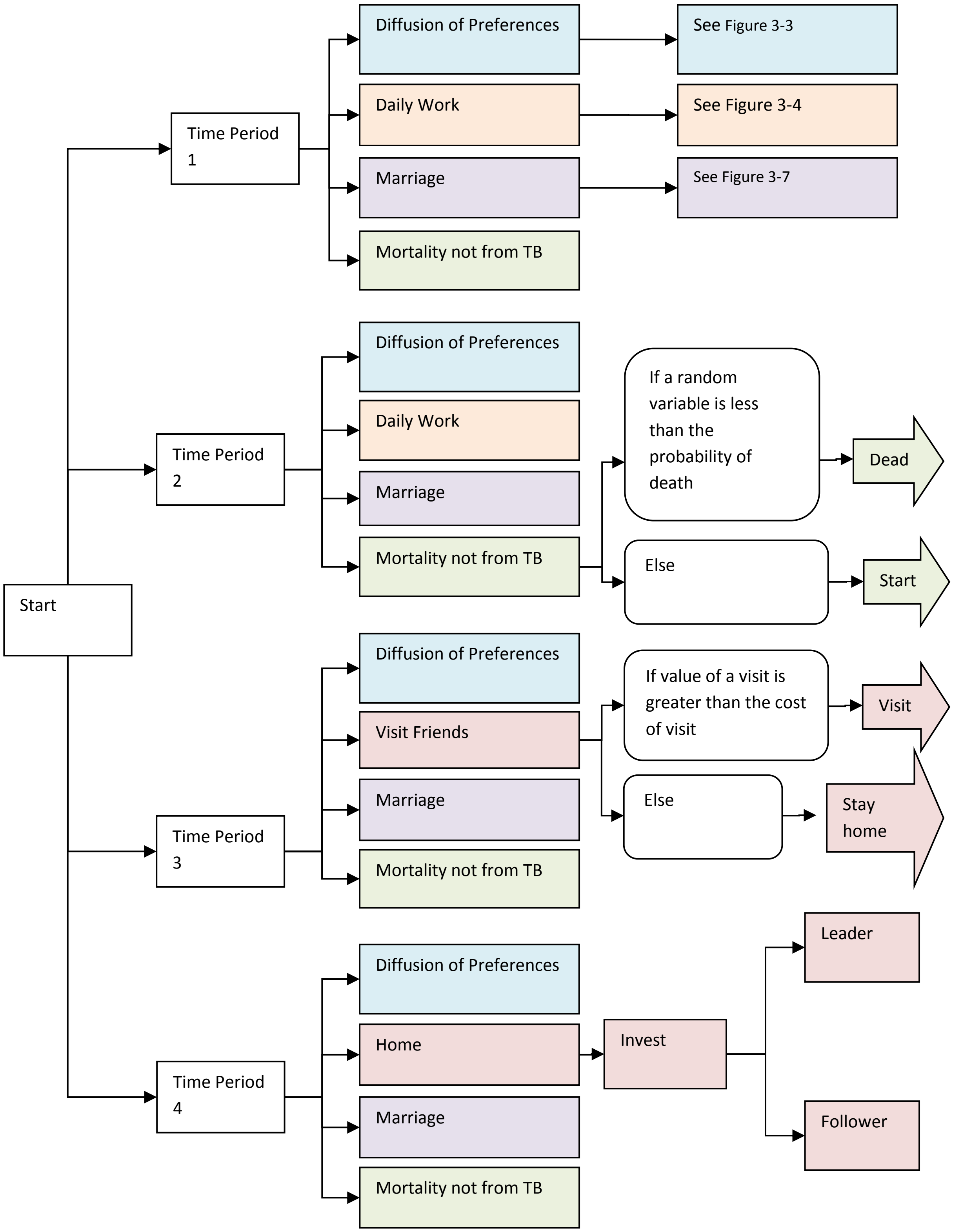

Figure 3-2: Model flow chart for adult men and unmarried adult women: First and selected

second level decisions

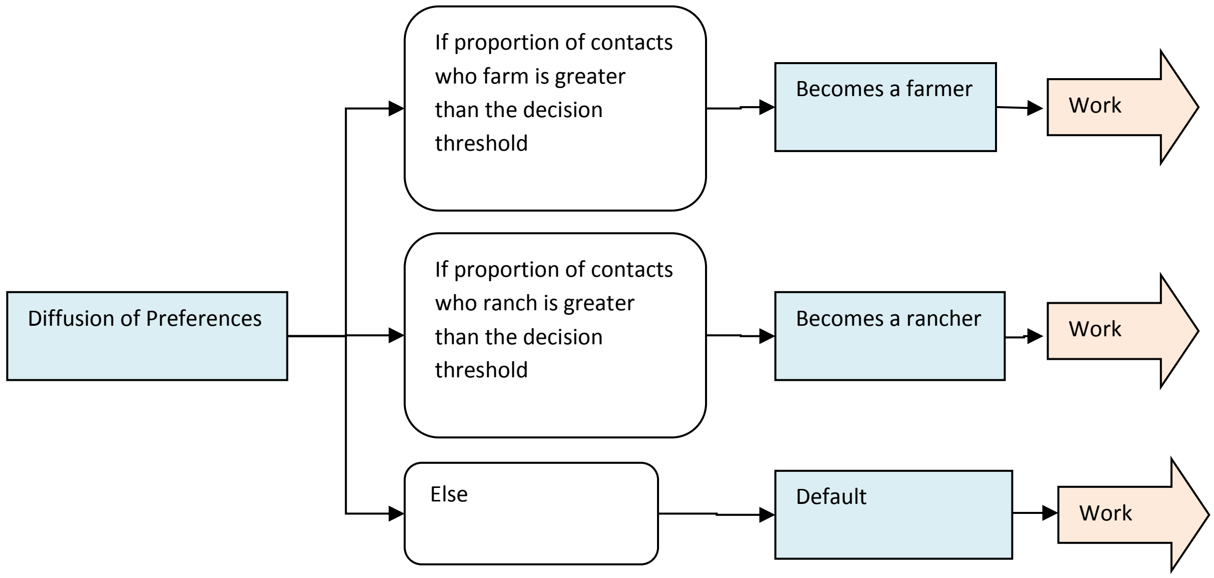

Figure 3-3: Diffusion of preferences

1. <u>Choice of farming and ranching</u>

Farming was one of the most important sources of income in the Qu'Appelle Agency. Unlike wage labour and the collection of dry wood, data on agricultural capital published in the Annual Reports as well as daily weather data allow for an econometric study of wheat and oats production as a function of actual levels of agricultural capital and climate. This in turn allows for simulation of the actual wheat and oats supply using an agent-based model.

Simulation of agricultural output is important for several reasons. First of all, it allows for a simulation of individual socioeconomic circumstances, which is important for exploring the link between social determinants of health and their distribution, and TB. Second, availability of detailed data allows for the comparison between the actual and simulated agricultural output, which is one of the ways to validate an agent-based model.

The process of growing wheat and oats is modelled as a four stage process that includes decisions to (1.1) farm, (1.2) plant, (1.3) harvest, and (1.4) the amount of wheat and oats harvested. The process of raising cattle is modelled as a two stage process that includes: (2.1) decision to ranch, and (2.2) increase in size of the herd. Both processes are described below.

(1.1) Decision to farm

The first decision that an agent has to make is whether to adopt farming. This decision is made on the basis of a personal threshold for making decisions, as well as return to farming. As was mentioned before, an agent, depending on his type, may follow one of two rules. As a leader, he will follow the rule "**if** return to farming is higher than return to other activities, **then** farm." As a follower, he will follow the rule "**if** a certain proportion of contacts farm and return to farming is higher than return to other activities, **then** farm."

A small number of agents adopt farming from the very start via initial conditions, with others following based on their individual contacts and preferences. This assumption has support from the actual history of farming in the Qu'Appelle Agency. Once the decision to farm is made, the farmer begins to grow oats and wheat.

(1.2) Decision to plant

Once the agent decides to farm, he needs to decide when to plant his seeds. While this decision does not necessarily have to be made the same day as a decision to adopt farming, it is an event related to the diffusion of preferences for these agents.

The decision to plant is modelled in the following way: leaders plant on the first sunny day in May, while followers also base their decision on the number of their contacts who have already planted. Once the proportion of contacts who have planted reaches their personal threshold for making decisions, the followers plant as well.

Similarly, raising livestock was another cornerstone economic activity in the Qu'Appelle Agency. Like growing oats and wheat, raising livestock became an important source of income in the first decade of the 20th century. Herds of oxen and horses are simulated in this model.

Simulation of the size of herd of oxen in the Qu'Appelle Agency follows the same set of steps, as the simulation of wheat and oats harvests, adjusted slightly to take into account the differences between raising livestock and growing wheat.

(1.3) Decision to harvest

Decision to harvest is made the same way as the decision to plant. Leaders decide to harvest on the first sunny day in September. Followers also base their decision on the proportion of their contacts who have harvested.

(1.4) Amount of oats and wheat harvested

Harvesting wheat, oats, and cutting hay occurs at specific times of the year. Harvesting starts in September, and is assumed to take one day.

Once the farmer decides to harvest total amount of wheat and oats is calculated according to Equation 4-2, Equation 4-3, and Equation 4-4 of Section 4.1.4.2.

(2.1) Decision to raise livestock

Agents fall into two categories regarding the decision to raise livestock: leaders, and followers. As described in the section on Statics, one rancher starts raising oxen as soon as the time starts running. Other agents adopt raising livestock on the basis of their personal thresholds for making decisions. When agents adopt raising livestock for the first time, they start with five horses. The number of oxen new ranchers get depends on the year: it is zero before 1896, and 15 thereafter. The reason for this is an apparent policy or social change that took place that year. In 1896, Indian Agent Lash talked about several band members starting to raise cows (DIA., 1886b).

(2.2) Increase in the size of the herd

This block of activities describes the increase in the size of horse and oxen herds. It is assumed that animals are born in April. The number of oxen and horses are calculated according to Equation 4-5, and Equation 4-6 described in Section 4.1.4.2.

2. Decision which activity to carry out on a particular day

Whether or not they decided to farm and ranch, agents have to decide what to do on any particular day. They have a choice of farming, ranching, collecting dry wood, cutting hay, and wage labour. Leaders make the choice between these activities based on their expected payoffs.

Followers first see how many of their contacts are occupied by each activity, and then make the choice based on their personal threshold for making decisions and payoffs to each activity.

There are certain trade-offs between activities. For example, when an agent decides to collect dry wood instead of planting, he or she foregoes one growing degree day which is used for calculating wheat and oats harvest. Once the number of growing degree days falls below a certain minimum, planting becomes no longer possible because of insufficient number of growing degree days to produce a crop.

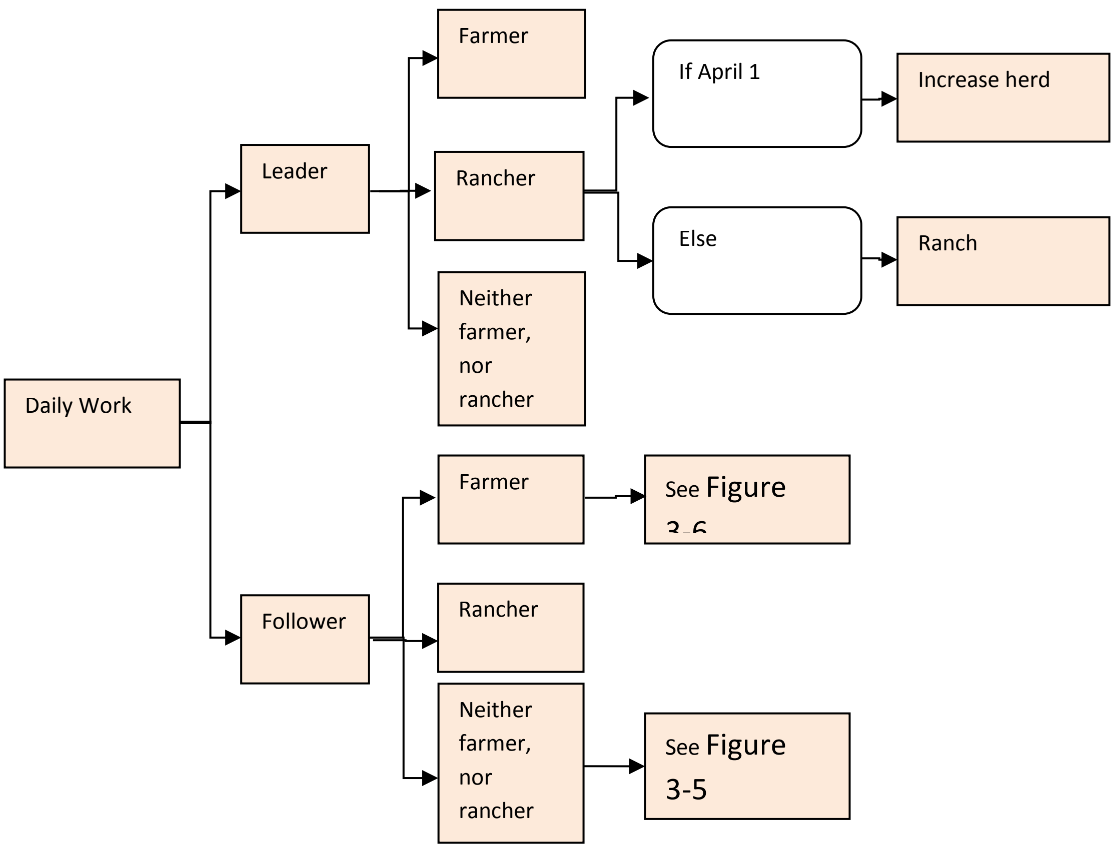

Figure 3-4: Work activities dynamics

<u>Step 2: Daily work activities</u>

Agents pursue their season-specific daily activities of choice. They spend the first two 6-hour blocks working. The choice of occupation described above matters not only for income, technological diffusion, and harvest, but also because it may have an implication for disease transmission. For example, wage labourers work outside of the Agency where, since the burden of TB was lower, the probability of disease transmission is assumed to be lower than in the Agency.

Daily activities of the two types of agents, leaders and followers, are presented in Figure 3-4.

Farming and ranching were described above, the other two types of economic activity are sale of dry wood and wage labour, as described below. Payoffs to farming and cutting hay are zero in the winter months.

(1) Sale of dry wood

Sale of dry wood to settlers was one of the earliest and most important types of economic activity in the Agency. While there are data on the income derived from the sale of wood, the quantity of wood sold and the number of people involved in this activity are not known. Reports of Indian Agents indicate that sale of dry wood lost its prominence as the major source of income because the supply of dry wood was exhausted by the end of the 19th century.

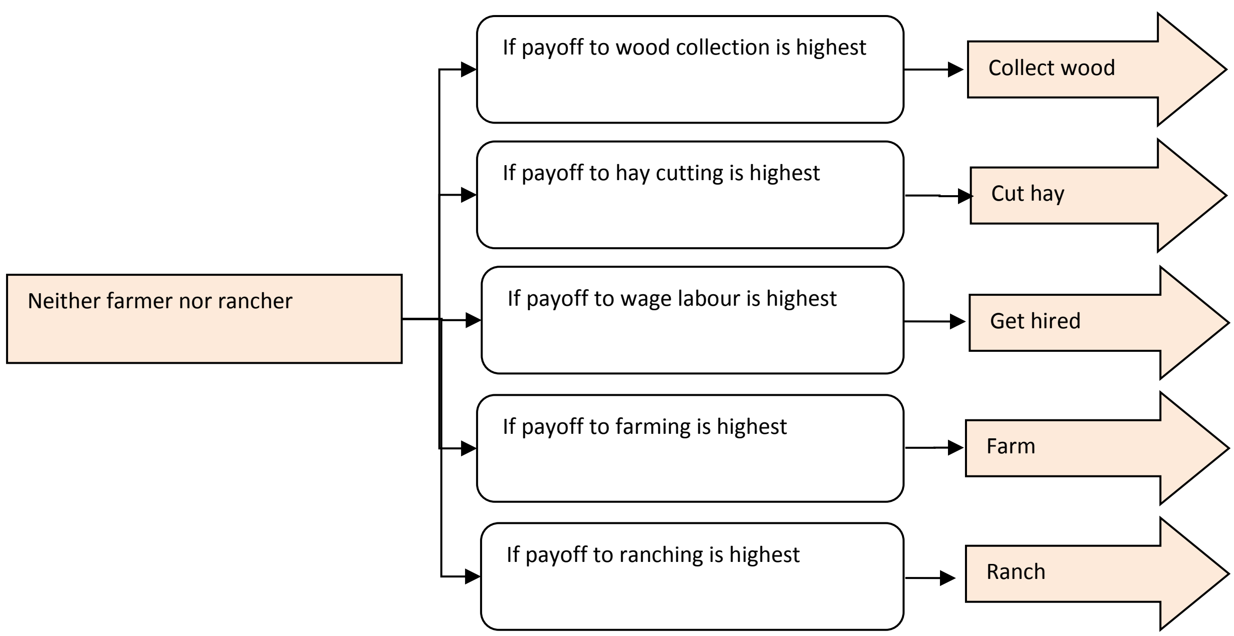

Figure 3-5: Other work activities decision tree in period 4

Because of the lack of data, the amount of wood is simulated as a fixed reserve that is being harvested by the residents of the Agency until it is fully exhausted.

(2) Wage labour

Importance of wage labour to the model derives from the fact that it took place outside of the Agency. While there are some data on the total amount of annual wage labour income, there are no data on the number of days worked, the hourly wage, or the number of wage labourers. Lack of data makes it difficult to model wage labour explicitly, like farming and ranching.

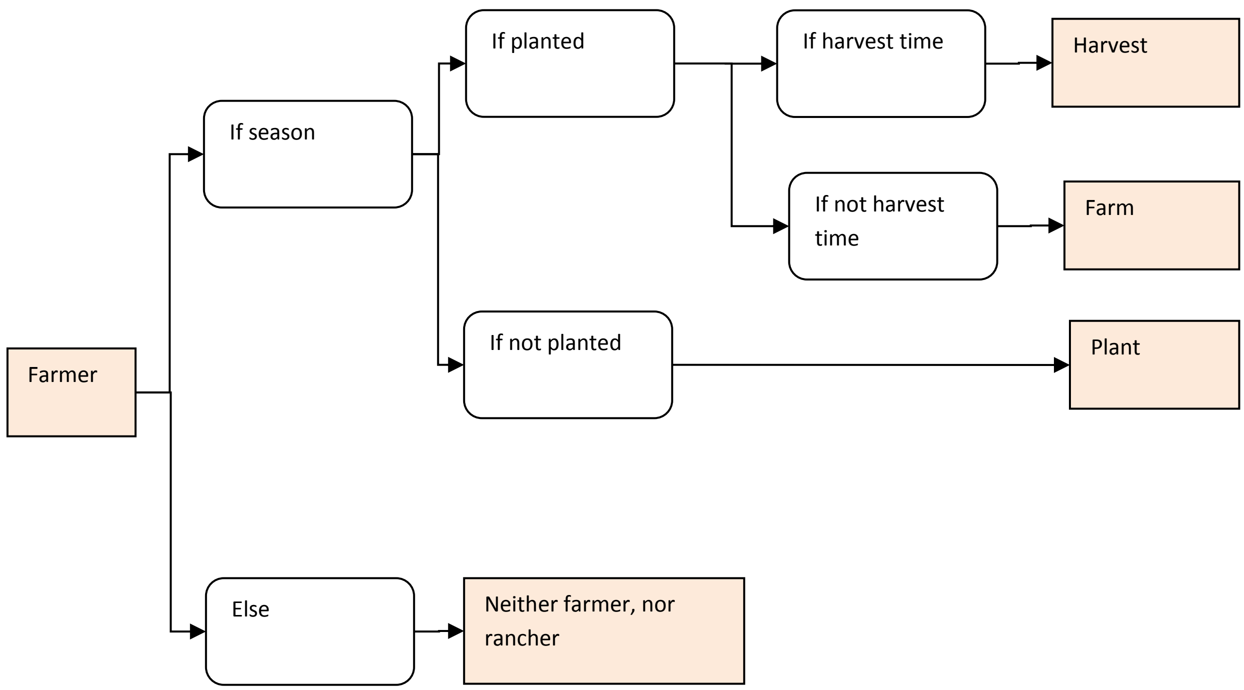

Figure 3-6: Farming activities dynamics

(3) Cost of food and rations

Expenditures on food are deducted daily. These expenditures are calculated based on the number of people in the household and per capita cost of food (Table 3-4). If household income is not sufficient to cover food expenditures, rations are provided in form of money to cover the gap between available income and daily expenditures on food. Once historical amount of rations is exhausted, all individuals in households that cannot cover their daily food expenditures are considered to be in a state of hunger with subsequent impacts on TB progression probabilities.

Step 3: Marriage and children

Only adults can marry in the model. This is a simplifying assumption that allows for an easier link between social processes, such as marriage and individual socioeconomic status. Given that the objective of this model is to explore the role of social determinants of health in TB outbreak, the linkage between social and economic processes, as well as disease transmission is crucial.

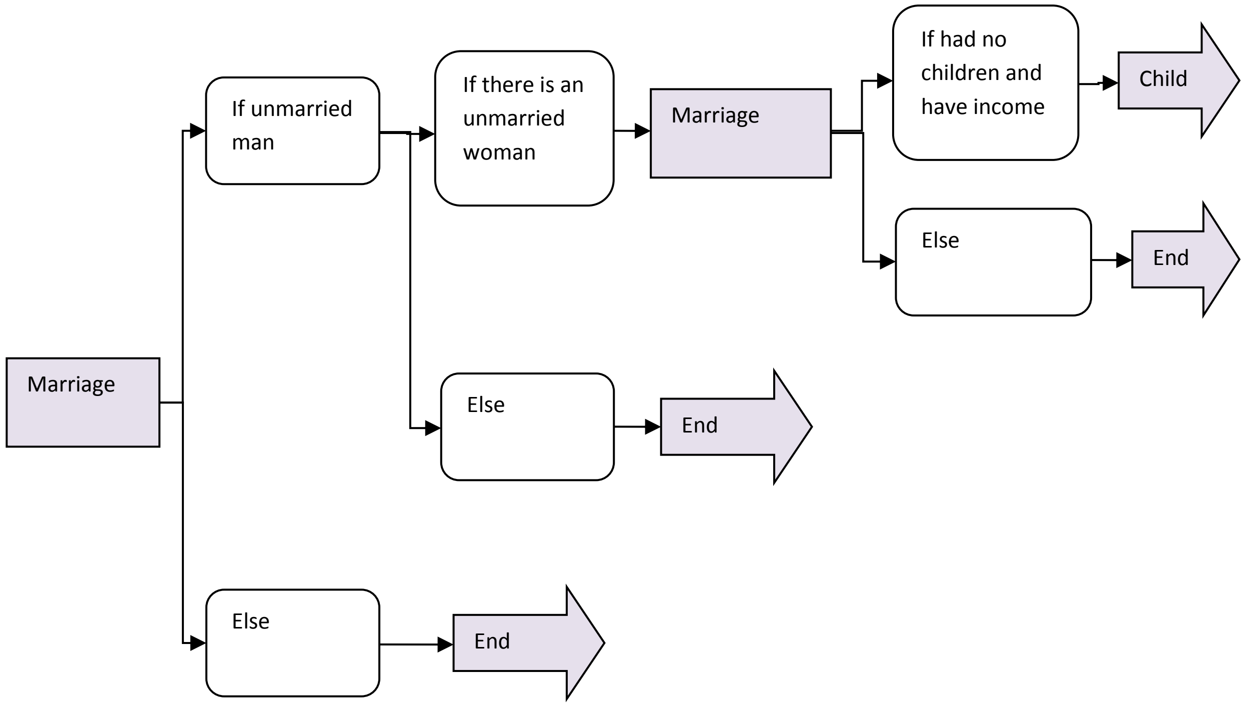

Figure 3-7: Marriage tree dynamics

Unmarried adult men may marry unmarried adult women. Once the marriage takes place, the newlyweds swap their model contact networks and build a new house by identifying an empty plot of land located a certain distance away from the nearest dwelling. Distance of choice is determined by their personal distance preference to the nearest dwelling. Once the house is built, a couple can start having children. The probability of having children depends on the historical birthrate for the Agency.

A simplifying assumption used in the model is that a couple below 50 can have only one child at a time if they have a certain amount of resources. The gap between children has to be one year. The couple can continue having children until they reach the age of 50. The number of children a couple can have is based on survey data (Table 3-4).

Children share “genetic” characteristics of their parents: the likelihood of the child being susceptible to TB depends on whether one or both parents are susceptible to the mycobacteria.

Step 4: Mortality from causes other than TB

Any person can die from causes unrelated to TB. Mortality from other causes is modelled as a random process that depends on historical mortality rate from causes other than TB.

Step 6: Visiting friends

Leisure activities take place in daily time periods 3 and 4, or between 6 pm and 6 am each day. Agents are social beings with networks of contacts with which they interact. These are the hours that agents can spend at home or visiting their friends. More specifically, time period 3 (6 pm to 12 am) is reserved for visitations, while period 4 (12 am to 6 am) is reserved for rest. Periods 3 and 4 are particularly important for modelling TB spread, given that agents tend to stay indoors.

In period 3, agents visit contacts from their social network depending on the distance to the contact, as well as an abstract utility derived from visiting the contact. Each agent calculates the distance from his or her own home to the home of the contact, compares it to the utility of visiting the contact, and decides whether to make the trip or not. This decision is made on the basis of individual characteristics that were assigned before the first iteration of the model. Given that each period is six hours long, agents can visit other agents as well as receive visitors in the same time period.

Step 10: Rest at home

In period 4, all agents return to their homes. Time at home is also the time when agents can plan their investment decisions.

Investments

This block allows agents to build a new house, and build or improve their stables. Like with other decisions, leaders and followers act differently. Leaders decide to build a new house based on their individual probability and level of income. If they have enough money to build a new house, there is a certain chance they will do so. They choose a location for their house based on the individual preference for the distance to the nearest dwelling. Once they find an empty spot with no house on it, they build a new house that is assumed to be of better quality than the one they had in the past. The flowchart of investment decisions is presented in Figure 3-8.

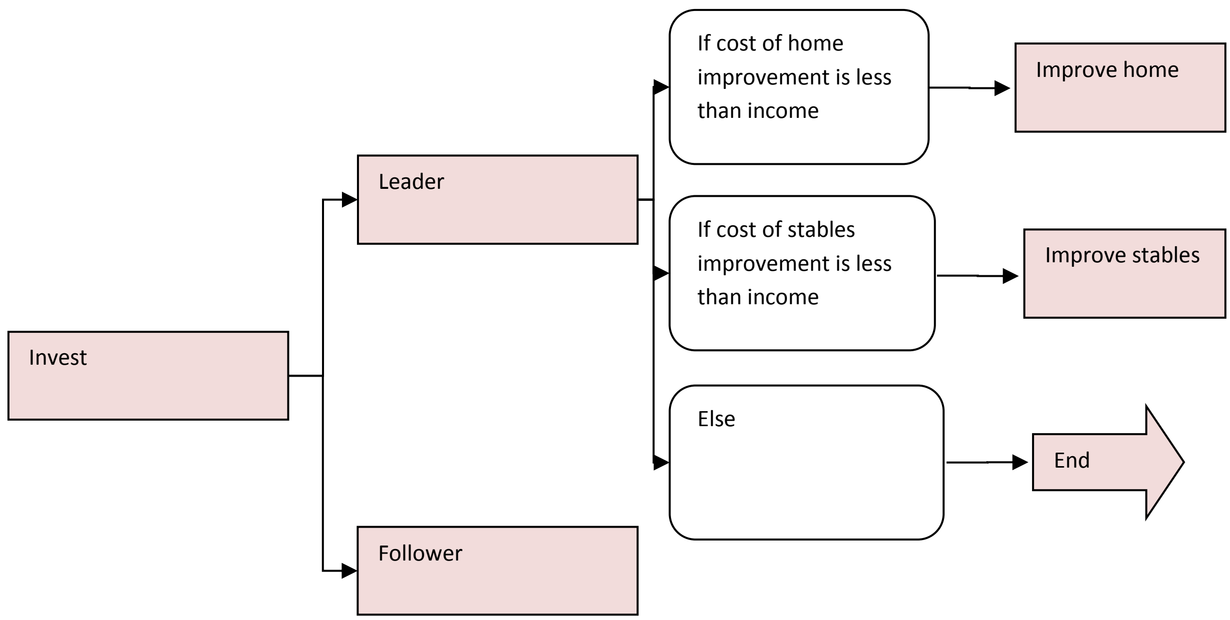

Figure 3-8: Investment decisions flowchart

Followers observe how many of their contacts have better houses than they do, and once the proportion of contacts with better houses exceeds an individual threshold, they decide to build a new house provided they have sufficient amount of money to do so. Then they follow the same steps that the leaders take.

The size of stables has an impact on the number of animals that they can hold. Leaders have a certain probability to increase the size of their stables if they can afford it. Followers observe how many of their contacts have bigger stables, and once the proportion of contacts with bigger stables exceeds their individual threshold, they may invest in bigger stables if they can afford it.

This last block completes the description of model blocks related primarily to adult men through which the model iterates.

3.3.2.2. *Married women and elderly*

Because men are assumed to be primarily engaged in economic activities which are the largest blocks of this ABM, other categories of agents play a much less prominent role in the model. It is true for married women and the elderly, and is even more so for the children.

The key events for this category of agents include the steps: (1) daily work activities; (2) daily leisure activities; and (3) mortality. These are discussed in more detail below.

Step 1: Daily work activities

Daily work activities of women and the elderly are not modelled explicitly, like they were for the men. It is assumed that women and the elderly also work outside during the first six hour block of the day, from 6 am to 12 pm. This is important for modelling disease spread, as being outside reduces the chances of disease transmission given a contact with infectious individual. Both women and the elderly spend the second 6 hour block indoors.

Step 2: Mortality from causes other than TB

Mortality from causes other than TB is modelled the same way as it is for adult men and unmarried adult women.

Step 4: Daily leisure activities

Daily leisure activities of women and the elderly are modelled the same way as those of men. They visit their friends between 6 pm and 12 am, and they are at home from 12 am to 6 am.

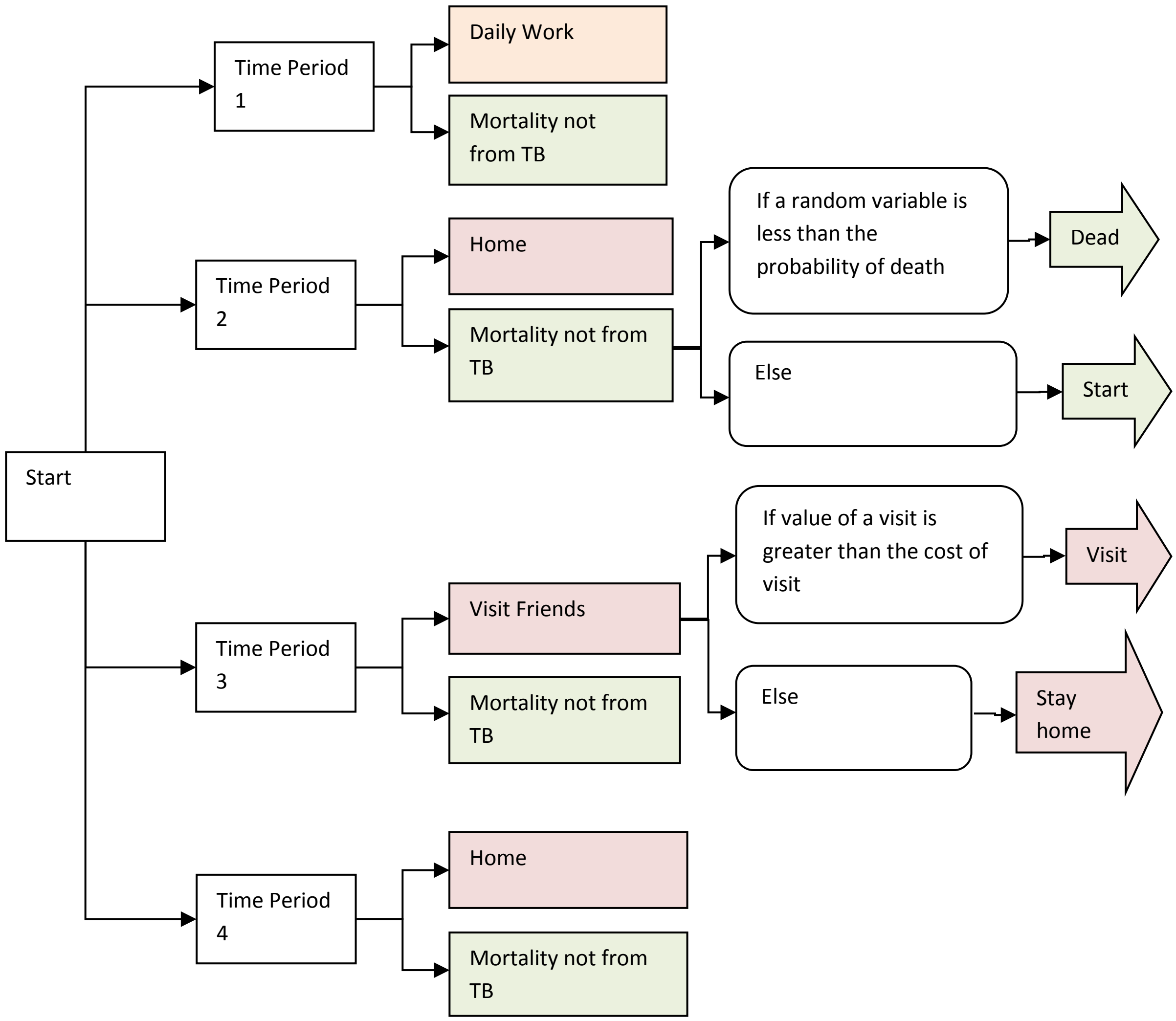

Figure 3-9: Model flow chart for women and elderly

3.3.2.3. Children

School-aged children attend a boarding school, where they divide their time between studying, working on the farm, and leisure. During holidays children return home to their parents, where they divide their time between working in the fields with adults, and leisure activities, such as visiting friends. Like adults and elderly, children can die from causes other than TB. These activities are described in detail below.

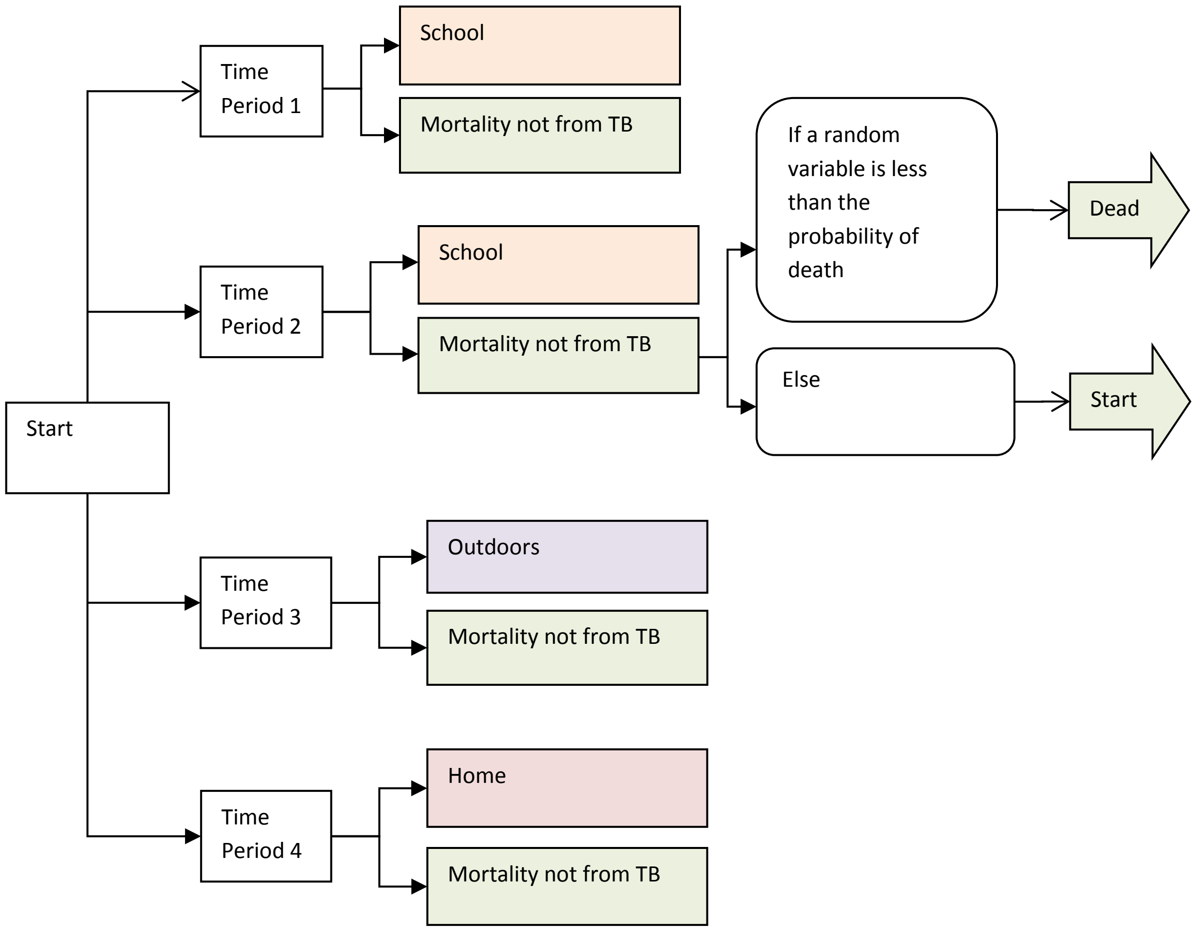

Figure 3-10: Model flow chart for children during school year

The flow chart presented in Figure 3-10 applies to school age children. Children under the age of 7 are assumed to stay at home.

Daily activities during school year

During school year children spend time periods 1, 2, and 4 indoors. They are studying or working indoors during the first two periods, and they work/play outside during the third period. They return home for period number four.

Daily activities outside of school year

Children return home during school holidays. It is assumed that boys spend time periods 1 and 2 working in the fields, while girls spend period 1 at home, and period 2 working in the fields. Both return home for periods 3 and 4.

Mortality from causes other than TB

Children also die from causes other than TB according to the procedure described above.

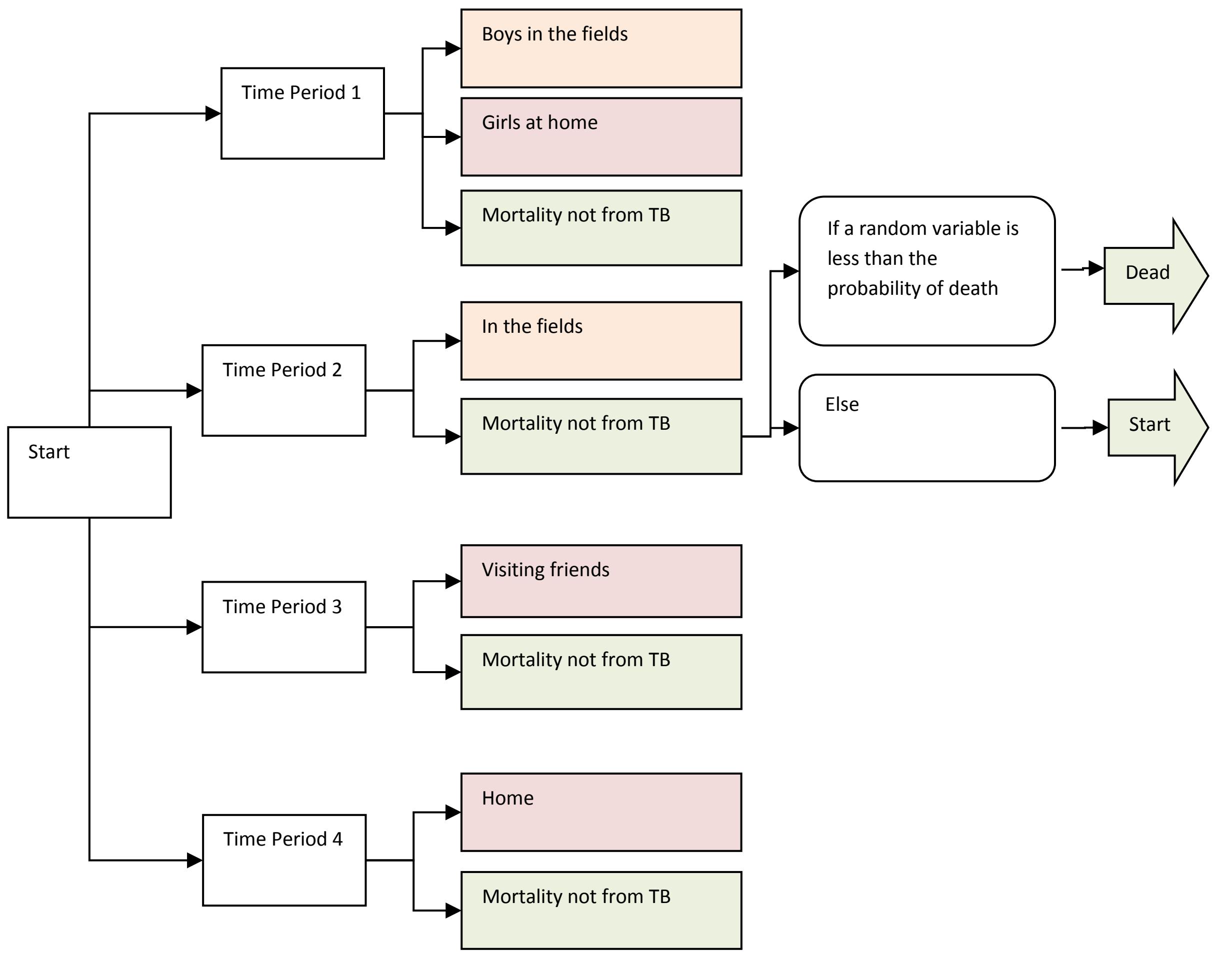

Figure 3-11: Model flow chart for children outside of the school year

This completes the description of major blocks through which the model iterates. The following sections describe how the model simulates the spread of TB through the population of the Agency as well as the progression of the disease of the infected individuals.

3.3.2.4. Disease transmission

Disease transmission refers to the acquisition of TB bacteria by an agent without TB from an agent with active TB. As was mentioned above, transmission of TB usually requires prolonged and close contact with an agent with active TB. Because of that, the probability of disease transmission depends on the setting: it is fairly low in an open field, and fairly high in the house, particularly among members of the same household.

3.3.2.5. Disease propagation

Once an individual is infected, disease progresses according to the natural history of TB (Figure 3-1). In order to link disease progression to social conditions, transition probabilities between disease states depend on individual socioeconomic status and susceptibility to the disease. Susceptibility to TB, hunger, and SES below the average for the Agency increase the speed of disease progression. Lack of susceptibility to TB and SES above the average reduce it.

Agents with latent infection have a specific probability of progressing to the active disease stage. This probability depends on several factors, such as individual probability of progression, susceptibility to TB, and socioeconomic status.

Once an agent develops active disease, he or she may either die, or return to the latent state. In both cases transition is modelled by modifying the basic individual probability of death and recovery by individual susceptibility to TB and socioeconomic status. It is assumed that agents with active disease stay at home at all times.

3.3.3. Model parameters

Model parameters are a key part of any model, as the quality of model output depends on the parameters used as inputs into the model. Baseline model parameters are presented in Table 3-2.

Table 3-2: Model parameters: Agent

Parameter description	Parameter value	Source
Probability of infection in a 6-hour period	0.000001027	(Stein, 1952)
Probability of disease progression from latent to active infection in a 6-hour period	0.00001712329	Calculated from Ferguson (1928)
Probability of dying from active TB in a 6-hour period	0.0000913242	Calculated from Ferguson (1928)
Multiplier for increased speed of disease progression among children as compared to adults	2.5	Calculated from Ferguson (1928)
TB susceptibility	28.3% [3]	(Larcombe et al., 2015)
Odds ratio of developing TB given TB susceptibility	1.35	(Li, Zhou et al., 2011)
Odds ratio of developing TB given malnutrition	12.43	(Cegielski, Arab, & Cornoni-Hutley, 2012)
Prevalence of latent TB cases in 1885	20%	Assumption
Prevalence of active TB cases in 1885	10%	Assumption
Multiplier for the risk of infection off reserve	0.05	Ferguson (1928)

[3] Estimate for the Dene population is used. Dene have a higher prevalence of alleles associated with higher risk of developing TB, than the Cree and Saulteaux population of the Qu'Appelle Agency. Since the Dene live further north and came into contact with TB later than Cree and Saulteaux, prevalence of these alleles among them is likely to be closer to what it was in the Cree and Saulteaux population in the past

Table 3-3: Model parameters: Host

Parameter description	Parameter value	Source
Demographics		
Initial population size	200	Chosen because of limits to processing power
Initial number of farmers	2	Annual Reports of the Indian Agent
Initial number of ranchers	1	Assumption based on the fact that they transitioned from no ranchers to some ranchers
Proportion of leaders	5-10%	(Couzin, Krause, Franks, & Levin, 2005)
Number of friends	Min 0, max 150, mean 7.3	General Social Survey, 2008 “Social Networks”. Question: How many close friends do you have?
Proportion of children in the population	33%	DIA(1901)
Proportion of adults in the population	63%	DIA(1901)
Number of children in a household (used to generate initial population only)	Min 0, max 8, mean 2	Total number of children raised by respondent. 2006, General Social Survey “Family Transitions”.
Number of elderly in a household (used to generate initial population only)	Min 1, max 3, mean 2	General Social Survey, 2006 “Family Transitions”. Question: Type of parents respondent has within household.
Proportion of families with children (used to generate initial population only)	69.1%	Total number of children raised by respondent. 2006, General Social Survey “Family Transitions”. Question: Total number of children raised by respondent
Age of children	1-15 years old	DIA reported data by age for the following age groups: under 6, 6-15, 16-20, 21-64, and over 65
Age of adults	16-64 years old	DIA reported data by age for the following age groups: under 6, 6-15, 16-20, 21-64, and over 65
Age of elderly	65 years old and over	DIA reported data by age for the following age groups:

		under 6, 6-15, 16-20, 21-64, and over 65
Proportion of females in the population	56%	DIA(1901)
Preferences		
Personal distance preference	1, 10	Assumption
Personal threshold for making decisions is the lowest proportion of contacts taking an action for the agent to take the same action	Random with a range Min 0.2, max 0.9	Assumption
Contact network		
Number of TB contacts	Min: 0, Mean: 2.5, Max: 20	(Al-Azem, 2007)
Transition probabilities		
Mortality from causes other than TB in a 6 hour block	Calculation using the formula: $hmort = 1 - (1 - ymort_i)\hat{}(\frac{1}{1460})$ where ymort is yearly mortality presented in Table 7-1 and i is year	Calculated from data in Ferguson (1928)
Probability of giving birth	Calculation using the formula: $pbirth = 1 - (1 - pChild_i)\hat{}(\frac{1}{1460})$ where pChild is a yearly birthrate presented in Table 7-1 and i is a year. This probability is adjusted for the number of couples in the Agency	Calculated from data on births presented in Annual Reports of the Indian Agent

Table 3-4: Model parameters: Environment

Parameter description	Parameter value	Source
Size of the Qu'Appelle Agency	200,000 acres	Annual Reports of the Indian Agent
Cost of stable improvement	$50	Assumption based on the cost of the house
Cost of house	$200	Assumption based on the fact that construction of an Agent's house cost $606 in 1885 (DIA., 1886a)
Cost of food, per person per day	$0.33	(DIA., 1902a) (p.1106)
Size of land for new farmers in the 1890s	10 acres	Assumption
Size of land for new farmers in the 1900s	100 acres	Assumption
Decrease in the size of area farmed in response to poor harvest	25%	Assumption
Increase in the size of area farmed in response to a good harvest in the 1890s	0%	Assumption
Increase in the size of area farmed in response to a good harvest in the 1900s	50%	Assumption
Initial wood supply	1000 cubic metres	Assumption
Probability of disease transmission when the number of people per house is below the mean	0.27	(Stein, 1952)
Probability of disease transmission when the number of people per house is above the mean	2.27	(Stein, 1952)
Probability of disease transmission outdoors	0	(Menzies et al., 2014) (p.36)
Reduction in the probability of disease progression because of above-average socioeconomic status as measured by income	30%	(Stein, 1952) As measured by mortality by socio-economic status
Increase in the probability of disease progression because of below-average socioeconomic status as measured by income	30%	(Stein, 1952) as measured by mortality by socio-economic status
Growing degree days (GDD)		(P. Miller, Lanier, & Brandt,

		2001)
Precipitation		(Environment Canada., 2013)
Daily temperature		(Environment Canada., 2013)
Wheat harvest	$Wheat_t = 1876.7 - 3.4 * Hay_t + 48.5 * Oxen_t + 39.2 * Precip_t + \varepsilon$	See section 3.1.4
Oats harvest	$Oats_t = 12173 + 1066.4 * Stables_t + 8.8 * Hay_t + 108.4 * Precip_t + \varepsilon$	See section 4.1.4
Hay cut	$Hay_t = 0.08 + 0.02 Oats_t - 0.05 Wheat_t$	See section 4.1.4
Number of oxen	$Oxen_t = 3.99 - 4.5 * Houses_t + 2.6 * Stables_t + \varepsilon$	See section 4.1.4
Number of stables	$Stables_t = -3.3 + 0.01 * Area_t + 1.06 * Houses_t - 0.5 * Ploughs_t + 0.04 * Oxen_t + 0.04 GDD_t$	See section 4.1.4

3.4. Model output analyses

Developing model structure and parameters are important steps in the creation of a model. Another step is the way the model will be validated and the output analysed. While validation of agent-based models has been an on-going issue in the social simulation literature, it appears that there has not been a systematic evaluation of whether different approaches to model validation are more preferable than others, especially given the differences in approaches to agent-based modelling (Moss, 2008). Among existing approaches to model validation, the one used to evaluate the Qu'Appelle Agency ABM is called the history-friendly approach. History-friendly calibration of agent-based models includes three steps as follows: (Moss, 2008)

1. The ABM is designed on the basis of available information, such as empirical or historical studies.

2. This information is also used to identify initial parameter values that are likely to generate observed historical data.
3. Model output is compared with actual history as represented by the data.

The three steps of the history-friendly approach to model validation raise the issue of how to compare simulated output with actual data.

Axtell et. al., (2002) in their seminal series of papers on the artificial Anasazi did not use specific metrics to compare model output with historical data (Axtell et al., 2002). While in some instances it may be easy to select model runs that are closer to the actual data just by looking at the graphs, in cases where different sets of model runs are close to each other, selecting the best candidate will be easier if the goodness of fit of model output to historical data were quantified.

Most often, model fit is evaluated using a least-squares based on squared distances between model output and actual data (Thorngate & Edmonds, 2013). While providing a good measure of average fit of the model output to actual data, least-squares methods are not preferable for measuring ordinal differences between two sets of data. Thorngate and Edmonds (2013) propose Ordinal Pattern Analysis to evaluate ordinal fit of model output to historical data.

Ordinal pattern analysis (OPA) creates an index of fit that "indicates the chances that an ordinal prediction derived from a simulation's output will be matched in the sample of observations at hand" (Thorngate & Edmonds, 2013). OPA contains two steps. First of all, ordered lists of all pairs of data points using historical data and model output are created by comparing each data point with the next one. At the second stage, the number of matches between the ordered pairs of model output and the historical data are compared (Thorngate & Edmonds, 2013).

Use of OPA can be illustrated with an example as presented in Figure 3-12: Series1 and Series2 have exactly the same shape, but each data point of Series2 is below Series1. A linear regression fit to Series1 has a completely different shape, but is a better representation of the average than Series2. According to OPA, Series2 would have a perfect fit to Series1, because the shape of the series is the same.

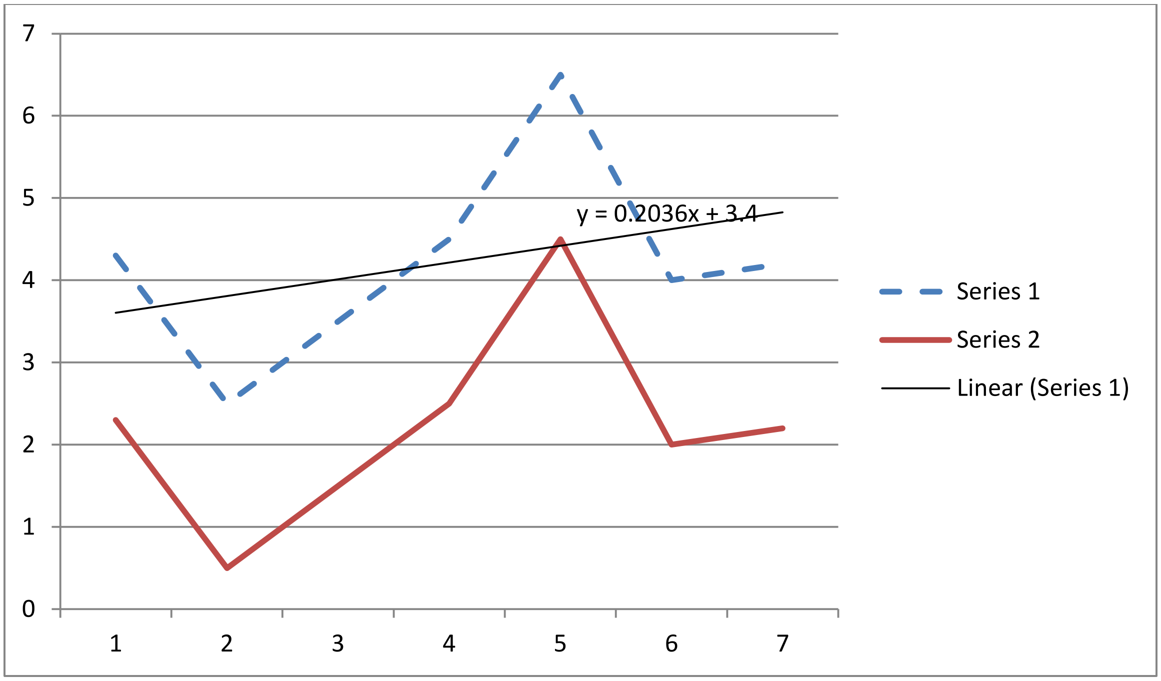

Figure 3-12: Example: application of OPA

OPA allows one to compare the topological properties of the two data series, but not other properties, such as the average. For this reason, OPA has to be used with another measure, such as area under the curve (AUC). In the following analyses, both OPA and AUC will be compared between model output and actual data to select the best fitting candidates. In order to reduce the running time of the ABM, all model runs used a population of 200, which was 0.134 of the actual population in 1885. For this reason, AUC estimates were scaled accordingly.

3.4.1. Uncertainty and sensitivity analyses

Uncertainty analyses are performed to evaluate the contribution of uncertainty in model inputs to uncertainty in model outputs (Marino et al., 2008). Sensitivity analyses refer to statistical methods for measuring the ability of parameters to influence the output of a model (Dancik, Jones, & Dorman, 2010) , identifying critical inputs (Marino et al., 2008), as well as to attributing uncertainty in output to different sources of uncertainty in model parameters (Saltelli, 2002). Given the complexity of ABMs, sensitivity analyses may impose a significant expense in terms of processing time and evaluation of results. Because of that there are various approaches to reduce the scope of the problem by focusing on variables of interest (Uusitalo et al., 2015), selecting time points for analysis (Marino et al., 2008), and efficient technique (Marino et al., 2008; Morris, 1991). Variables directly related to TB mortality are presented in Table 3-5.

While there are a number of methods for performing sensitivity analyses, the one used here is based on partial rank correlation coefficient (PRCC). PRCC describes the linear relationship between each input parameter and output after linear effects of other inputs have been discounted (Marino et al., 2008). In this case, both inputs and outputs have been rank transformed. For more on the PRCC methodology, please see Marino et al. (2008), including the supplement. PRCC analysis was implemented in MATLAB using code provided by Marino et al. (2008).

Table 3-5: Variables selected for sensitivity analyses

Variable name	Description	Range	Source
pInf	Probability of infection in a 6-hour time period	Min: 0.000001027 Mean: 0.00001027 Max: 0.0000684932	Assumption (Stein, 1952) Calculated from Ferguson (1928)
pPrg	Probability of disease progression in a 6-hour period	Min: 0.000001712329 Mean: 0.00001712329 Max: 0.0001712329	Assumption Calculated from Ferguson (1928) Assumption
Suscept	Susceptibility to TB	Min: 0 Mean: 0.283 Max: 0.8	(Larcombe et al., 2015) (Larcombe et al., 2015) Assumption
prevA	Prevalence of active TB cases in 1885	Min: 0.025 Mean: 0.1 Max: 0.5	Assumption Assumption Assumption
prevL	Prevalence of latent TB cases in 1885	Min: 0.05 Mean: 0.2 Max: 0.8	Assumption Assumption Assumption
TBcmin	Minimum number of TB contacts	Min: 0 Mean: 0 Max: 1	(Al-Azem, 2007) (Al-Azem, 2007) (Al-Azem, 2007)
TBcmean	Mean number of TB contacts	Min: 2.4 Mean: Max: 11	(Al-Azem, 2007) Based on the number of

			contacts (Table 3-4)
TBcmax	Maximum number of TB contacts	Min: 20 Mean: Max: 81	Based on the number of contacts (Table 3-4)
cFood	Cost of food per person per day	Min: \$0 Mean: \$0.33 Max: \$0.5	(DIA., 1902a) (p.1106) Based on the cost of a hospital day in 1902. (DIA., 1902a)

Latin Hypercube Sampling (Blower & Dowlatabadi, 1994) was used to sample parameter values from a uniform distribution with ranges presented in Table 3-5. A total of 43 samples was drawn and PRCCs between selected variables (see Table 3-5) and model output were calculated.

3.5. Research process

The research process includes the following steps:

1. Data are tested for stationarity and outliers. Non-stationary data are transformed by fitting an ARIMA model. Potential outliers are examined in light of other evidence. Confirmed outliers are dropped.
2. Correlation, univariate, and multivariate regression analyses are carried out to study the relationship between TB mortality and the social determinants of health at the aggregate level, as well as to develop the ABM as described in Section 3.2.1
3. Baseline scenario (Table 3-2, Table 3-3, and Table 3-4) of the ABM is run.
4. Uncertainty and sensitivity analyses are conducted as described in Section 3.4.1
5. The output of the model is aggregated across the agents, and aggregate yearly output is evaluated as described in section 3.4.
6. Policy analysis is carried out

With methodology described and data presented, the next chapter presents results of the analyses.

4. Analyses and Results

4.1. Aggregate data analysis

4.1.1. Stationarity

Analyses to determine stationarity of data and the final form of each ARIMA model, where necessary, are presented in Table 4-1.

Table 4-1: Final form of ARIMA models fit

Data series	Order of Differencing	AR	MA
TB mortality	1	0	2
Potato harvest	0	1	0
Hay cut	1	0	1
Oats harvest	1	3	1
Wheat harvest	1	0	1
Number of houses	1	1	0
Number of ploughs	1	0	1
Number of stables	1	0	0
Number of horses	1	0	1
Number of oxen	1	0	0
Area farmed			
Expenditures	1	0	0

The autocorrelation functions (ACF) of TB mortality, potato, hay, oats, and wheat harvest, the number of houses, ploughs, stables, horses, and oxen, as well as expenditures presented in Appendix C suggest that these data might be non-stationary. As a result, these data were transformed in the way described in Table 4-1until stationarity was attained. Subsequent analyses were carried out using stationary data.

4.1.2. Outliers

Out of the total of 41 years of observation only 15 have no outliers in any of the data series (Figure 4-1). While statistical methods help identify outliers, they do not specify what is to be done with them. Decisions about what to do with outliers must be taken on non-statistical grounds (Grubbs, 1969), and while "...it is tempting just to throw them away. This would be a major mistake. Often influential observations are the most valuable observations in a data set."(Kennedy, 2003)(p.374) With this in mind, and adopting a cautious approach in order not to drop potentially valuable observations (Kennedy, 2003), outliers were examined starting with years 1914 and 1915 and proceeding with each individual data series. Detailed discussion of outliers can be found in Section 7.1.2 of Appendix C, while the subsequent paragraphs summarize that discussion.

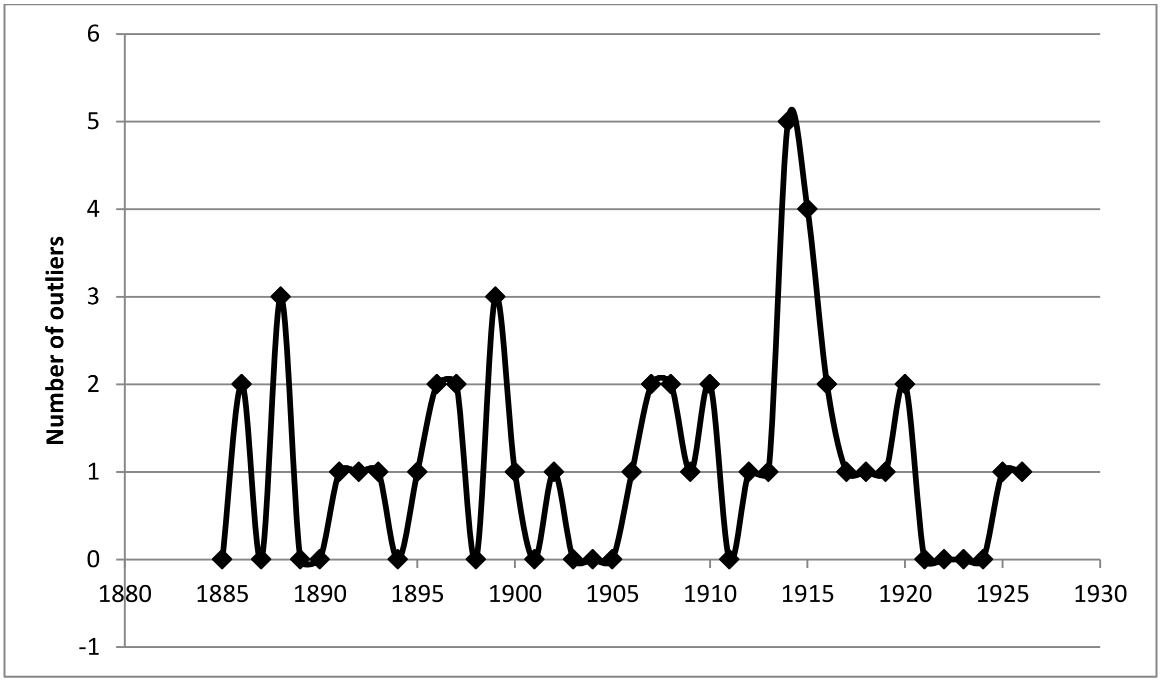

Figure 4-1: Number of outliers by year

Analyses presented in Appendix C suggest that data points for the number of houses in years 1891, 1892, 1896, 1897, 1913, and 1914, the number of horses in year 1906, and the number of ploughs in year 1916 were likely real outliers. These data points were dropped from the analysis as they lacked support from other data sources.

Other outliers had supporting evidence, suggesting that it would not have been prudent to drop them from analyses. Outliers in agricultural indicators often had support from the weather data. For example, there were 9 outliers in years 1914 and 1915 (Figure 4-1), and all of them, with the exception of the number of houses, were related to agriculture. When environmental data were examined, years 1914 and 1915 turned out to be warmer (Figure 7-41), and wetter (Figure 7-42), than years immediately before and after. Also, agricultural indicators had a similar profile in the Blackfoot Agency (Figure 7-37, Figure 7-38, and Figure 7-39), suggesting that the high agricultural output in years 1914/15 was real, and not attributable to error.

Weather data had support for outliers in the area farmed in year 1920; the number of ploughs in 1920; hay cut in year 1907; wheat harvest in years 1907, 1908, and 1910; and oats harvest in 1910. Other outliers had support from the Annual Reports of Indian Agents, such as the number of oxen in 1896 and oats harvest in year 1916.

TB mortality data had three outliers corresponding to the two early peaks and the subsequent drop in mortality. TB mortality in those years was very high, but not as high as mortality recorded among the Inuit (Grygier, 1994) (p.56), suggesting that it would be more prudent to retain these data points.

The outcome of the outlier analyses was to retain all data points except the outliers in the number of houses, horses, and ploughs mentioned above. With the outlier analysis completed, the next steps included correlation and regression analyses.(DIA., 1888; DIA., 1889)

4.1.3. Correlations

Visual examination of plots of agricultural data series (Figure 7-37, Figure 7-39, Figure 7-40, Figure 7-43) suggest that years 1885 – 1899 and 1900 – 1926 were different. Since the Qu'Appelle Agency had an agrarian economy in a sense that agriculture was the main source of income (Figure 2-9), it makes sense to examine the two periods separately.

Table 4-2: Table of correlations (years 1885-1899)

		Capital			*Agricultural indicators*						
	TB	*Houses*	*Ploughs*	*Stables*	*Potato*	*Hay*	*Wheat*	*Oats*	*Horses*	*Oxen*	*Area*
TB	1										
Houses	0.161	1									
Ploughs	-0.078	**0.68**	1								
Stables	-0.232	0.41	0.138	1							
Potato	-0.06	0.218	**0.48**	-0.4	1						

	8										
Hay	-0.104	-0.149	<u>-0.34</u>	0.098	**-0.42**	1					
Wheat	0.032	0.046	**0.53**	<u>-0.36</u>	**0.7**	-0.107	1				
Oats	0.018	-0.317	0.283	**-0.65**	**0.69**	-0.326	**0.71**	1			
Horses	-0.137	-0.185	-0.017	<u>-0.35</u>	-0.103	**0.44**	<u>0.35</u>	0.237	1		
Oxen	0.111	<u>-0.37</u>	**-0.47**	-0.073	-0.185	0.189	-0.059	0.012	0.180	1	
Area	<u>0.34</u>	-0.226	0.176	<u>-0.39</u>	0.005	-0.105	0.053	0.240	0.211	0.265	1
Expenditures	**0.66**	0.068	**-0.47**	0.13	**-0.6**	**0.41**	-0.20	**-0.61**	0.103	0.213	0.086

Correlation coefficients between TB mortality and other data series are presented in Table 4-2 and Table 4-3. Coefficients significant at 5% are bold, and those significant at 10% are underlined.

TB mortality was positively and significantly correlated with area farmed and expenditures in years 1885-1899 (Table 4-2), suggesting that as one of these indicators increased, so did TB mortality. Since area farmed can be thought of as a proxy for income in an agrarian society, and expenditures represent "supplies for destitute Indians", these results are unexpected given the literature reviewed in Chapter 1.

In addition to examining correlations of various data series with TB mortality, it is also important to examine cross-correlations between these data series, as one data series may impact TB mortality indirectly through another.

The number of houses was positively correlated with the number of stables and ploughs, suggesting that in years 1885-1899 investment in "capital" was broad, and not necessarily limited to one type only. It is also expected, Qu'Appelle Agency being an agrarian economy, that construction of a new homestead would require the construction of stables as well.

There were a number of significant cross-correlations between capital and agricultural indicators. For example, the number of houses was negatively correlated with the number of oxen (Table 4-2), suggesting that in years 1885 – 1899, as the number of houses went up, the number of oxen declined. Since the number of oxen and horses was not correlated, it seems that perhaps oxen were sold to finance the construction of houses.

The number of ploughs was positively correlated with potato and wheat harvests, as expected. It was negatively correlated with amount of hay cut as well as the number of oxen, suggesting that there was a trade-off between different types of agricultural activities. Trade-off between farming and ranching could explain the negative correlation between the number of stables and potato, wheat, and oats harvests, as well as area farmed. The negative correlation between the number of stables and the number of horses is more difficult to explain.

There were also a number of significant cross-correlations between agricultural indicators themselves. For example, potato harvest was negatively correlated with hay cut, but positively correlated with wheat and oats harvests. This supports the previous conclusion that there was a trade-off between various agricultural activities. In this case, farming and cutting of hay. Potato harvest was negatively, significantly, and strongly correlated with expenditures, suggesting that expenditures were likely to decline when potato harvest increased.

As expected, the amount of hay was positively correlated with the number of horses: hay was cut to provide food for horses and horses drew mowers that cut hay. The amount of hay was positively correlated with expenditures, meaning that both data series moved in the same direction.

Wheat harvest was positively correlated with oats harvest and the number of horses, but not area farmed, suggesting that factors other than size of available land drove agricultural output. Like potato harvest, oats harvest was negatively correlated with expenditures, suggesting that an increase in oats harvest was associated with a decline in expenditures.

The pattern of correlations changed in years 1900-1926 (Table 4-3). TB mortality was no longer correlated with expenditures, but instead was positively and significantly correlated with the number of stables, hay cut, oats harvest, and the size of area farmed.

Table 4-3: Table of correlations (years 1900-1926)

		Capital			*Agricultural indicators*						
	TB	*Houses*	*Ploughs*	*Stables*	*Potato*	*Hay*	*Wheat*	*Oats*	*Horses*	*Oxen*	*Area*
TB	1										
Houses	0.140	1									
Ploughs	-0.075	**0.36**	1								
Stables	**0.64**	0.046	**-0.53**	1							
Potato	0.225	0.142	0.176	0.26	1						
Hay	0.32	-0.057	0.076	0.142	0.32	1					
Wheat	0.132	-0.186	-0.28	0.31	-0.094	-0.26	1				
Oats	**0.39**	-0.056	-0.35	**0.73**	**0.34**	**0.47**	0.31	1			
Horses	0.024	**0.36**	0.131	0.250	0.172	-0.085	0.29	0.114	1		
Oxen	0.114	-0.29	-0.235	**0.46**	-0.161	0.024	**0.46**	0.28	0.063	1	
Area	**0.60**	-0.073	-0.186	**0.62**	0.076	0.191	**0.33**	**0.38**	**0.33**	0.208	1
Expenditures	0.092	-0.008	0.068	-0.051	0.184	**0.53**	**-0.55**	0.121	-0.242	-0.044	0.073

Among the capital indicators, the number of houses was still cross-correlated with the number of ploughs, but not with the number of stables. Like in years 1885-1899, the number of houses was still negatively correlated with the number of oxen, but was also positively correlated with the number of horses. The number of ploughs was not correlated with the number of stables in years 1885-1899, but was significantly and negatively correlated with the number of stables in years 1900-1926. The pattern of cross-correlation of the number of ploughs with agricultural indicators also changed. The number of ploughs was no longer correlated with either potato harvest or the amount of wheat, and the sign of correlation with wheat harvest changed from positive to negative (Table 4-3). The number of ploughs was negatively correlated with oats harvest.

The number of stables was significantly correlated with a number of agricultural indicators, but unlike years 1885-1899 when all these correlations were negative, all cross-correlations between the number of stables and agricultural indicators became positive in years 1900-1926. This might mean either that trade-offs between different types of agricultural activities that existed in the 19th century were no longer there in the 20th, or it could mean simply that the role of horses was growing. There is evidence that under pressure from Indian Agents, people of the Agency were switching to using horses when ploughing (DIA., 1901).

Like in years 1885-1899, there were a number of significant cross-correlations between agricultural indicators. Potato harvest was no longer correlated with expenditures or wheat harvest, but was significantly and positively correlated with hay cut and oats harvest. The amount of hay cut was still positively correlated with expenditures, but was no longer correlated with the number of horses. Instead, the amount of hay cut was positively correlated with oats and negatively with wheat harvests.

Similar to years 1885-1899, wheat harvest was positively correlated with oats harvest and the number of horses, but unlike the earlier period, it was also positively correlated with the number of oxen and the size of area farmed. While correlation between wheat harvest and expenditures was negative but not significant, this correlation became significant in years 1900-1926.

Oats harvest was positively correlated with the number of oxen and the size of area farmed in years 1900-1926, and no longer correlated with expenditures. Finally, the number of horses became significantly correlated with the size of area farmed, perhaps consistent with the growing role of horses mentioned above.

Table 4-4: Correlation coefficients between TB mortality and first lags of other data series

	Houses	*Ploughs*	*Stables*	*Potato*	*Hay*	*Wheat*	*Oats*	*Horses*	*Oxen*	*Area*	*Expenditures*
1885-	<u>-0.35</u>	-0.18	-0.17	**-0.47**	<u>0.33</u>	-0.05	0.18	<u>0.36</u>	**0.53**	0.06	0.3

1899											
1900-1926	-0.28	-0.03	-0.19	-0.04	-0.13	0.29	0.03	-0.31	0.19	-0.02	-0.31

While Table 4-2 and Table 4-3 presented same period correlations, lagged values of one data series may be correlated with present values of another data series. Similar to same period correlations, correlations with lagged values of data series are presented for the period 1885-1899 and 1900 – 1026 (Table 4-4). Lagged number of houses and potato harvest were negatively correlated with TB mortality in years 1885 – 1899. TB mortality was positively correlated with lagged number of horses and oxen.

The pattern of correlations with lagged values changed in years 1900-1926. TB mortality was still negatively correlated with lags of the number of houses, but also negatively correlated with the number of horses and expenditures. TB mortality in years 1900-1926 was positively correlated only with lags of the wheat harvest data series.

Relationship between expenditures and agricultural output is an important one to understand. Expenditures were "supplies for the needy", and in the agricultural society such as the Qu'Appelle Agency, income was measured in terms of agricultural output. Correlations between expenditures and lags of agricultural output may provide insights into how Indian Agents responded to changes in agricultural output.

Table 4-5: Correlation of Expenditures with first lags of other data series

	Houses	*Ploughs*	*Stables*	*Potato*	*Hay*	*Wheat*	*Oats*	*Horses*	*Oxen*	*Area*
1885-1899	-0.40	**-0.47**	0.16	**-0.51**	0.27	-0.38	**-0.56**	-0.3	0.26	-0.25
1900-1926	0.06	0.04	0.08	-0.27	-0.01	-0.22	-0.09	**0.40**	0.15	-0.2

In years 1885-1899, lags of the number of houses and ploughs, as well as potato, wheat, and oats harvest were negatively correlated with expenditures (Table 4-5). In years 1900-1926, lags of potato harvest were still significantly and negatively correlated with expenditures. Unlike years

1885-1899 when expenditures were not positively correlated with lags of any data series, in years 1900-1926 expenditures were positively correlated with the lags of the number of horses.

Table 4-6: Correlations of agricultural output with first lags of expenditure data series

	Potatoes	*Hay*	*Wheat*	*Oats*	*Horses*	*Oxen*	*Area*
1885-1899	0.039	0.012	0.065	-0.203	0.222	0.194	**-0.48**
1900-1926	-0.071	0.060	0.074	-0.008	**-0.37**	0.006	**-0.42**

Since expenditures often included agricultural assistance in the form of seeds, understanding how agricultural output responded to lagged expenditures may be important. In years 1885-1899, expenditure data series was negatively correlated with lags of area farmed data series. In years 1900-1926 this negative correlation remained, with the addition of the negative correlation with lags of the number of horses (Table 4-6).

4.1.4. Regression analyses

Results of regression analyses introduced in Section 2.3.2 are presented in the following sections.

4.1.4.1. TB mortality and the SDH

Results of univariate regressions of TB mortality against explanatory variables (Table 4-7) and first lags (Table 4-8) are presented below.

Table 4-7: Summary of univariate regressions

Indicator	R-squared	Parameter estimate	Pr > t
Number of houses	0.009498	0.121921	0.544395
Number of ploughs	2.57E-05	-0.00164	0.975903
Number of stables	0.001712	0.040811	0.825112
Potato harvest	0.000588	-0.0003	0.88039
Hay cut	0.00738	0.00155	0.593296
Wheat harvest	0.001208	6.21E-05	0.829219
Oats harvest	0.048005	8.79E-05	0.180208
Number of horses	0.001797	-0.00864	0.792454
Number of oxen	0.011832	0.013262	0.498421
Area farmed	0.02945	0.003304	0.283338
Expenditures	0.260343	0.000309	0.000654

Results indicate that expenditures, first lags of potato harvest, and the number of oxen were significant in univariate analyses (Table 4-7, Table 4-8). Expenditures and the first lag of the number of oxen have positive coefficients: an increase in those two was associated with an increase in TB mortality. First lag of potato harvest has a negative coefficient, suggesting that an increase in potato harvest was associated with a decline in TB mortality next year (Table 4-8).

The following variables were included in a multivariate regression:

- Variables significant in univariate regressions
- A dummy variable was included to evaluate a potential break in TB mortality data series. The dummy variable, D1, equals zero in years 1885-1899, and 1 otherwise.
- A categorical variable a indicating years of known infectious disease outbreaks (Ferguson, 1928) was included to evaluate the contribution of infectious diseases to TB mortality.
- Two variables for expenditures in years 1885-1899 and years 1900-1926 to evaluate different behaviour of expenditures in the two time periods

Table 4-8: Summary of univariate regressions: TB and 1st lags

Indicator	R-squared	Parameter estimate	Pr > t
Number of houses	0.046134	-0.14932	0.183202
Number of ploughs	7.02E-05	-0.00268	0.960742
Number of stables	0.021234	-0.07959	0.442288
Potato harvest	0.08592	-0.00361	0.062895
Hay cut	0.008323	-0.00091	0.57556
Wheat harvest	0.026557	0.000166	0.315026
Oats harvest	0.001294	1.34E-05	0.830252
Number of horses	0.003138	-0.00634	0.731344
Number of oxen	0.084529	0.019722	0.068755
Area farmed	0.008952	0.001031	0.561428
Expenditures	0.03252	6.05E-05	0.265487

Multivariate regression equation is presented below (Equation 4-1). Model selection was based on adjusted R-squared, Akaike information (AIC) and Bayesian Information criteria (BIC) (Table 7-29). Results of multivariate regression analyses are presented below (

Table 4-9, Table 4-10, Table 4-11).

Equation 4-1: Multivariate regression

$$TBmort = \propto\ + \beta_1 Exp + \beta_2 L.Potato + \beta_3 L.Oxen + \beta_4 D1 + \beta_5 Infect + \beta_5\ Exp1800 + \beta_6 Exp1900 + e$$

where Exp is expenditures, L.Potato is the first lag of potato harvest data series, L.Oxen is the first lag of the number of oxen data series, Exp1800 are expenditures in years 1885 - 1899, Exp1900 are expenditures in years 1900 – 1926, and Infect is a variable indicating years of infectious disease outbreaks (equals 1 in years of recorded outbreaks, 0 otherwise).

Table 4-9: Analysis of variance, Model 1 (Table 7-29)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1064.497	354.8323	5.41	0.0035
Error	36	2359.467	65.54076		
Corrected Total	39	3423.964			

Adjusted R-squared, AIC and BIC suggest that the best model included expenditures in years 1885-1899, first lag of the number of oxen, and infectious disease outbreaks (Table 7-29). Since D1 is not statistically significant (Table 7-31), it was omitted from results presented below.

Table 4-10: Model diagnostics, Model 1 (Table 7-29)

Root MSE	Dependent Mean	Coeff Var	R-Square	Adj R-Sq
8.096	-2.56	-316.2	0.31	0.25

All coefficients are positive, suggesting that an increase in any of these factors led to an increase in TB mortality (Table 4-11). Multivariate regression explained approximately 30% of variation in TB mortality (Table 4-10).

Table 4-11: Parameter estimates, Model 1 (Table 7-29)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	95% Confidence Limits

Intercept	1	-3.44	1.49	-2.31	0.027	0	.	-6.46061
L.Oxen	1	0.02	0.01	2.34	0.025	0.33	0.96	0.003
Infect	1	7.69	3.12	2.46	0.019	0.35	0.97	1.36
Exp1800	1	0.0002	6.66E-05	2.48	0.018	0.34	0.99	2.99E-05

4.1.4.2. Agricultural and cattle production

Drivers of agricultural and cattle production were analyzed to develop rules to determine individual level of agricultural output and the number of horses and oxen in the ABM. Only the final form of regressions are presented here. Details can be found in Appendix C.

Equation 4-2: Oats harvest

$$Oats_t = 12173 + 1066.4 * Stables_t + 8.8 * Hay_t + 108.4 * Precip_t + \varepsilon$$

Equation 4-3: Wheat harvest

$$Wheat_t = 1876.7 - 3.4 * Hay_t + 48.5 * Oxen_t + 39.2 * Precip_t$$

Equation 4-4: Hay cut

$$Hay_t = 0.08 + 0.02 Oats_t - 0.05 Wheat_t$$

Equation 4-5: Number of oxen

$$Oxen_t = 3.99 - 4.5 * Houses_t + 2.6 * Stables_t$$

Equation 4-6: Number of stables

$$Stables_t = -3.3 + 0.01 * Area_t + 1.06 * Houses_t - 0.5 * Ploughs_t + 0.04 * Oxen_t + 0.04 GDD_t$$

, where Stables are the number of stables, Hay is the amount of hay cut, and Precip is precipitation. Heat is the number of heat days; GDD is the number of growing degree days, wheat, oats, and potatoes are respective harvests.

4.1.5. Discussion

Preceding analyses examined correlations, univariate, and multivariate regressions between TB mortality and socioeconomic indicators in years 1885 – 1926. It is clear that the relationship between TB mortality and socioeconomic indicators on the one hand, as well as relationships between socioeconomic indicators on the other, were different in periods 1885 – 1899, and 1900 – 1926. This is clear from the difference in patterns of correlations as well as the results of multivariate regression analysis. It is not unusual for the relationship between TB mortality and socioeconomic environment to change over time: TB incidence rates in Glasgow were correlated with different sets of social factors before and after World War II (Stein, 1952).

TB mortality was correlated in a statistically significant way with several non-lagged (Table 4-2, Table 4-3) as well as lagged data series (Table 4-4).

In years 1885-1899, expenditures were significantly and negatively correlated with the number of ploughs, potato, wheat, and oats harvests, and positively correlated with amount of hay cut (Table 4-2). Relationship between expenditures and other socioeconomic indicators completely changed in years 1900-1926, when expenditures were negatively correlated only with wheat harvest, and still positively correlated with hay cut. Expenditures were no longer correlated with TB mortality in years 1900-1926 (Table 4-3). This result was also supported by the multivariate regression – interaction term for expenditures in years 1885 – 1899 was retained, while interaction term for expenditures in years 1900-1926 was dropped (Table 7-29).

Cross-correlations between various socioeconomic indicators showed that the structure of the economy was changing. While reports of the Indian Agent mention that availability of dry wood and hay prevented engagement in agricultural activity and ranching (DIA., 1902a), it appears that there were trade-offs between various types of agricultural activity. In years 1885-1899, potato harvest was positively correlated with wheat and oats harvest, but negatively correlated with hay cut (Table 4-2). In years 1900-1926 it was positively correlated with hay cut and oats harvest (Table 4-3). The number of stables was negatively correlated with most agricultural indicators in years 1885-1899, but positively correlated with them in years 1900-1926.

Univariate regressions supported the correlation analysis by showing that expenditures, potato harvest, and the number of oxen were related to TB mortality (Table 4-7, Table 4-8).

Multivariate regression added another important result – the impact of infectious diseases on TB mortality. In years where infectious disease outbreaks were recorded, TB mortality was almost 8 deaths per 1,000 higher (Table 4-11) than in other years.

Despite these changes in the socioeconomic environment of the Qu'Appelle Agency, TB mortality was positively correlated with socioeconomic indicators in both periods, suggesting that changes in the socioeconomic environment did not change the relationship between TB mortality and that environment. Since a positive relationship between TB mortality and socioeconomic indicators that persisted for 40 years is unexpected, it is possible that TB mortality was driven by another factor that was not included in these analyses. One of such factors could be susceptibility to TB.

Another possibility is that socioeconomic factors examined here actually accounted, at least in part, for variation in TB mortality, in which case the challenge is to understand how that relationship could have been positive.

In years 1885 – 1899, expenditures were negatively correlated with oats and potato harvest, number of ploughs, and negatively, but not significantly correlated with wheat harvest. This result suggests that a reduction in these indicators was associated with an increase in expenditures. Increase in expenditures in turn was correlated with increase in TB mortality (Table 4-2). Given that the Agency could not produce a sufficient amount of food in the early years and the level of agricultural production was low compared to the 1900-1926 period, a reduction in expenditures in years of good harvest may have put pressure on the people of the Agency. Expenditures would increase in years of poor harvest, but this increase took place in the context of declining expenditures (Figure 7-52) and the general push to cut costs by the DIA. Given that it is known that DIA would withhold rations even during hunger (J. Daschuk, 2013), it is possible that increase in expenditures was either insufficient or would come late, particularly that there is a negative correlation between lags of agricultural indicators and expenditures (Table 4-5), suggesting that this relationship persisted for more than a year.

In addition to agricultural output and expenditures, housing played a role in the spread of TB in the Agency. The first lag of the number of houses was negatively correlated with TB mortality in years 1885-1926 (Table 4-8). This suggests that an increase in the number of houses in one year was associated with a reduction in TB mortality the following year. Increasing the number of houses as well as quality of houses resulted in reduction of TB mortality. This is expected, as lower housing density is related to lower TB incidence (M. Clark et al. 2002; Stein, 1952).

It appears then, that in years 1885-1899 interaction of expenditures and agricultural output were driving increases in TB mortality, while housing construction acted to reduce it.

In years 1900-1926 TB mortality was significantly and positively correlated with the number of stables, hay cut, and area farmed (Table 4-3), suggesting that TB mortality and these indicators moved in the same direction. TB mortality was negatively correlated with first lags of the number of houses and expenditures (Table 4-4). This result suggests that an increase in the number of houses or expenditures was correlated with a reduction in TB mortality next year.

The relationship between agricultural output and expenditures continued to be important as many agricultural indicators were positively and significantly correlated with TB mortality (Table 4-3). Relationship between expenditures and agricultural indicators (Table 4-3) as well as their lags (Table 4-5) do not have a clear pattern. Table 4-6 suggests that agricultural output and lags of expenditures moved in different directions: a reduction in expenditures was followed up by an increase in area farmed and the number of horses.

It is possible that in years 1900-1926 agricultural output followed cyclical fluctuations which were more significant than in years 1885 – 1899 simply because the level of agricultural activity was much higher. TB mortality may have fluctuated with fluctuations in the agricultural output, increasing after a year of poor harvest, and declining after a good one. The negative correlation between TB mortality and lag of expenditures (Table 4-4) suggests that expenditures may have played a role in cyclical fluctuations of agricultural output – either contributing to it, as some of the expenditures took form of seeds, or not providing enough of support in years of poor harvest.

It must be kept in mind that even after the start of agricultural growth in the Qu'Appelle Agency, average income in the Agency was still far below average for Canada (see (Urquhart, 1986)), meaning that a cut in expenditures must have had an impact on the people of the Agency. Bearing that in mind, it is possible that the following relationship existed between expenditures, agricultural output, and TB mortality. A cut in expenditures, possibly because of a good crop

(Table 4-3, Table 4-5), was correlated with increased TB mortality in the following year (Table 4-4), as even a good crop would not necessarily last the whole year. At the same time, a cut in expenditures was correlated with expansion of area farmed, as people of the Agency had to replace the drop in their incomes (Table 4-6). Expansion of area farmed was positively correlated with most agricultural indicators, which, in turn, were positively correlated with TB mortality. In this scenario, positive correlation between TB mortality and agricultural indicators was an outcome of two impacts of the drop of expenditures – direct impact on TB mortality through reduction in nutritional levels, and indirect one through agricultural expansion in response to reduction in nutritional levels.

4.1.6. Summary

TB mortality and socioeconomic data analysed in this section contain a number of potential outliers. While statistical methods can help identify outliers, decisions on what to do with them lie in other domains. Many of potential outliers can be explained or supported by other data. For that reason they were retained for subsequent analysis. Some of the outliers, notably in the number of houses and horses time series, were dropped.

Correlation analyses demonstrated that correlations differed between years 1885-1889 and 1900-1926, suggesting that the economic reality was different in the two periods. Lags of the number of houses was negatively correlated with TB mortality in both periods, suggesting that an increase in the number of houses was correlated with subsequent reduction in TB mortality.

TB mortality was significantly and positively correlated with a number of indicators of agricultural output and expenditures. This unexpected result can be explained by the cyclical nature of agriculture and the nature of expenditures which were reduced after a good crop.

Multivariate regression analysis supported the earlier conclusion of a difference between years 1885-1899 and 1900-1926 and introduced another important predictor of TB mortality – infectious disease outbreaks.

These results indicate the interconnectedness of external factors, such as climate, with the SDH in the form of agricultural output, policy decisions, and health. The endogenous and dynamic nature of these relationships may be more easily analysed using alternative methods that preserve these dynamic interactions. This can be done using agent-based modelling.

4.2. Agent-based modelling

The following sections present results of the ABM. Section 4.2.1 presents results of the base-case scenario. Section 4.2.2 presents uncertainty analysis. Section 4.2.3 presents sensitivity analyses. Section 4.2.4 presents analysis of the micro data.

4.2.1. Base-case scenario

Base-case scenario presents the model output given model parameters specified in Table 3-2, Table 3-3, and Table 3-4.

4.2.1.1. Area farmed

Given the difference in population size between the actual and the simulated population of the Agency (Table 3-2), all model outputs should be lower as well. As expected, simulated area farmed under the base case scenario is lower than the actual size of area farmed. Simulated area farmed is less than 1/7th of the actual area farmed, suggesting that the model underestimated the simulated area farmed. Given that the initial number of farmers and the size of their plots of land are not known, this result is not surprising. At the same time, the shape of the simulated area farmed was a close match to the actual area farmed. The base case scenario resulted in area under the curve (AUC) of 0.03 and the ordinal pattern analysis (OPA) indicator of 0.54.

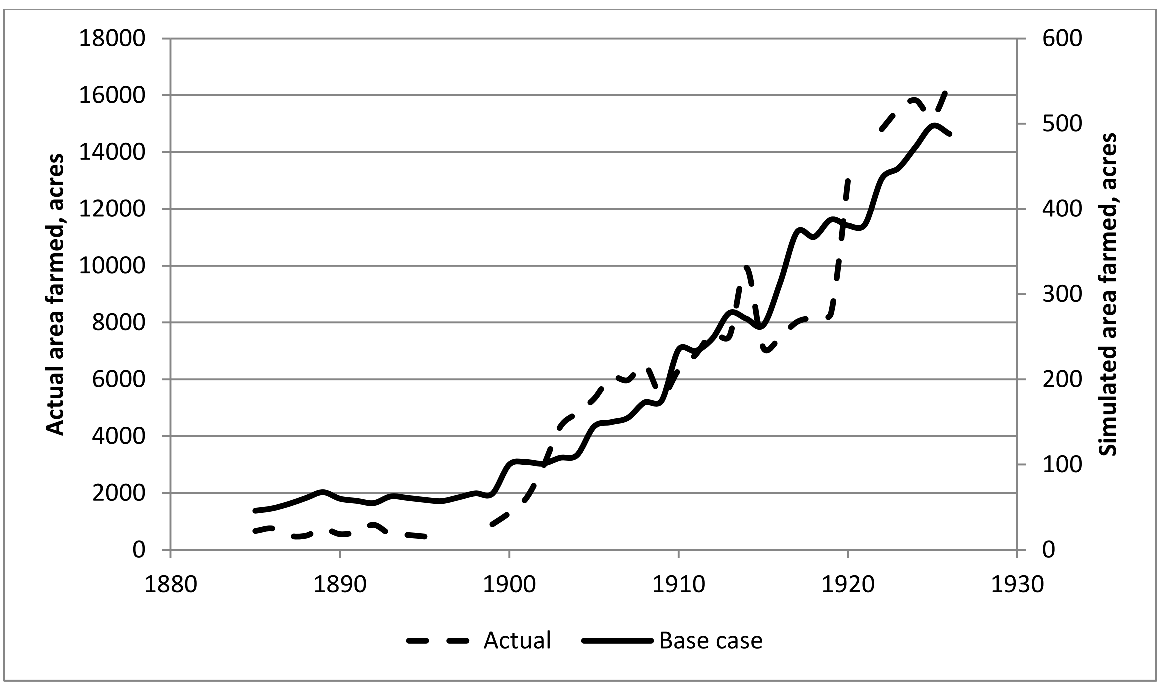

Figure 4-2: Area farmed

Given differences in population sizes between the actual population of the Agency and the one used in base case scenario, AUC of area farmed should have been about 0.14, suggesting that the base case scenario underestimated the area farmed. The OPA indicator of 0.54 suggests that the simulated area farmed correctly reproduced 54% of the ups and down of the actual data.

4.2.1.2. Wheat harvest

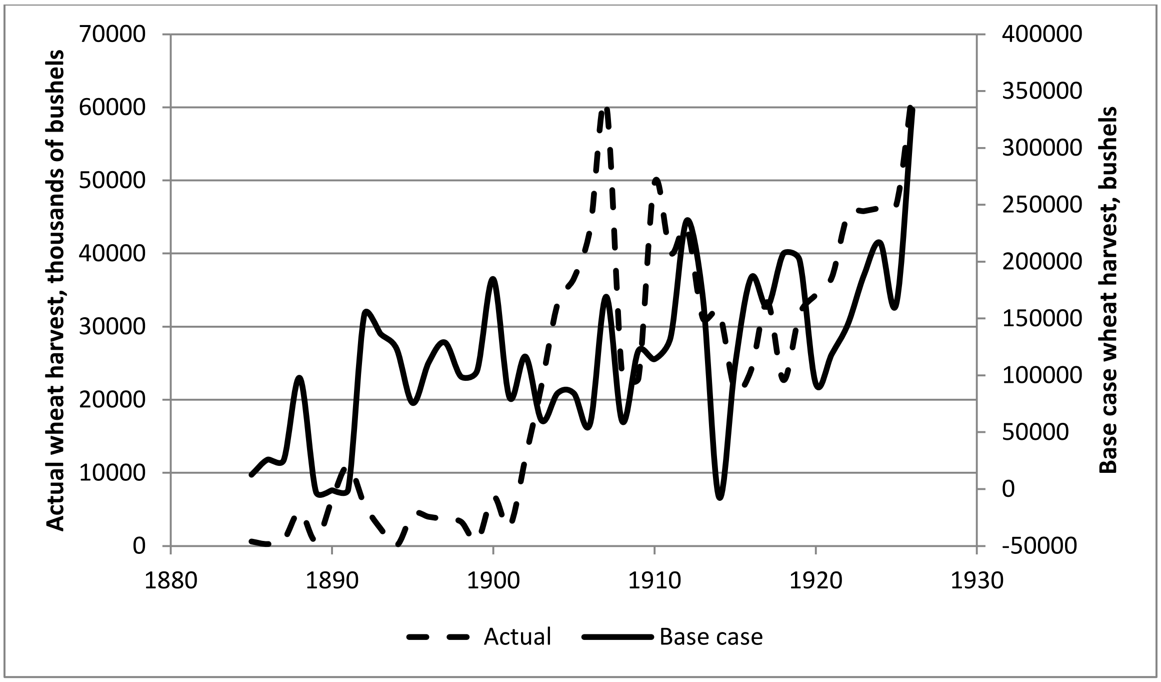

Figure 4-3: Wheat harvest

Unlike simulated area farmed, simulated wheat harvest does not have a particularly good fit in years 1885-1900. This result stresses the fact that years 1885-1899 and 1900-1926 were very different. Simulated wheat harvest is a function of precipitation (Table 3-4) among other factors. Given that weather data for years 1889-1901 were missing (Table 3-1), a relatively poor fit in years 1885-1899 could be expected. The fit of the base case scenario is improved in years 1900-1926, when weather data are available. The implication of the poor fit of the base case scenario in years 1885-1899 is that the simulated income of the Agency was higher than the actual one. Since agricultural output formed a major part of income, higher income would be expected to lead to lower TB mortality rates.

Base case scenario produced the AUC of 5.5, suggesting that simulated wheat harvest significantly overestimated the actual one. Simulated wheat harvest had the OPA indicator of

0.46, suggesting that the ups and downs of the simulated wheat harvest matched those of the actual data about 46% of the time.

4.2.1.3. Oats harvest

Like wheat harvest, oats harvest is a function of weather (Table 3-4), and the fit in years 1885-1899 suffers because of missing weather data. The fit of the base case scenario is again improved in the years 1900-1926, reproducing the significant swings of the original data. Like the wheat harvest, the higher than expected oats harvest in years 1885-1899 could lead to the lower simulated TB mortality rate.

An alternative simulation produces a significant improvement over the base case scenario (Figure 4-4). Alternative definition of oats harvest is conditional on choice of statistical methods to transform the raw data (Appendix D). Area farmed, included in alternative definition of oats harvest, improved model fit in years 1885-1899 and alignment of the peaks of oats harvest in years 1900-1926. Base case scenario produced the AUC of 0.05 suggesting that the simulated oats harvest underestimated the actual one. The OPA indicator of 0.44 suggests that the ups and downs of the simulated oats harvest correctly matched about 44% of the actual ones.

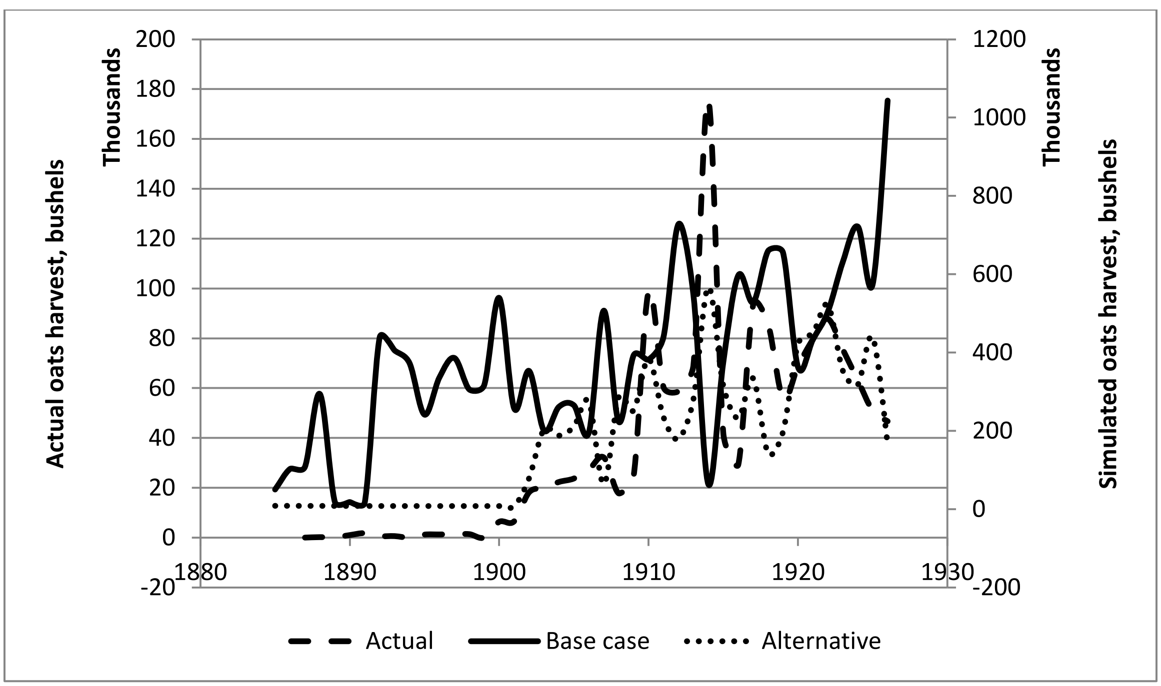

Figure 4-4: Simulated and actual oats harvest

4.2.1.4. *Number of oxen*

Unlike wheat and oats harvests, the simulated number of oxen has a better fit in years 1885-1899 than in years 1900-1926. A lower than expected number of oxen in years 1900-1926 leads to a lower than expected income and potentially increased TB mortality. An alternative definition of the number of oxen (see Appendix D) improved model fit in years 1900-1926. The base case scenario produced an AUC of 0.73, suggesting it overestimated the actual number of oxen. The base case scenario correctly matched approximately 42% of ups and downs of the actual data.

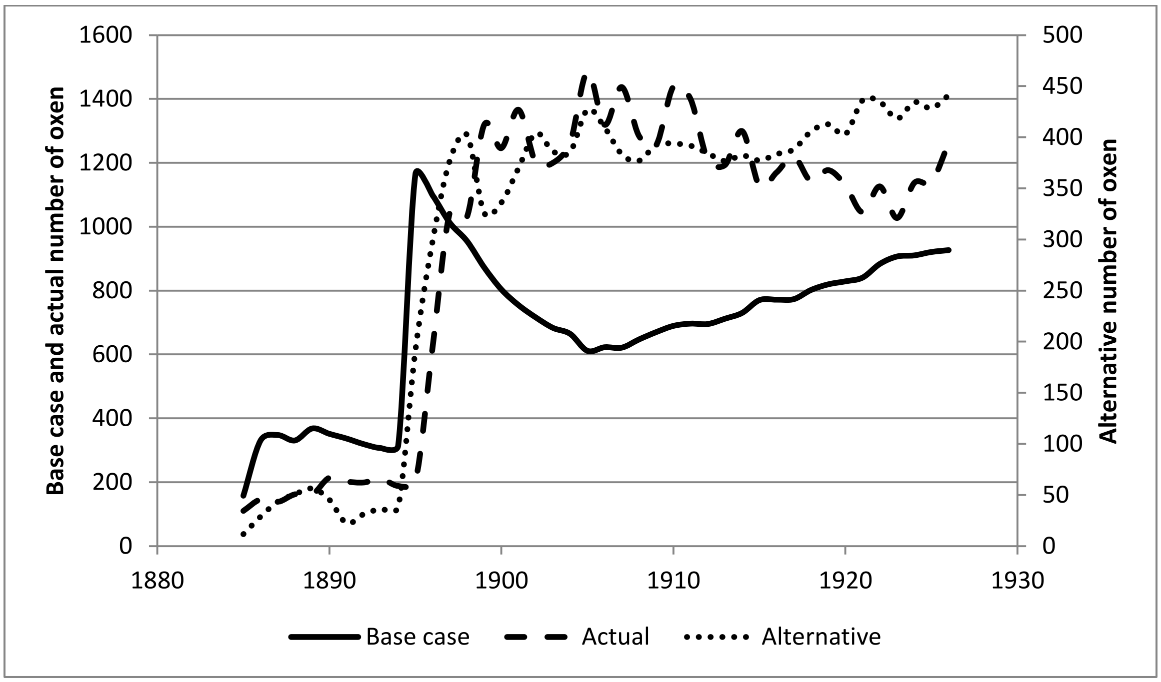

Figure 4-5: Simulated and actual number of oxen

4.2.1.5. *Hunger*

Correlation and regression analysis demonstrated that correlation between government expenditures "on destitute Indians", or rations, and their interactions with agricultural production likely played a role in driving TB outbreaks. ABM allows one to estimate the prevalence of hunger given agricultural production levels simulated by the model, assumptions about the cost of food, and historical level of government expenditures, or rations. Taken together, these three elements make it possible to define hunger as a state in which individual is not able to cover the cost of food and the government rations are no longer available. In the subsequent graph and analyses, hunger was defined as a number of people in this condition, while TB prevalence was defined as a proportion of active and latent TB cases in the population.

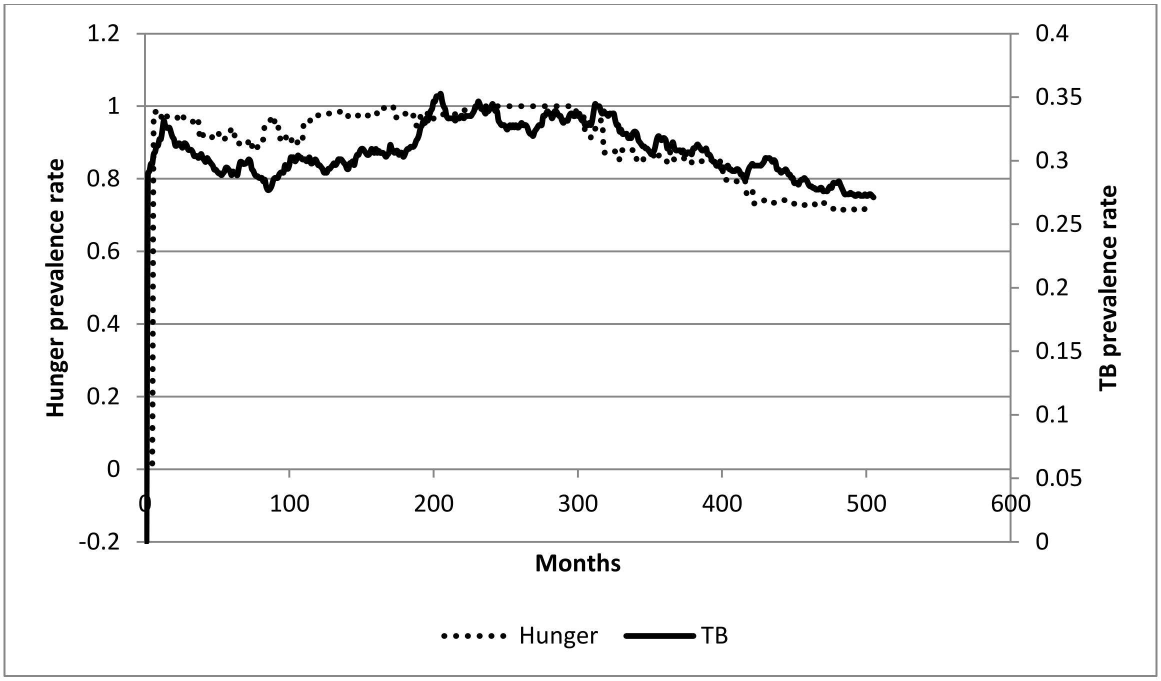

Figure 4-6: Simulated monthly hunger and TB prevalence rates

The two data series are highly correlated with a correlation coefficient of 0.68. A linear regression of TB prevalence against hunger suggest that hunger alone can explain about 47% of variation in TB prevalence (Table 4-12, Table 4-13, Table 4-14).

Table 4-12: TB and hunger prevalence regression statistics

Multiple R	0.683946
R Square	0.467782
Adjusted R Square	0.466714
Standard Error	0.0145
Observations	500

Table 4-13: ANOVA statistics, TB and hunger prevalence regression

	df	SS	MS	F	Significance F
Regression	1	0.092034	0.092034	437.7074	3.26E-70
Residual	498	0.104711	0.00021		
Total	499	0.196746			

Table 4-14: Regression coefficients, TB and hunger prevalence regression

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.19	0.0057	32.76	2.4E-126	0.17	0.19	0.17	0.19
Hunger	0.13	0.0063	20.9	3.26E-70	0.12	0.15	0.12	0.15

4.2.1.6. Population of the Agency

Population of the Agency is a summary indicator that depends on inputs, such as TB mortality, birth rate, and income, among others. The base case scenario overestimates the population of the Agency (Figure 4-7). An alternative model run presents a much closer fit to the simulated population than the base case. The reasons for this will be discussed below. The base case scenario produced an AUC of 0.36, suggesting that it overestimated the actual population, and correctly matched approximately 46% of ups and downs of the actual data.

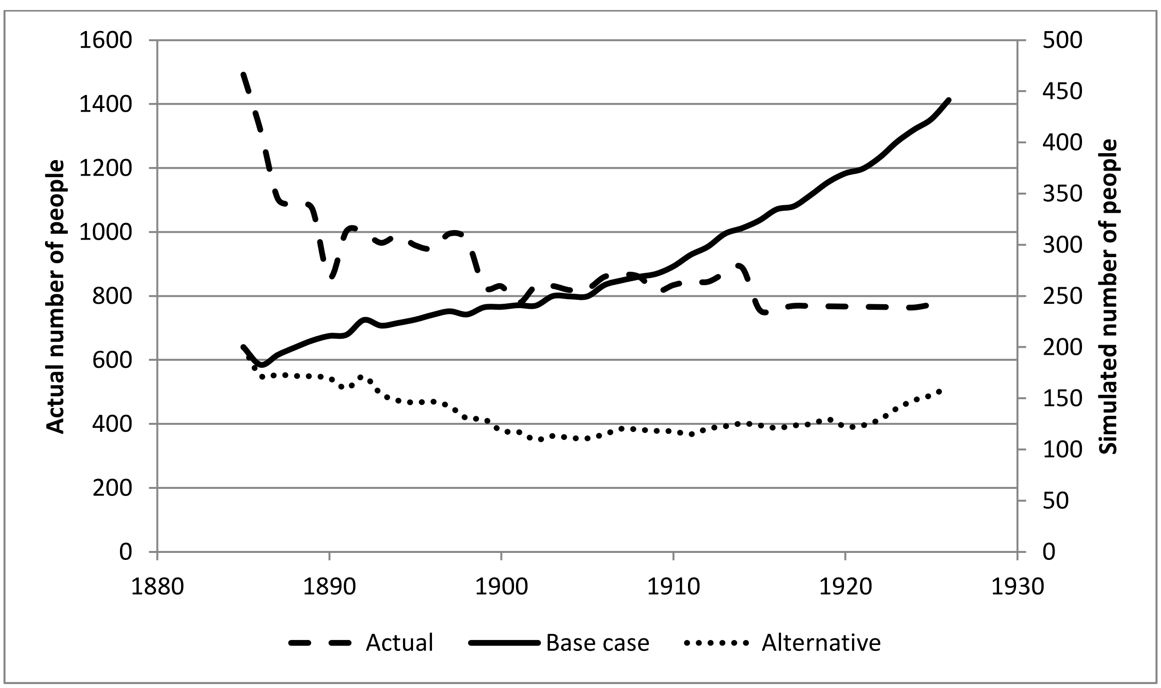

Figure 4-7: Actual and simulated population of the Agency

4.2.1.7. TB mortality

The base case scenario significantly underestimated TB mortality in the Qu'Appelle Agency. While the peaks of the simulated TB mortality are often aligned with those of historical data, the overall level of TB mortality is too low (Figure 4-8). An alternative model run presents a much closer fit to the historical TB mortality. This will be discussed further below. The base case scenario produced an AUC of 0.12, suggesting that it underestimated TB mortality. The base case scenario correctly matched approximately 54% of ups and downs of the actual TB mortality. Alternative representation of TB mortality based on maximum values of TB mortality in a given year attained an AUC of 0.42, which still underestimated the actual TB mortality.

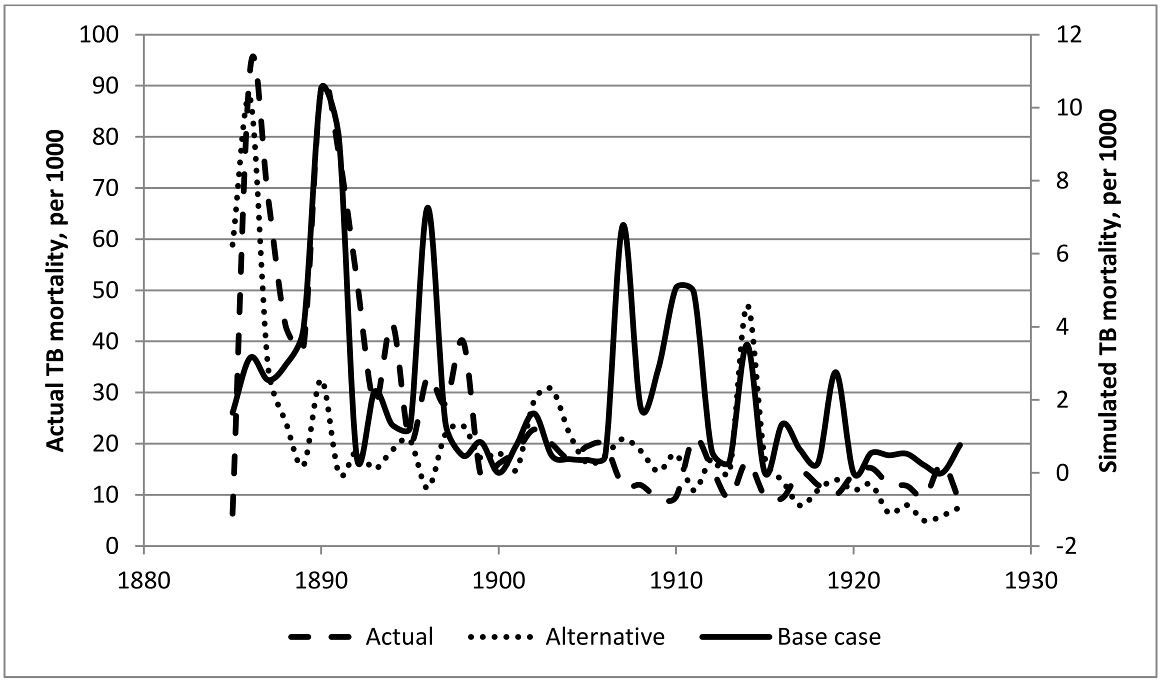

Figure 4-8: Actual and simulated TB mortality

4.2.2. Uncertainty analysis

Uncertainty analysis evaluates uncertainty in the model output that is generated by uncertainty in model input parameters (Marino et al., 2008). This ABM is a stochastic model, and as such it has inherent randomness. Figure 4-9 and Figure 4-10 present results of repeated sampling of input parameter values from a distribution specific to each parameter. In this analysis, 30 samples were drawn using ranges presented in Table 3-2, Table 3-3, and Table 3-4.

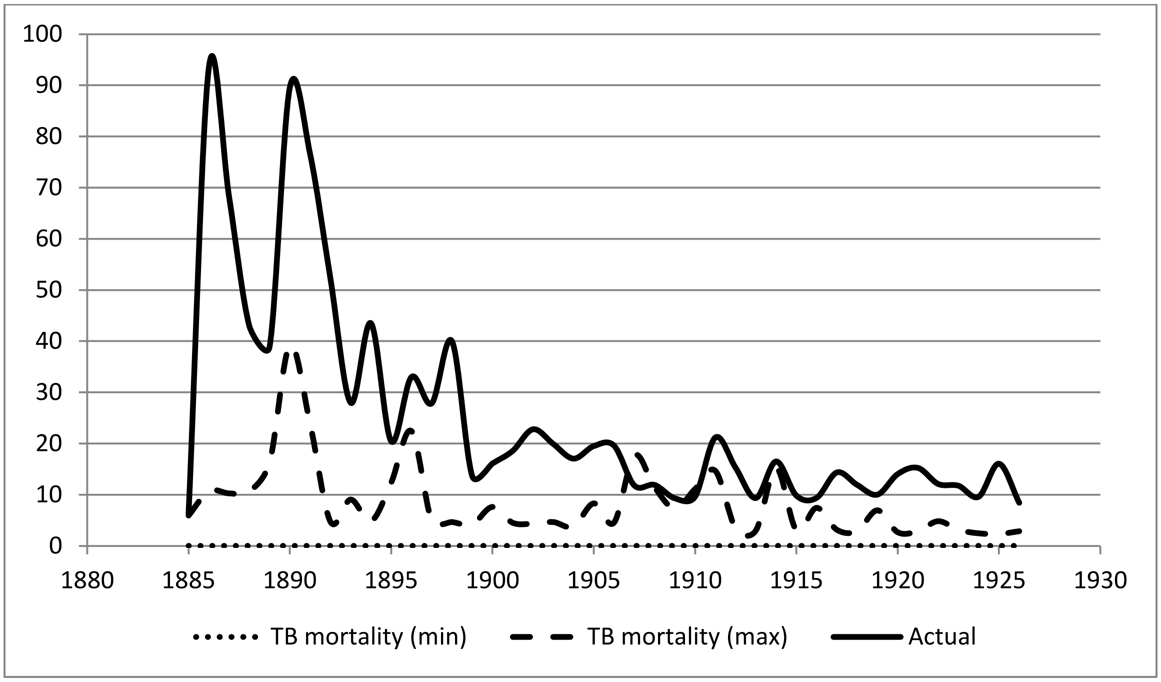

Figure 4-9: Simulated lowest and highest TB mortality

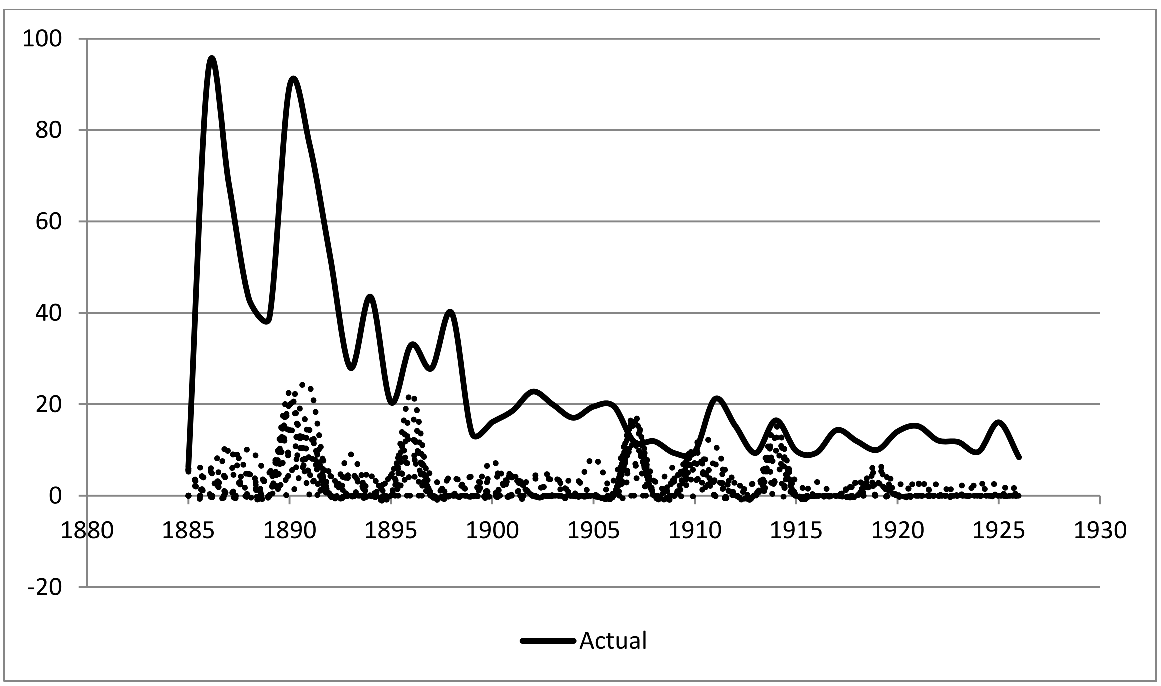

Figure 4-10: Measure of uncertainty: actual and simulated TB mortality

As Figure 4-9 demonstrates, there is a significant variation in model outputs across model runs. In a given year, the difference between the lowest and the highest TB mortality across the 30 samples can be very significant: in year 1890 the lowest TB mortality across model runs was 0, while the highest TB mortality was 39.2. Maximum values of simulated TB mortality have a better fit than the base case scenario (Figure 4-9).

Figure 4-10 presents a measure of uncertainty: there is a significant variation in simulated TB mortality across model runs. The next section evaluates contribution of individual model parameters to the overall model uncertainty.

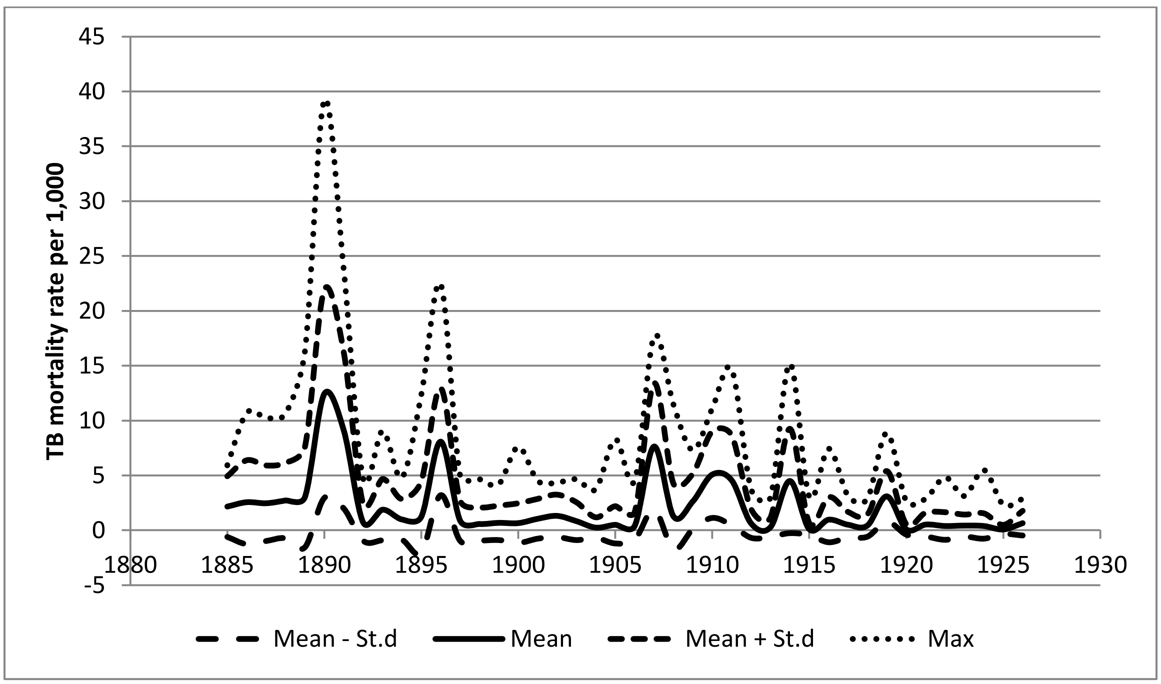

Figure 4-11: Simulated TB mortality rate: mean, +/- standard deviation

4.2.3. Sensitivity analysis

P-values of the PRCCs of selected variables and model outputs are presented in Appendix E.

Table 4-15: Summary of sensitivity results: number of significant p-values by variable and model output

	Active TB	Latent TB	Area	Oats	Wheat	Oxen	Population	TB mortality	Row Total
pInf			2			4	1	1	**8**
pPrg				1	2			9	**12**
Suscept.	10		4			23	1		**38**

prevL								2	**2**
prevA	2			4	4			1	**11**
Tbcmin			1					7	**7**
Tbcmean	3			22	27			1	**56**
Tbcmax	10	2	2	1	1	1	1	1	**19**
cFood								3	**3**
Column total	**25**	**2**	**9**	**28**	**34**	**28**	**3**	**25**	

Table 4-15 presents the number of significant p-values for PRCCs between selected variables and model outputs. The number of significant p-values can be greater than one, as a p-value was calculated for each year of model output. For example, TB susceptibility was significantly correlated with the number of active cases of TB in 10 out of 42 years (Table 4-15).

Among the variables (row totals), the mean number of TB contacts has the strongest correlation with model output with 56 significant p-values, followed by susceptibility to TB with 28, and the maximum number of TB contacts with 19 significant p-values. The prevalence of latent TB in 1885 had the fewest significant correlations. These results stress the importance of contact network and biological factors in the spread of TB. On the model output side (column totals), agricultural indicators (wheat and oats harvest, the number of oxen) followed by the number of active cases and TB mortality had the most significant correlation with model parameters.

TB mortality was positively correlated with the probability of disease progression and the cost of food: as one of these parameters increased, so did TB mortality (Table 4-16). This result is both intuitive and in line with previous findings regarding the impact of hunger on TB mortality.

Table 4-16: Sign of significant correlations

	Active TB	Latent TB	Area	Oats	Wheat	Oxen	Population	TB mortality
pInf			-			-	+	+
pPrg				+	+			+
Suscept.	-		+ (-)			-	+	
prevL								+
prevA	-			-	-			-
Tbcmin			+					-
Tbcmean	+			+	+			-
Tbcmax	-	-	+	-	-	-	+	-

cFood								+

Correlation between TB mortality and the minimum number of TB contacts was negative, suggesting that as the minimum number of TB contacts decreased, TB mortality increased. This is not immediately intuitive, as a lower minimum number of TB contacts reduces TB transmission opportunities, which should result in a lower TB mortality.

The number of cases of active TB was negatively correlated with susceptibility to the disease, prevalence of active TB in 1885, and maximum number of TB contacts (Table 4-16). The number of latent cases was negatively correlated to the maximum number of TB contacts.

4.2.4. Policy analysis

While preceding sections presented results of the base case scenario, the following section presents model application to policy analysis. As was discussed above, hunger significantly contributed to the burden of TB in the Qu'Appelle Agency. It also presented an opportunity to evaluate policy interventions, particularly so because DIA had a policy on hunger management – it was distributing rations. The amount of rations distributed declined over the years, as the government was interested in cutting costs. This policy had a significant and dramatic impact on mortality in the Prairies (J. Daschuk, 2013).

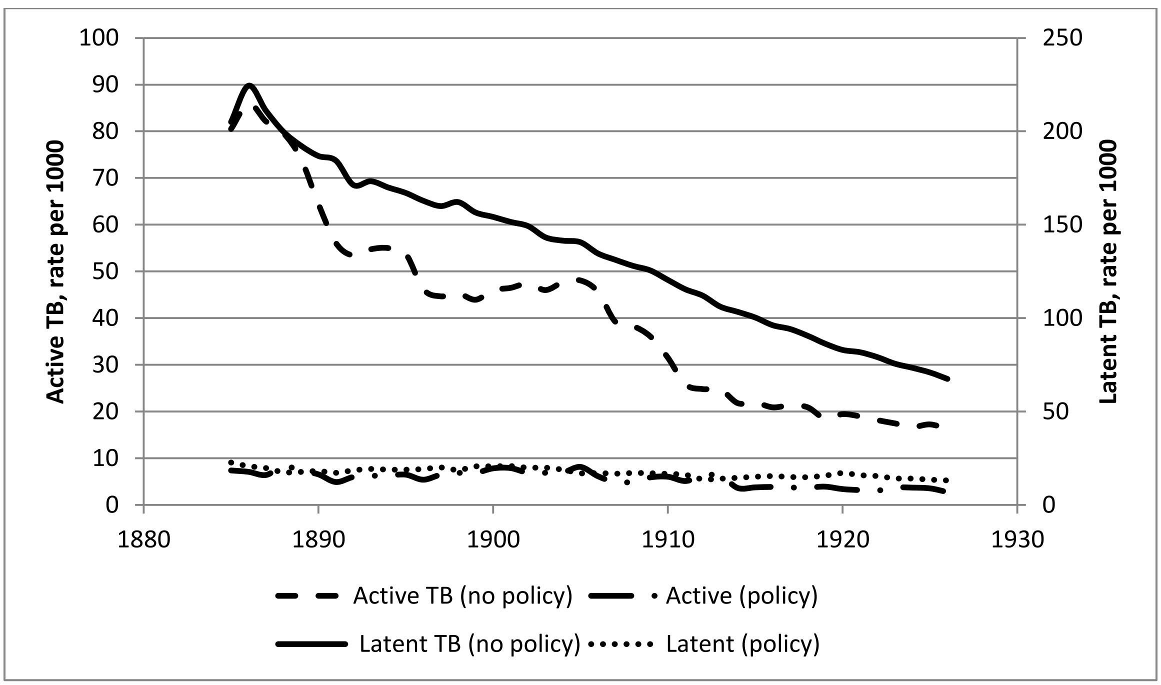

Figure 4-12: Prevalence of latent and active TB with and without anti-hunger policy.

ABM allows one to evaluate the cost of eliminating hunger and the impact of such a policy on TB prevalence by not capping the rations at their historical maximum. In this model run, anyone who could not afford the daily cost of food would be provided food even after the historical level of rations has been reached. Based on this run of the model, the prevalence of both active and latent TB is much lower under the anti-hunger policy than without it (Figure 4-12).

The total amount of per capita funds required to eliminate hunger varied from a low of $46.65 in 1901, to a high of $79.50 in 1925. Given that the DIA was spending on rations, additional per capita expenditures varied from a low of $38.50 per year, to a high of $77.50 per year in 1902 prices. In 2016 prices, $77.05 would equal $2,171.5 (Statistics Canada, 2014).

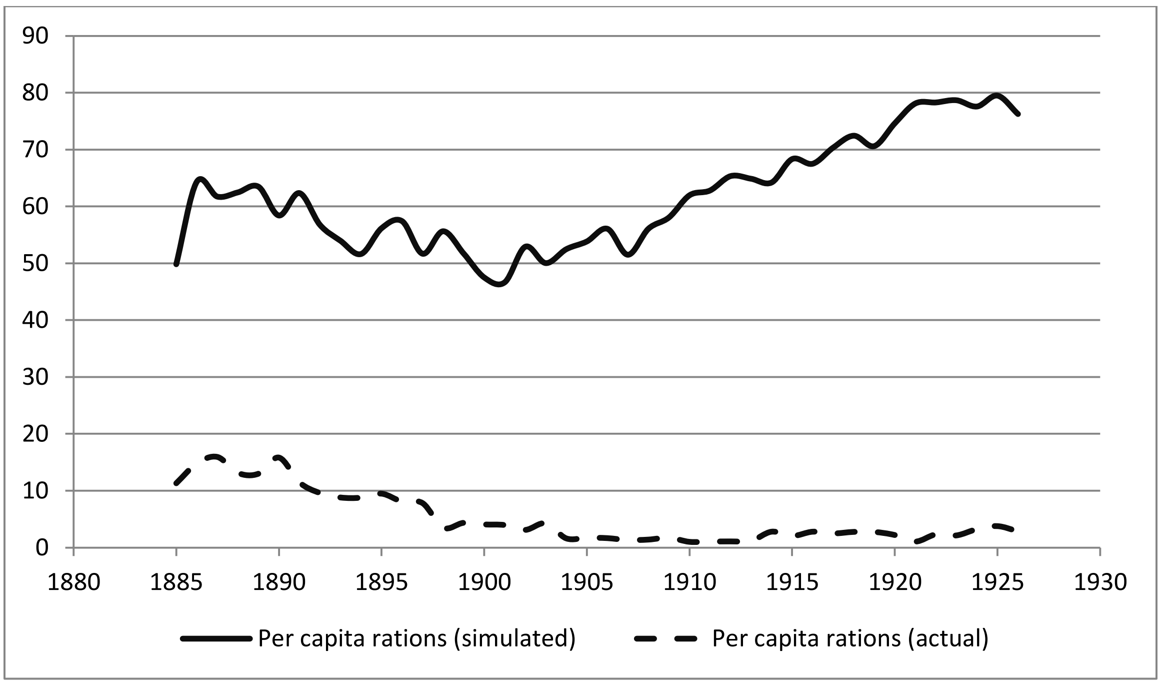

Figure 4-13: Estimate of the per capita cost of eliminating hunger

While the DIA did not spend sufficiently to eliminate hunger, it was providing rations, albeit a very small amount. To put that amount in perspective, there were years in which DIA spent just over a dollar per person, while its employees were allocated $0.33 per day for food (Table 3-4). Without that expenditure, TB mortality would have been even higher (Figure 4-14).

4.2.5. Discussion

The base case scenario presented here simulated a part of the socioeconomic environment of the Qu'Appelle Agency, namely its agricultural output, population size, and TB mortality. The base case scenario relied on model parameters presented in Table 3-2, Table 3-3, and Table 3-4. Results of the base case scenario show that there are differences between how well various model outputs fit historical data. Some, such as area farmed, fit well. Others, such as oats and wheat harvest, fit better in years 1900-1926 than in years 1885-1889. Still others, such as population, do not fit particularly well.

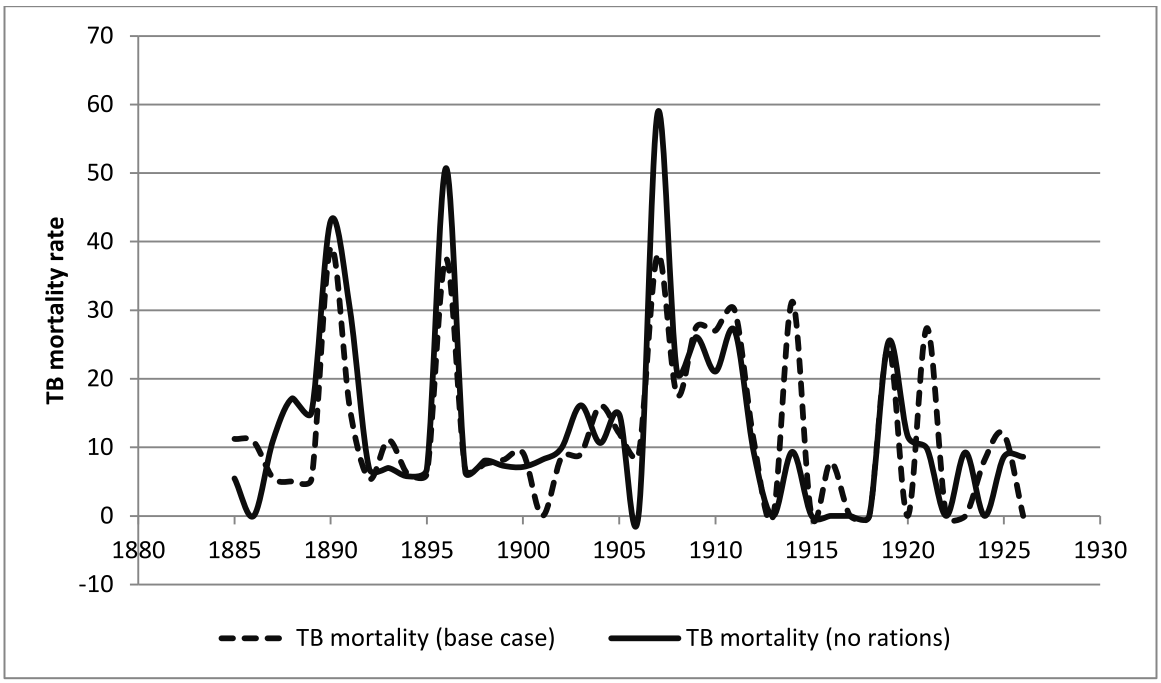

Figure 4-14: TB mortality rate with and without rations

Identifying sources of problems in an ABM is a complex task because of the interdependences of various parts of the model. This means that a change in one relationship precipitates a cascading effect through the model, affecting many other outputs. For example, the way a number of stables is simulated will impact agricultural output, which will in turn affect income and, ultimately, TB mortality and population counts.

The base case scenario had a low TB mortality. As was mentioned above, it could in part be explained by the higher than expected wheat and oats harvests. At the same time, there might be more directly relevant reasons for this low mortality, particularly given that alternative model runs can produce a much better fit (Appendix D).

The simulated prevalence of both active and latent TB was declining from 1885 to 1926 (Figure 4-12). This would be expected, as TB mortality was also declining in general throughout this period. There are no historic data on latent and active TB for the whole population of the Qu'Appelle Agency, but medical evaluations of students in boarding schools conducted by Ferguson showed that there was almost 100% prevalence of TB infection by age 15 (Figure 2-18).

Figure 4-15 demonstrates that the simulated incidence of active TB was often zero; suggesting that little TB transmission was going on. Low TB transmission could occur for at least two reasons:

- Low TB transmission probability
- Small number of transmission opportunities

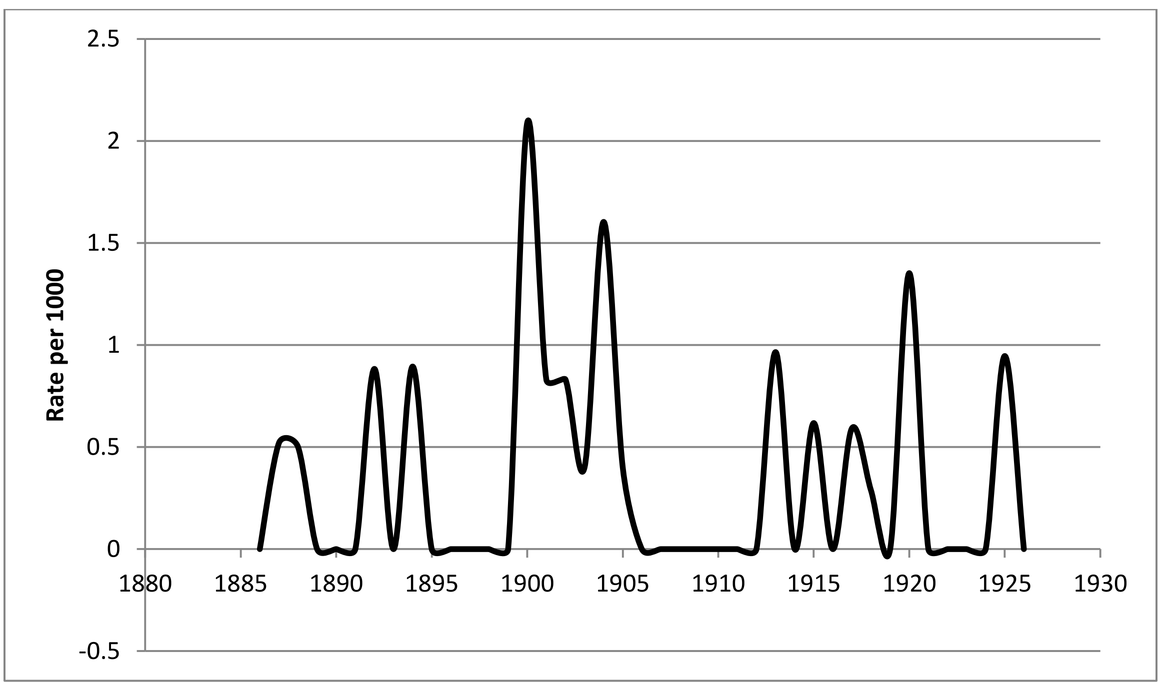

Figure 4-15: Simulated incidence of active and latent TB, rate per 1000.

TB transmission probability was modelled as a product of the baseline rate and a number of risk factors such as susceptibility to TB, low SES, malnutrition, and housing density (Table 3-2,

Table 3-3, Table 3-4). The baseline TB infection rate comes from Glasgow (Table 3-2) and therefore could be too low. The baseline TB transmission rate is modified by a number of risk factors mentioned above. Estimates of these risk factors come from studies that were published in the late 20th century, and mostly apply to the general population. As such, these estimates may be too low for the realities of the late 19th century Prairies. An example may be required to illustrate this point. The odds ratio for TB transmission given the above-average number of people in the house came from 1930s Glasgow. 1930s were relatively close to the study period considered here, yet the housing quality in 1930s Glasgow cannot be compared to the housing quality in the 19th century Qu'Appelle Agency. For this reason, this odds ratio might be too low.

Low TB transmission could also occur because of a low number of people with active TB. One of the reasons this could occur is that the estimate of the prevalence of TB susceptibility came from a modern population. Since TB susceptibility leads to higher mortality, prevalence of TB susceptibility is likely to have been higher in the past.

The second potential reason for the low incidence rate is the low number of transmission opportunities. This could occur for at least two reasons. First of all, the number of TB contacts used in this simulation came from a TB outbreak investigation in a First Nations community (Table 3-2). With zero as a minimum number of contacts and 2 as a mean number of contacts, most people have a low chance of transmitting TB. The number of TB contacts may have been higher in the 19th century Qu'Appelle Agency, given inferior housing among other reasons.

While contact network parameters, some disease transmission parameters and odds ratios used to model disease transmission came from the Indigenous context, this context might have been different from that of the 19th century, and resulted in a low TB mortality. ABM is a stochastic model with a variation in model output that can be attributed to the uncertainty in model parameters as well as the stochastic nature of the model. When only highest simulated TB mortality levels are retained, the model output has a much better fit to the actual data. This suggests that it is possible that ranges of parameter values obtained from recent studies cover the "true" but unknown parameter ranges appropriate to the Qu'Appelle Agency, but these "true" parameter values might be located close to the maximums of "modern" parameter ranges.

This suggests that a model based on contemporary parameters may be appropriate, but the focus should not be on model output averaged over many runs as is usually the case with ABMs, but rather on model runs with output located at the extremes of relevant distributions. In the case of TB mortality that would be model runs with high mortality. In the case of agricultural output that would be model runs with low output.

The second reason for the low number of transmission opportunities comes from limitations of the current model. The model does not include instances related to cultural practices of the Indigenous people that may have presented additional opportunities of transmission. Such instances include pipe sharing ceremonies and sweat baths.

Hunger is a summary indicator that significantly increases the likelihood of developing active TB given infection (Cegielski et al., 2012). Hunger was defined as not being able to cover the cost of food once the historical amount of rations for that particular year ran out. It is a composite indicator that depends on a number of parameters, such as the agricultural output, welfare support, and cost of food. In these analyses cost of food estimate was obtained from a budget allocated to workers employed by the DIA. Given its source, this estimate might be higher than the actual cost of food in the Agency, particularly given the fact that the people of the Agency could and did hunt and fish to supplement their diet. On the other hand, access to food, like access to everything else, was regulated by the Indian Agent. Since people of the Agency had to seek permission of the Indian Agent to sell and purchase goods, there were non-monetary barriers to access food.

The number of people in the state of hunger fluctuated throughout the year. This is to be expected, as harvest would reduce the number of people in that condition. Clearly, the definition of income is critical. Simulated income included major components, such as agricultural production and value of beef (Figure 2-9), but not smaller income categories, such as the value of meat and fish provided by hunting. Together with the definition of the cost of food, this suggests that the number of people in the state of hunger might have been overestimated.

Hunger was a significant contributor to the burden of TB in the Agency. It was significantly and positively correlated with the prevalence of TB. Regression analysis demonstrated that hunger explained approximately 47% of total variation in the prevalence of TB (Table 4-14).

Model output shows significant variability from one run to another, typical of stochastic models. Sensitivity analyses identified parameters with a significant impact on model output: probability of disease progression and the mean number of TB contacts. While these two parameters did not have a significant impact on TB mortality, they were significantly correlated with other model outputs. This result is not surprising, given that PRCCs were used in sensitivity analyses. These correlation coefficients estimate the strength of relationship between two variables net of other influences. It is possible that susceptibility to TB and the mean number of TB contacts impact TB mortality through other model outputs, such as the number of active TB cases.

TB mortality was significantly correlated with the probability of disease progression, the minimum number of TB contacts, and the cost of food. It was less strongly correlated with all other model parameters examined in sensitivity analyses, with the exception of susceptibility to TB. TB mortality was positively correlated with the probability of disease progression and cost of food, as would be expected, particularly in light of the discussion of hunger above. It was negatively correlated with the minimum number of TB contacts. This result is not intuitive.

The Mean and maximum numbers of contacts were also correlated with the number of active and latent TB cases. While the mean number of TB contacts was positively correlated with the number of active cases as one would expect, the maximum number of TB contacts was negatively correlated with both the number of active and latent cases (Table 4-16).

Negative correlations between contact network parameters, TB mortality and the number of active and latent TB cases could exist simply because PRCC measures residual correlations net of correlations with other parameters, and that the expected positive relationship between network parameters and disease outcomes could be mediated by another model parameter. Once the correlation between that parameter and disease outcomes is taken into account, the expected positive relationship between disease outcomes and network parameters does not exist anymore.

4.2.6. Summary

The base case scenario presented here simulated agricultural output and disease outcomes in the Qu'Appelle Agency. The TB mortality rate produced by the base case scenario was low, even when compared with alternative model runs as well as those runs that selected only the highest TB mortality. This suggests that models that rely on contemporary parameter values to simulate environments that differed from those that produced those values can be successful. This is particularly so if instead of focusing on model outputs averaged out over many runs, they focus on individual extreme outcomes that might in reality come closer to the actual ones.

ABM is a stochastic model, and as such produces a significant amount of variability in outcomes. Sensitivity analyses identified the mean number of TB contacts and susceptibility to TB as the two variables that had the most impact on model output. When disease outcomes were examined, the probability of disease progression and the minimum and the mean numbers of TB contacts were correlated with the output. The sign of these relationships was as expected, aside from the impact of the network parameters on disease outcomes, which tends to be unintuitive. These results indicate the importance of socioeconomic environment summarized as hunger, as well as biological and network parameters to the spread of TB in the Agency. These results present opportunities to explore policy interventions that target hunger reduction.

5. Conclusions

Research presented in this thesis attempted to answer three questions raised in the beginning:

1. Can a TB outbreak that took place over a century ago be reconstructed?
2. What can we learn about the relationships between the disease, the population it afflicted, and the environment in which the outbreak took place?
3. How can reconstruction of a TB outbreak be used to evaluate policy interventions?

The following sections discuss each of these questions.

5.1. Can a TB outbreak that took place over a century ago be reconstructed?

An outbreak of an infectious disease that took place in a distant past can be reconstructed if several challenges are overcome. First of all, the length of time between the outbreak and the time of its reconstruction puts a significant barrier to our ability to understand the context within which the outbreak took place, significantly complicating the reconstruction. In case of the Qu'Appelle Agency, in addition to some 130 years of social, technological, and economic change that took place since the recorded outbreak began, the Indigenous context requires additional effort to understand. This must be done in order to construct a faithful representation of the socioeconomic context within which the outbreak took place. In case of the ABM presented here, the model did not include certain cultural practices such as pipe sharing that are both important to the Indigenous communities from the cultural and spiritual point of view, but could also have facilitated the spread of TB. This omission could be one of the reasons why the base case TB mortality was low compared to its historic level.

The second challenge is that of the data. Understanding the socioeconomic context is a necessary, but not a sufficient condition for reconstructing an outbreak. This understanding has to be given a concrete form by expressing it using data.

Socioeconomic data collected by the DIA can be used to model social determinants of health on reserve. This is a rich source of data, although it has been criticized for having been assembled by and for the benefit of European Canadians, and as such not necessarily capturing that which was important to the Indigenous population of the Agency. From the European point of view, economy and material well-being might be very important. From the Indigenous point of view, European economy based on production has to fit into the Indigenous economy based on sharing (Lutz, 2008).

The third challenge is to make the data alive and talking across the decades by applying appropriate methodology. Agent-based modelling used here is only the tip of the iceberg. Other research methodologies were used to prepare and analyze the data along the way. Each research methodology comes with its own set of assumptions and shapes the data its own unique way. This means that the research process is path-dependent, and the choice of methodology in the early stages of research process might mean that the destination may change.

Each model relies on parameters, and the ABM presented here relied heavily on parameters obtained from studies that were carried out in times and contexts far removed from those of the Qu'Appelle Agency. Often it was the only choice, as studies located closer to the time and space of the Qu'Appelle Agency either did not exist or lacked in quality. A researcher who attempts to reconstruct an outbreak of an infectious disease that took place in a distant past and in a different socioeconomic and cultural environment is forced to make a choice between the quality of studies that provided parameter values, and applicability of their findings to the context being modelled. For example, the impact of malnutrition on TB was obtained from a study that took place in the US in the 1970s. Needless to say, even though from the methodological point of view the study was strong, the context was very different from the context of the Qu'Appelle Agency in the 19th century. There were other studies that might have been more applicable, but the methodology was not appropriate (Cegielski & McMurray, 2004).

Despite this challenge, one can still use parameters derived from contemporary studies. But instead of relying on model output averaged over a number of runs, one may have to use model outputs located towards the tails of its distribution.

Agent-based modelling with its explicitly defined environment, host, and agent, is uniquely suited to be an interdisciplinary exercise. The results of the model developed in this research for the Qu'Appelle Agency TB outbreak 1885-1926 provide an individual-based mechanism that gives a relatively accurate picture of how the outbreak and its related social determinant variables evolved over time as observed. Further, this work demonstrates how the model can be tuned to develop further precision to actual data, thereby offering hypotheses for understanding how the outbreak may have actually evolved.

5.2. What can we learn about the relationship between the disease, the population it afflicted, and the environment in which the outbreak took place?

The research process included analysis of aggregate data, model development, analysis of model output, and policy analysis. Aggregate data analysis demonstrated that correlations differed between years 1885-1889 and 1900-1926, suggesting that socioeconomic reality was different in the two periods. Lags of the number of houses was negatively correlated with TB mortality in both periods, suggesting that housing improvement was instrumental in reducing TB mortality.

TB mortality was significantly and positively correlated with a number of indicators of agricultural output and government rations. This unexpected result can be explained by the cyclical nature of agriculture and the fact that rations would be reduced after a good crop. Agent-based modelling demonstrated that hunger was a key determinant of TB mortality that is able to explain 47% of its variation.

Multivariate regression analysis identified another important predictor of TB mortality – infectious disease outbreaks.

These results indicate the interconnectedness of external factors, such as the climate, with social determinants of health in the form of agricultural output, policy decisions, and health. The endogenous and dynamic nature of these relationships suggests that they should be analysed in a way that preserves these dynamic interactions. This can be done using agent-based modelling.

Sensitivity analyses of the ABM output identified the mean number of TB contacts and susceptibility to TB as the two variables that had the most impact on model output. When disease outcomes were examined, probability of disease progression, the minimum and the mean number of TB contacts were correlated with the model.

Taken together, these results indicate that TB outbreaks in the Qu'Appelle Agency were driven by an interaction of biologic, socioeconomic, and policy factors. Genetic susceptibility to TB likely played a role, as did hunger and the DIA policy to deal with it. Housing construction – another DIA policy – worked to reduce TB mortality in the Agency.

5.3. How can reconstruction of a TB outbreak be used to evaluate policy interventions?

The existing model can be used to evaluate a number of different policy options. In light of its impact on TB mortality, hunger presented a natural candidate for policy intervention. This was particularly so since the DIA had a rations policy.

DIA had a number of other policies as well. Housing construction was an important factor in reducing TB mortality. It was encouraged by the DIA and its impact on TB mortality can be further evaluated. What would an optimal housing policy look like?

In addition to housing, outbreaks of other infectious diseases, such as smallpox, had a significant impact on TB mortality. DIA provided smallpox vaccinations to the people of the Agency. While there is little information on this policy in the Annual Reports, the ABM presented here can evaluate the impact of this policy by simulating TB mortality without the protective effect of smallpox vaccination. Evaluating the impact of housing policy and smallpox vaccinations looks like a promising next step in this research.

5.4. Model limitations and further research

The agent-based model presented here has important limitations. First of all, the complexity of the model requires significant processing power to run the model. Depending on the model scenario, one model run takes from an hour to slightly over two hours. It is possible that some inconsistencies in model output could be caused by stochastic effects. At the same time, given that some model parameters came from a different context and that the "true" parameter values may be closer to the extremes of the parameter ranges explored here, it may not be appropriate to focus on an average output of the model, and special attention should be paid to output of model runs lying in the extremes of their respective distributions.

Data limitations present an important set of challenges. As was discussed above, DIA data were criticised for a number of reasons, and at least in one case DIA fabricated data to make the government look good (J. H. Daschuk et al., 2006) (p. 136). At the same time, DIA data are a unique source of community-level data on Indigenous communities (Burrows, 1986). This research project would benefit from additional sources of data, such as death certificates.

Any model is only as good as the model parameters that support it. The ABM presented in this thesis relies on many parameters that were derived from studies carried out in a different context. This may have resulted in a low TB mortality estimate.

Related to the issue of understanding the cultural context is the way economic activity was modelled. Unmarried adults were assumed to carry out economic activity, farming or ranching on their own. That assumption may not be correct. It is possible that the approach of Gumerman et., al. who assumed that all men and women are married at the age of 15 (Gumerman et al., 2003) and create new households is more appropriate.

The limitations of data, model structure, and model parameters mentioned above stress the importance of having a good understanding of the socioeconomic, historical, and political context of Indigenous societies in which the outbreak took place. This is a clear limitation of this research.

Sensitivity analyses are key to evaluating performance of the model. The size of the model presented in this thesis given limitations of available computing power requires that a subset of all possible values be explored. In subsequent work the scope of sensitivity analyses should be expanded to ensure that model performs well. One of the parameters that should be entered into sensitivity analyses is the proportion of people with active TB that stay at home. Currently, it is assumed that all people with active TB stay at home. Sensitivity analyses should evaluate the impact of this assumption on results by including other proportions, such as 50%.

Another important limitation has to do with TB transmission probabilities indoors and modelling of housing in general. It was a common practice to spend the summer months in tepees rather than houses. The poor quality of houses in the early years of the Agency meant that with their better ventilation, tepees provided protection against TB. Currently this protective effect of tepees is not incorporated into the agent-based model because of data limitations. While reports of Indian Agents mention the practice of moving out into tepees, it is not clear when it started, when it ended if it did, and how widespread it was. Potential importance of this protective effect of tepees to the dynamics of TB suggests that it should be examined further.

Modelling housing and its impact on TB transmission in Qu'Appelle Agency is a challenging task. Lack of standardization of housing stock complicates its modelling. Yet, given the importance of housing in TB transmission, this challenge needs to be overcome.

Hunger was another important driver of TB mortality. Hunger was modelled using an estimate of the cost of food. This approach simplifies the process of modelling, yet may not be totally adequate for a population that produced its own food and supplemented it with hunting and fishing. While this limitation was mitigated by the fact that agricultural output was evaluated in monetary terms, given that hunting and fishing were not modelled, hunger prevalence was likely overestimated. This important factor should be further investigated in the future.

The model fits better in the second half of the period being studied, roughly from 1900 onwards. Years 1885-1900 require further study in order to identify factors that could improve model fit.

Next steps in this research fall into three categories: model calibration, output exploration, and model validation.

Model output suggests that model fit is not perfect. Improvements should include a more realistic modelling of housing and hunger – two key drivers of TB mortality. In addition, while this model focused primarily on adult males, behaviour of other types of agents should be modelled explicitly. For example, boarding schools are not modelled explicitly in the current ABM. Yet, they were key environments in which TB transmission took place (Ferguson R. G., 1934).

The ABM generated a lot of data that describes demographic characteristics, contact network, SES, location and activity, among other things, of each agent at four different points in a day, but only a fraction of it was examined. Further research should include examination of the relationship between other types of output generated by the model and TB mortality. For example, while the focus of this thesis was largely on the economic determinants of health, TB mortality may be strongly related to one's contact network and type of environment in which interactions take place. Further research should include examination of the relationship between TB and the SDH at the micro level.

Finally, model validation using a different data set, ideally from another reserve, but from the same time period, would provide important insights into the ability of the ADM to explain historic TB outbreaks in First Nations communities. Further research may involve a collaboration with Aboriginal Affairs and Northern Development Canada and Library and Archives Canada in order to find TB mortality data from another Agency that could be used to validate results of this research.

6. Bibliography

Al-Azem, A. (2007). *Social network analysis in tuberculosis control among the Aboriginal population of Manitoba.* Unpublished PhD, University of Manitoba, University of Manitoba.

Axtell, R. E., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., et al. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in long house valley. *PNAS, 99(Suppl, 99*(supp. 3), 7275-7279.

Bandini, S., Manzoni, S., & Vizzari, G. (2009). Agent based modeling and simulation: An informatics perspective. *Journal of Artificial Societies and Social Simulation, 12*(4)

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K. P. W. J., Whittle, H. C., & Hill, A. V. S. (1998). Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *New England Journal of Medicine, 338*, 640.

Bishai, J. D., Bishai, W. R., & Bishai, D. M. (2010). Hightened vulnerability to MDR-TB epidemics after controlling drug-susceptible TB. *PLOS One, 5*(9), 1-11.

Blaine, E. M. (1991). Charles Bishop and Northern Ojibwa. In Abel,K., Friesen, J., editors (Ed.), *Aboriginal resource use in Canada: Historical and legal aspects* (). Winnipeg: University of Manitoba Press.

Blower, S. M., & Dowlatabadi, H. (1994). Sensitivity and uncertainty analysis of complex models of disease transmission: An HIV model, as an example. *International Statistical Review, 62*(2), 229.

Burrows, J. K. (1986). "A much-needed class of labour": The economy and income of the Southern Interior Plateau Indians, 1897-1910. *BC Studies*, 71:27-46.

Cegielski, J. P., Arab, L., & Cornoni-Hutley, J. (2012). Nutritional risk factors for tuberculosis among adults in the United States, 1971-1992. *American Journal of Epidemiology,* 176(5);409-422

Cegielski, J. P., & McMurray, D. N. (2004). The relationship between malnutrition and tuberculosis: Evidence from studies in humans and experimental animals. *International Journal of Tuberculosis and Lung Disease, 8*(3), 286-298.

Cilfone, N. A., Perry, C. R., Kirschner, D. E., & Linderman, J. J. (2013). Multi-scale modeling predicts a balance of tumor necrosis factor-alpha and interleukin-10 controls the granuloma environment during mycobacterium tuberculosis infection. *PLOS One, 8*(7), 1-16.

Clark, M., Riben, P., & Nowgesic, E. (2002). The association of housing density, isolation and tuberculosis in Canadian First Nations communities. *Int J Epidemiol, 31*, 940.

Clark, G. A., Kelley, M. A., Grange, J. M., Cassandra Hill, M., Katzenberg, M. A., Klepinger, L. L., et al. (1987). The evolution of mycobacterial disease in human populations: A reevaluation. *Current Anthropology*, 28(1):45-62.

Connolly, M. A. (Ed.). (2005). *Communicable disease control in emergencies. A field manual.* Geneva: WHO.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature, 433*, 513.

Dancik, G. M., Jones, D. E., & Dorman, K. S. (2010). Parameter estimation and sensitivity analysis in an agent-based model of leishmania major infection. *Journal of Theoretical Biology, 262*, 398-412.

Daschuk, J. (2013). *Clearing the Plains: Disease, politics of starvation, and the loss of Aboriginal life.* Regina: University of Regina Press.

Daschuk, J. H., Hackett, P., & McNeil, S. (2006). Treaties and tuberculosis: First Nations people in late 19th-century western Canada a political and economic transformation. *Canadian Bulletin of Medical History, 32*(2), 307-330.

Dean, J. G., Gumerman, G. J., Epstein, J. M., Axtell, R., & Swedlund, A. C. (2006). *Understanding the Anasazi culture change through agent-based modeling*. Santa Fe: Santa Fe Institute.

DIA. (1877). *Annual report of the Department of the interior for the year ended 30th june 1876;*. Ottawa: MacLean.

DIA. (1886). *Annual report of the Department of Indian Affairs for the year ended march 31, 1886.* Ottawa, ON: Dawson, S.E.

DIA. (1888). *Annual report of the Department of Indian Affairs for the year ended december 31, 1887.* Ottawa, ON: MacLean, Roger & Co.

DIA. (1889). *Annual report of the Department of Indian Affairs for the year ended december 31st, 1888.* Ottawa, ON: A Senecal.

DIA. (1890). *Annual report of the Department of Indian Affairs for the year ended 31st december, 1889.* Ottawa, ON: Brown Chamberlin.

DIA. (1891). *Annual report of the Department of Indian Affairs for the year ended 31st december, 1890.* Ottawa, ON: Brown Chamberlin.

DIA. (1892). *Annual report of the Department of Indian Affairs for the year ended 31st december 1891.* Ottawa, ON: S.E. Dawson.

DIA. (1894). *Annual report of the Department of Indian Affairs 30th june 1893.* Ottawa, ON: S.E. Dawson.

DIA. (1896). *Annual report of the Department of Indian Affairs for the year ended 30th june 1895.* Ottawa, ON: S.E.Dawson.

DIA. (1897.). *Annual report of the Department of Indian Affairs for the year ended 30th june 1896.* Ottawa, ON: S.E.Dawson.

DIA. (1898). *Annual report of the Department of Indian Affairs for the year ended 30th june 1897.* Ottawa, ON: S.E.Dawson.

DIA. (1899.). *Annual report of the Department of Indian Affairs for the year ended 30th june 1898.* Ottawa, ON: S.E.Dawson.

DIA. (1900). *Annual report of the Department of Indian Affairs for the year ended 30th june 1899.* Ottawa, ON: S.E.Dawson.

DIA. (1901). *Annual report of the Department of Indian Affairs for the year ended june 30 1900.* Ottawa, ON: S.E.Dawson.

DIA. (1902). *Annual report of the Department of Indian Affairs for the year ended june 301902.* Ottawa, ON: S.E.Dawson.

DIA. (1906). *Annual report of the Department of Indian Affairs for the year ended june 30 1906.* Ottawa, ON: S.E.Dawson.

DIA. (1907). *Annual report of the Department of Indian Affairs for the year ended march 31 1907.* Ottawa, ON: S.E.Dawson.

DIA. (1909). *Annual report of the Department of Indian Affairs for the year ended march 31 1909.*C.H. Parmelee.

DIA. (1913). *Annual report of the Department of Indian Affairs for the year ended march 31 1912.* Ottawa, ON: C.H.Parmelee.

DIA. (1915). *Annual report of the Department of Indian Affairs for the year ended march 31 1915.* Ottawa, ON: J. De L. Tache.

DIA. (1917). *Annual report of the Department of Indian Affairs for the year ended march 31, 1916.* Ottawa, ON.: J. De L. Tache.

DIA. (1930). *Annual report of the Department of Indian Affairs for the year ended march 31;*. Ottawa, ON: Dawson, S.E.

DIA. (1966). *Treaty texts - treatyno.4 treaty no.4 between Her Majesty the Queen and the Cree and Saulteaux tribes of Indians at the Qu'Appelle and Fort Ellice* No. Ci 72-0466). Ottawa, ON.: ROGER DUHAMEL, F.R.S.C.

Dickason, O. (2002). *Canada's First Nations: A history of founding peoples from earliest times* (3rd edition ed.). Oxford: Oxford University Press.

Dominion Bureau of Statistics. (1939). *Annual report of tuberculosis institutions*. Ottawa, ON: J.O. Patenaude, I.S.O. Printer to the King's Most Excellent Majesty.

El-Kebir, M., van der Kuip, M., van Furth, A. M., & Kirschner, D. E. (2013). Computational modeling of tuberculous meningitis reveals an important role for tumor necrosis factor-alpha. *Journal of Theoretical Biology, 328*, 43-53.

Environment Canada. (2013). *Climate.* Retrieved 01/01, 2013, from http://climate.weather.gc.ca/

Epstein, J. (2006). *Generative social science.* Princeton.: Princeton University Press.

Epstein, J. A., & Axtell, R. L. (1996). *Growing artificial societies: Social science from the bottom up* Washington, D. C; The Brookings Institution.

Ferguson, R. G. (1928). *Tuberculosis among the Indians of the Great Canadian Plains* Preliminary report of an investigation being carried out by the National Research Council of Canada.

Giesecke, J. (2002). *Modern infectious disease epidemiology. second edition*. New York: Oxford University Press.

Gomez i Prat, J., & Mendoca de Souza, S. M. (2003). Prehistoric tuberculosis in america: Adding comments to a literature review. *Memórias do Instituto Oswaldo Cruz, 98*(Supp 1), 151-159.

Gracey, M., & King, M. (2009). Indigenous health part 1: Determinants and disease patterns. Lancet, 374, 65.

Greenwood, C. M. T., Fujiwara, M., Boothroyd, L. J., Miller, M. A., Frappier, D., Fanning, A., et al. (2000). Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large Aboriginal Canadian family. *American Journal of Human Genetics, 67*, 405.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11*(1), 1.

Grygier, P. S. (1994). *A long way from home: The tuberculosis epidemic among the inuit*. Quebec: McGill-Queen's University Press.

Grzybowski, S. A., & Allen, E. A. (1999). Tuberculosis: 2 History of the disease in canada. *Canadian Medical Association Journal, 160*(7), 1025.

Gumerman, G. S., Swedlund, A. C., Dean, J. S., & Epstein, J. M. (2003). The evolution of social behavior in the prehistoric american southwest. *Artificial Life, 9*, 435-444.

Hamilton, J. D. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Krepakevich, J. D. (Producer), & Harper, R. (Director). (1992). *Coppermine: Consequences of contact with the outside.* [Motion Picture] Ottawa,ON.: National Film Board of Canada.

Hethcote, H. W. (1989). Three basic epidemiological models. In Levin, Hallam & Gross. (Eds.), *Biomathematics* (vol. 8. ed., pp. 119). Berlin: Springer-Verlag.

Heymann, D. L. (Ed.). (2015). *Control of communicable diseases manual.* (20th ed.). Washington, DC: American Public Health Association.

Ho, D. (1996). The influence of coinfections on HIV transmission and disease progression. *The AIDS Reader, 64*(4), 114-116.

Karanikolos, M., Mladovsky, P., Cylus, J., Thomson, S., Basu, S., Stuckler, D., et al. (2013). Financial crisis, austerity, and health in europe. *The Lancet., 381*(9874), 1323-1331.

Kasaie, P., Andrews, J. R., Kelton, D. W., & Dowdy, D. W. (2014). Timing of tuberculosis transmission and the impact of household contact tracing. *Am J Resip Crit Care Med, 189*(7), 845-852.

Kennedy, P. (2003). *A guide to econometrics.* (5th ed.). Cambridge, MA: MIT Press.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London*, 700-721.

Larcombe, L. A., Mookherjee, N., Lodge, A. M., Brown, J. S., Denechezhe, L., Nickerson, P. W., et al. (2015). Frequency of NRAMP1 gene polymorphisms among canadian first nations peoples experiencing endemic tuberculosis. *Mycobacterial Diseases., 5*(5), 193.

Li, X. W., Yang, Y., Zhou, F., Zhang, Y., Lu, H., Jin, Q., et al. (2011). SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: Updated systematic review and meta-analysis. *PLOS One, 6*(1), e15831.

Long, R. E., & Ellis, E. (Eds.). (2007). *Canadian tuberculosis standards* (6th edition; ed.). Ottawa: Public Health Agency of Canada.

Lonnroth, K., Jaramillo, E., Williams, B.G., Dye, C., Raviglione, M. (2009). Drivers of tuberculosis epidemics: The role of risk factors and social determinants. *Soc. Sci. Med.* 68(12); 2240-2246

Loppie Reading, C. W., Wien, F. (2009). Health inequalities and social determinants of Aboriginal people's health. National Collaborating Centre For Aboriginal Health.

Lutz, J. S. (2008). *Makuk: A new history of Aboriginal-White relations*. Vancouver: UBC Press.

Makridakis, S., & Hibon, M. (1995). *ARMA models and the box- jenkins methodology.* No. 95/45/TM). Cedex, France.: INSEAD.

Marino, S., El-Kebir, M., & Kirschner, D. (2011). A hybrid multi-compartment model of granuloma formation and T cell priming in tuberculosis. *Journal of Theoretical Biology, 280*(1), 50-62.

Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analisys in systems biology. *Journal of Theoretical Biology, 254*, 178-196.

McKeown, T., & Record, R. G. (1962). Reasons for the decline of mortality in England and Wales during the ninetheenth century. *Popul Stud, 16*(2), 94.

Menzies, D., Ellis, E., Long, R., Madhukar, P., & Wong, T. (Eds.). (2014). *Canadian tuberculosis standards.* (7th edition. ed.). Ottawa,ON: PHAC.

Miller, P., Lanier, W., & Brandt, S. (2001). *Using growing degree days to predict plant stages.* No. MT200103 AG). Bozeman,MT: Montana State University.

Miller, R. (2002). *Epidemiology for health promotion and disease prevention professionals*. New York: Haworth Press.

Morris, M. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics, 33*(2), 167.

Moss, S. (2008). Alternative approaches to the empirical validation of agent-based models. *Journal of Artificial Societies and Social Simulation, 11*(1), 5-20.

Ndzinge, L. O., Marsh, J. M., & Greer, C. (1984). Herd inventory and slaughter supply response of Botswana beef cattle producers. *Journal of Agricultural Economics, 35*(1), 97-107.

Omran, A. (2005). The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly., 85*(4), 731-737.

Parunak, V. D. H., Savit, R., & Riolo, R. L. (1998). Agent-based modeling vs. equation-based modeling: A case study and users' guide. In J. S. Sichman, R. Conte & N. Gilbert (Eds.), *Multi-agent systems and agent-based simulation.* (pp. 10-25) Springer Berlin Heidelberg. doi:10.1007/10692956_2

Pepperell, C. S., Granka, J. M., Alexander, D. C., Behr, M. A., Chui, L., Gordon, J., et al. (2012b). Dispersal of mycobacterium tuberculosis via the Canadian fur trade. *PNAS*, 108(16):6526-6531.

PHAC. (2015). *Tuberculosis in Canada 2013 - pre-release.* Ottawa (Canada): Minister of Public Works and Government Services Canada;.

Prohorov.B.B. (2000). Общественное здоровье за 100 лет (1897-1997). In N. N. Marfenin, N. N. Moiseev & S. A. Stepanov (Eds.), *Россия в окружающем мире: 2000 (Аналитический ежегодник).* [Russia in the world: 2000.] (). Moscow.: MNEPU.

Ray, A. J. (1990). *The Canadian fur trade in the industrial age.* Toronto: University of Toronto Press.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis, 22*(3), 579.

Segovia-Juarez, J. L., Ganguli, S., & Kirschner, D. (2004). Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model. *Journal of Theoretical Biology, 231*(3), 357-376.

Singer, M., & Clair, S. (2003). Syndemics and public health: Reconceptualizing disease in bio-social context. *Medical Anthropology Quarterly, 17*(4), 423-441.

Sproule-Jones, M. (1996). Crusading for the forgotten: Dr. Peter Bryce, public health, and the prairie native residential schools. *Canadian Bulletin of Medical History, 13*, 199-224.

Statistics Canada. (2014). *Historical statistics of Canada - section K: Price indexes.* Retrieved 08/25, 2016, from http://www.statcan.gc.ca/pub/11-516-x/sectionk/4057753-eng.htm#1

Statistics Canada. (2015). *Microsimulation*. Retrieved 30/10 2017, from https://www.statcan.gc.ca/eng/microsimulation/index.

Stein, L. (1952). Tuberculosis and the "social complex" in Glasgow. *British Journal of Social Medicine, 6*, 1.

Taylor, J. L. (1985). *Treaty research report - treaty four*. Ottawa: Treaties and Historical Research Centre, Indian and Northern Affairs Canada.

Thorngate, W., & Edmonds, B. (2013). Measuring simulation-observation fit: An introduction to ordinal pattern analysis. *Journal of Artificial Societies and Social Simulation, 16*(2), 4-18.

Tian, Y., Osgood, N. D., Al-Azem, A., & Hoeppner, V. H. (2013). Evaluating the effectiveness of contact tracing on tuberculosis outcomes in Saskatchewan using individual-based modeling. *Health Educ Behav, 40*(1 Suppl), 98S-110S.

Tough, F. (1992). Regional analysis of Indian aggregate income, Northern Manitoba: 1896-1935. *The Canadian Journal of Native Studies*, 12(1):95-146.

Troisi, A., Wong, V., & Ratner, M. A. (2005c). An agent-based approach for modelling molecular self-organization. *PNAS*, 102(2):255-260.

Urquhart, M. C. (1986). New estimates of gross national product, Canada, 1870-1926: Some implications for Canadian development. In Engerman,S.L., Gallman, R.E., editors (Ed.), *Long-term factors in American economic growth* (pp. 9-94). Chicago: University of Chicago Press.

Usher, P. J. (1965). Economic basis and resource use of the Coppermine - Holman region, N.W.T. (Master's of Art, McGill University).

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling and Software., 63*, 24.

Waldram, J. B., Herring, D. A., & Young, T. K. (2006). *Aboriginal health in Canada: Historical, cultural, and epidemiological perspectives. 2nd edition*. Toronto: University of Toronto Press.

Watson, M. W. (1986). Univariate detrending methods wit stochastic trends. *Journal of Monetary Economics, 18*(1), 49-75.

WHO. (2007). Social determinants and indigenous health: The international experience and its policy implications. Report on specially prepared documents, presentations and discussions at the international symposium on the social determinants of indigenous health Adelaide, 29-30 April 2007 for the commission on social determinants of health (CSDH).

Zaidi,A., Rake, K. (2001). *Dynamic microsimulation models: A review and some lessons for SAGE.* London: SAGE.

Zimmer, T. A. M. (2008). *Abandoned rail lines in saskatchewan as of september 2008.*

7. Appendices

A. Timeline of TB outbreak in the Qu'Appelle Agency

The following graph presents the timeline of important events in the Qu'Appelle Agency. The left side of the graph presents the course of TB outbreak and various stages of economic development. The right side presents important events, such as the creation of the Agency and changes in living arrangements.

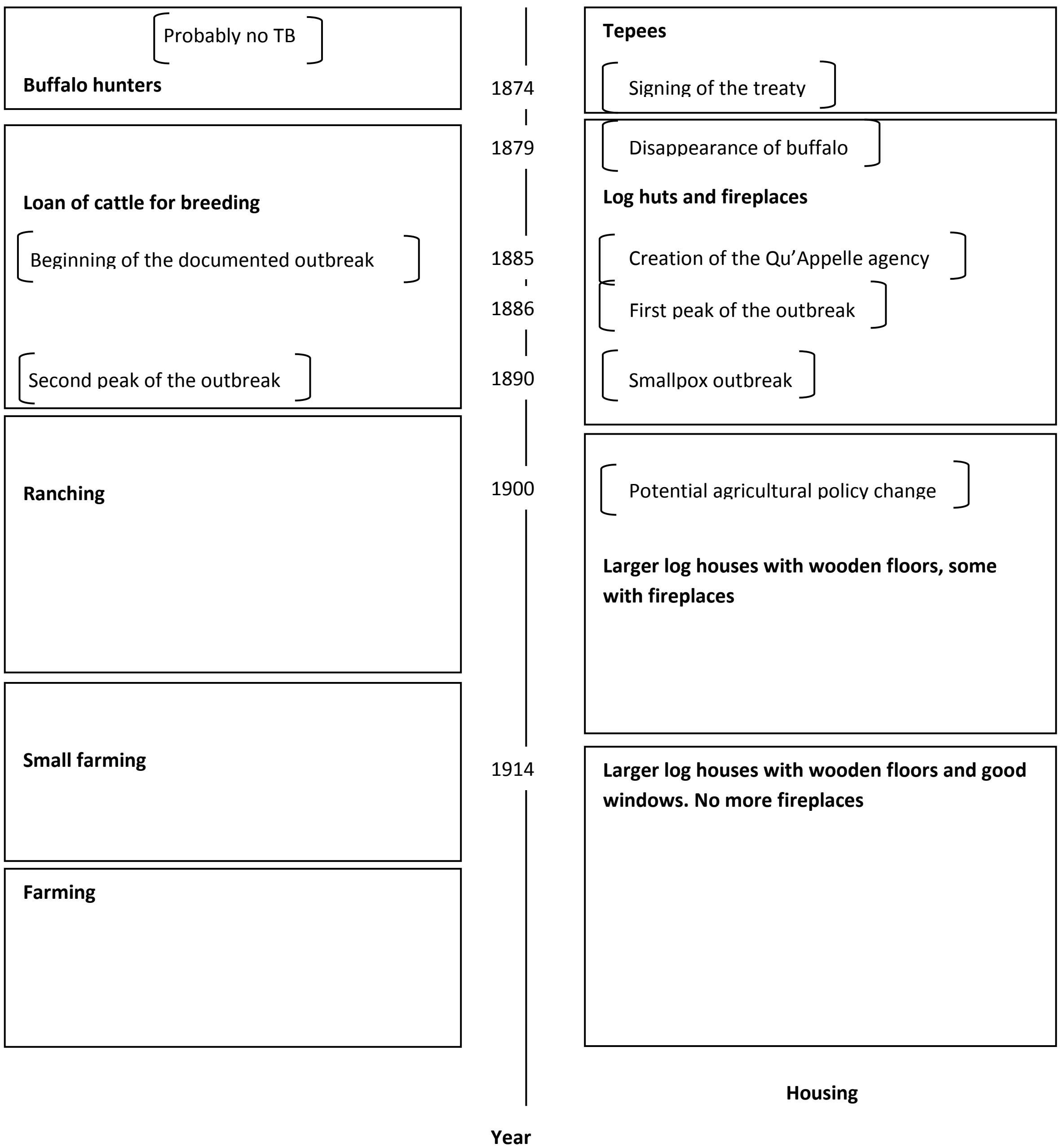

Figure 7-1: Timeline of historical events of the Qu'Appelle Agency. Source Ferguson (1928)

B. Data

The following tables present data on mortality and population (Table 7-1), total income and value of property (Table 7-2), population by age and sex (Table 7-3 and Table 7-4), agricultural statistics (Table 7-5), and capital statistics (Table 7-6).

Table 7-1: Mortality and population. Source: Annual Report of the Department of Indian Affairs.

Year	TB mortality rate per 1000	Total mortality rate per 1000	Population	Births by year, File Hills part of the Agency
1885	6.3	116	1492	
1886	93.4	94	1317	21
1887	68.2	88	1104	13
1888	42.9	66	1088	9
1889	38.9	62	1073	19
1890	89.6	138	857	12
1891	76.5	92	1003	11
1892	52.2	100	998	
1893	28.0	64	966	10
1894	43.5	67	988	10
1895	20.4	58	958	
1896	33.0	64	949	12
1897	27.8	83	995	13
1898	40.0	79	975	15
1899	13.6	68	828	5
1900	16.1	62	830	7
1901	18.6	57	778	7
1902	22.8	38	831	12
1903	19.9	50	831	
1904	17.0	42	818	
1905	19.5	32	822	
1906	19.6	29	860	
1907	11.8	40	867	
1908	11.9	42	860	
1909	9.3	32	815	
1910	9.6	46	834	
1911	21.2	33	845	
1912	15.3	28	844	
1913	9.3	30	871	
1914	16.5	52	888	
1915	9.8	36	757	
1916	9.4	38	760	
1917	14.3	29	769	

1918	11.9	31		
1919	10.0	52		
1920	14.0	50		
1921	15.2	24		
1922	12.1	18		
1923	11.7	22		
1924	9.6	33	764	
1925	16.0	21		
1926	8.4	22		

Table 7-2: Total income and value of property, Qu'Appelle Agency. Source: Department of Indian Affairs (1885-1926)

Year	Total income (dollars)	Value of property, dollars
1885		
1886		
1887		
1888		
1889		
1890		
1891		
1892		
1893		
1894		
1895		
1896		
1897	33173	191322
1898	36306	602927
1899	33529	673085
1900	39431	704608
1901	39519	720204
1902	50996	726238
1903	81861	1355161
1904	89363	1372557
1905	85660	1433643
1906	94559	1699270
1907	102462	1932295
1908	71940	1464750

1909	89952	1365095
1910	99527	1358836
1911	128114	1485660
1912	127545	1768419
1913	128730	1762005
1914	205744	2286689
1915	152669	1921400
1916	114999	1954356
1917	167949	1984690
1918	226053	1994302
1919	233301	2087228
1920	218818	2355063
1921	249498	1808350
1922	144449	2280320
1923	150306	2259950
1924	138229	2224396
1925	157897	2189775
1926	183778	2237525

Table 7-3: Population by age, males, Qu'Appelle Agency. Source: Annual Report of the Department of Indian Affairs.

	Under 6	6-15	16-20	21-65	65+
1901	57	73	35	168	13
1902	70	63	37	170	34
1903	68	61	33	181	33
1904	78	58	30	179	32
1905	90	57	35	166	33
1906	93	64	39	172	31
1907	75	73	40	182	38
1908	77	72	41	176	36
1909	75	69	47	151	36
1910	64	63	44	190	32
1911	83	85	38	210	28
1912	80	91	39	209	29
1913	88	92	43	209	30
1914	88	101	21	240	35
1915	69	83	30	218	30
1916	81	92	31	211	39
1924	96	123	32	192	47
1929	101	101	57	214	45

Table 7-4: Population by age, females, Qu'Appelle Agency. Source: Annual Report of the Department of Indian Affairs.

	Under 6	6-15	16-20	21-65	65+
1901	59	83	50	222	18
1902	76	81	35	227	38
1903	85	80	29	224	37
1904	86	76	27	219	33
1905	98	73	29	210	31
1906	113	76	33	212	27
1907	98	81	41	197	42
1908	100	79	39	198	42
1909	98	75	48	178	38
1910	73	72	50	210	36
1911	86	93	43	231	39
1912	84	92	51	231	39
1913	87	91	52	231	40
1914	95	92	24	256	52

1915	95	78	23	229	48
1916	89	76	32	218	50
1924	80	122	41	198	53
1929	98	107	66	203	55

Table 7-5: Agricultural statistics, Qu'Appelle Agency. Source: Annual Report of the Department of Indian Affairs.

Year	Wheat harvest, bushels	Oats harvest, bushels	Potato harvest, bushels
1885	608		1705
1886	258		2475
1887	824	18	5167
1888	4716	200	6257
1889	703	97	945
1890	6331	1022	5159
1891	10932	1801	5652
1892	5718	624	3322
1893	2403	610	1176
1894	56	14	213
1895	4323	1218	2181
1896		1278	
1897		1338	
1898	3310	1398	1556
1899	703	47	590
1900	6748	6348	4004
1901	2610	6541	3686
1902	12076	18173	6048
1903	22019	19911	3834
1904	33209	22333	4014
1905	36598	23739	4075
1906	43300	27278	2673
1907	60491	32186	2930
1908	23286	17759	2120
1909	22887	24845	2075
1910	49642	98730	2360
1911	40084	60396	3575
1912	43475	58723	3990
1913	31127	69195	1800
1914	32136	175475	6430
1915	21346	41939	3875
1916	24094	29228	3485
1917	33352	93466	3189
1918	22684	85703	4711
1919	31688	55826	3246
1920	34338	68880	1947
1921	36729	79566	1996
1922	45372	87775	3495

1923		75503	3180
1924		63231	2865
1925	46652	50959	2550
1926	62250	46703	2715
1927	82535	40758	2745
1928	82764	35187	6905
1929	73440	41511	2509
1930	26335	5832	475

Table 7-6: Capital statistics, Qu'Appelle Agency. Source: Annual Report of the Department of Indian Affairs.

Year	Horses	Oxen	Ploughs	Mowers	Houses
1885	76	110			161
1886	130	146			166
1887	174	138			180
1888	226	162	118		153
1889	305	161	125		156
1890	251	214	135		167
1891	375	202	142		298
1892	430	199	133		307
1893	349	210	133		309
1894	363	188	136	30	308
1895	605	202	136	30	304
1896		623	136	32	
1897	646	1044	125	44	183
1898	650	1031	125	53	182
1899	714	1319	124	56	174
1900	729	1246	134	60	179
1901	704	1366	139	61	177
1902	701	1200	138	63	174
1903	701	1200	138	57	174
1904	748	1264	118		178
1905	761	1485	110	66	177
1906	168	1319	119	70	184
1907	688	1437	126	76	182
1908	592	1284	131	82	171
1909	334	1256	108	61	165
1910	443	1439	105	69	145
1911	426	1399	122	75	160
1912	611	1208	143	77	182
1913	612	1193	140	81	68
1914	738	1299	172	116	224
1915	663	1128	252	89	204
1916	652	1170	118	90	229
1917	712	1219	228	240	235
1918	754	1142	240	244	230
1919	746	1177	286	218	204
1920	800	1129	298	287	238
1921	859	1045	298	289	249
1922	896	1126	286	293	250

1923	826	1027	296	302	251
1924	916	1137	303	312	249
1925	986	1140	303	312	249
1926	1035	1262	309	310	243
1927	1049	1101	312	314	247
1928	1190	952	312	318	249
1929	1010	788	563	250	260
1930	968	1058	588	258	287

Table 7-7: Expenditures on destitute (Rations). Source : DIA(1885-1926)

Year	Expenditures on destitute Indians, Manitoba	Expenditures on destitute Indians, Qu'Appelle Agency*
1885	478038	*16872.48*
1886	541826.3	*19470.59*
1887	484122.6	*17596.59*
1888	372070	*14289.26*
1889	362657.5	*13936.7*
1890	352734.3	*13565.51*
1891	289153.3	*11493.74*
1892	231302	*9577.551*
1893	202243.5	*8508.613*
1894	206107	*8674.318*
1895	215416.9	*9084.677*
1896	176092.8	*7682.284*
1897	178103.1	7771
1898	182718.3	3415.98
1899	190773.6	3605.66
1900	205267.1	3377.07
1901	199481.2	3080.2
1902	188325	2584.92
1903		3545.07
1904	186342.7	1320.28
1905	173118.4	1424.82
1906	154843.6	1425.71
1907	113299	1148.05
1908	143034.4	1207.33

1909	163151.8	1405.28
1910	119851.9	841.44
1911	115525.3	906.84
1912		912.16
1913	115480.6	1032.07
1914	152383.8	2480.48
1915	152588.9	1546.32
1916	145277.8	2119.74
1917	143833.7	1910.01
1918	150439.5	2120.15
1919	156098.4	2108.38
1920	153100.5	1691.1
1921	127921.5	806.83
1922	154871.9	1776.15
1923	164597	1635.12
1924	137855.8	2476.09
1925	141482.7	2913.73
1926	160480.6	2239.63

*Estimated in years 1885-1896 assuming the same profile as expenditures in Manitoba

C. Aggregate Data Analyses

7.1.1. ARIMA modelling

7.1.1.1. TB mortality

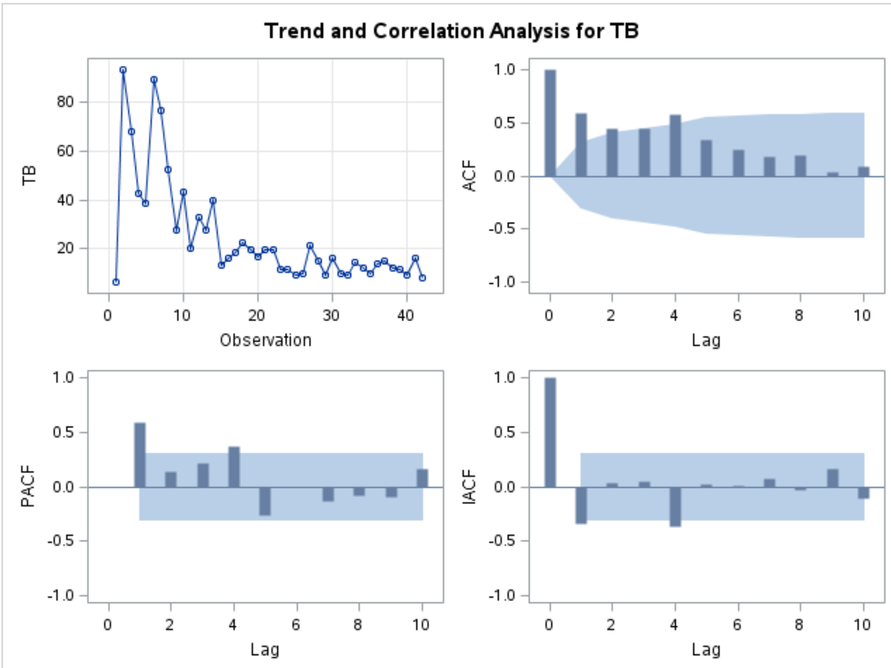

Figure 7-2: ARIMA modelling and forecasting results: TB mortality

Table 7-8: Conditional least squares estimation results, TB mortality ARIMA(0,1,2)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	0.45223	3.15637	0.14	0.8868	0
MA1,1	0.23725	0.13747	1.73	0.0925	1
MA1,2	-0.46857	0.13809	-3.39	0.0016	4

Table 7-9: Autocorrelation check of residuals, TB mortality ARIMA(0,1,2)

Autocorrelation check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.81	4	0.5909	-0.011	-0.191	-0.104	0.058	-0.083	-0.055
12	6.18	10	0.7997	-0.114	0.097	-0.151	0.065	0.083	0.07
18	7.18	16	0.9697	-0.118	-0.015	-0.009	-0.002	0.032	-0.019
24	7.87	22	0.9975	0.02	0.038	-0.061	-0.018	-0.044	-0.004

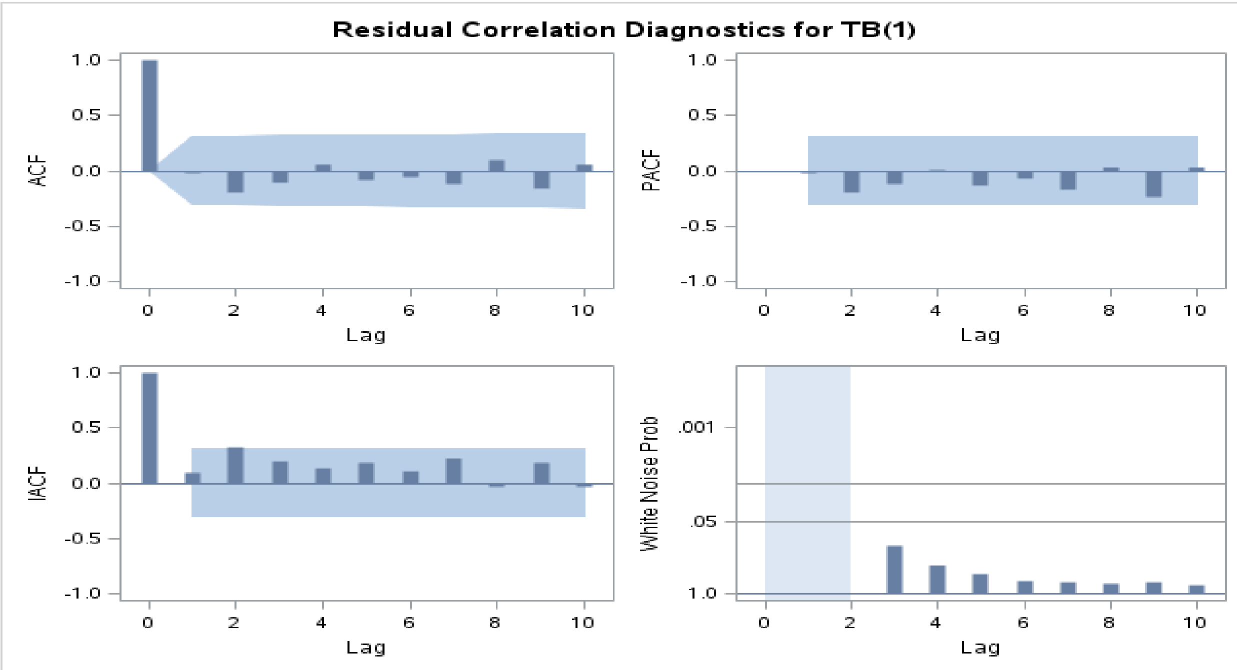

Figure 7-3: Residual correlation diagnostics for TB mortality, ARIMA(0,1,2)

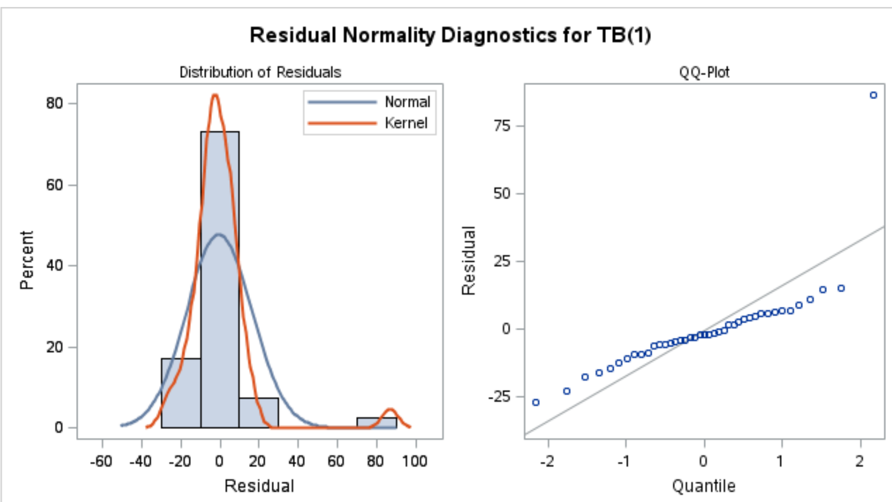

Figure 7-4: Residual normality diagnostics for TB mortality, ARIMA(0,1,2)

7.1.1.2. Potato harvest

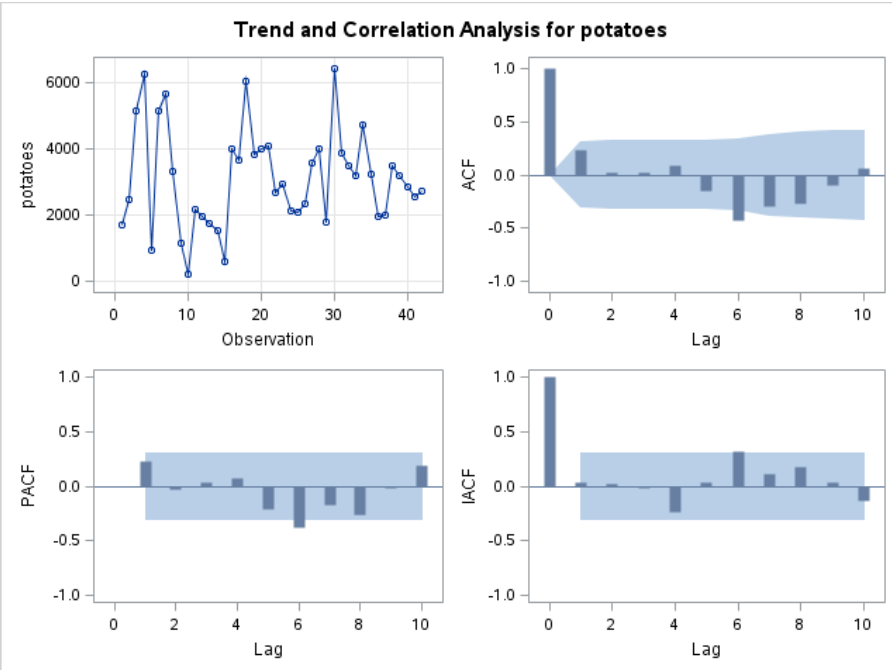

Figure 7-5: Trend and correlation analysis for potato harvest

Table 7-10: Autocorrelation check for white noise, potato harvest

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	13.43	6	0.0366	0.229	0.02	0.024	0.084	-0.156	-0.429

Table 7-11: Conditional least squares estimation, potato harvest, ARIMA(1,0,0)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	3117.8	152.36028	20.46	<.0001	0
AR1,1	-0.43846	0.14408	-3.04	0.0041	6

Table 7-12: Autocorrelation check of residuals, potato harvest, ARIMA(1,0,0)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.54	5	0.7702	0.085	-0.073	-0.002	0.166	-0.11	0.022

12	10.84	11	0.4567	-0.24	-0.221	-0.092	0.134	-0.134	0.014
18	14.54	17	0.6288	-0.031	0.21	-0.024	-0.003	-0.086	-0.043
24	20.12	23	0.6346	-0.213	0.001	-0.018	-0.063	-0.039	0.117

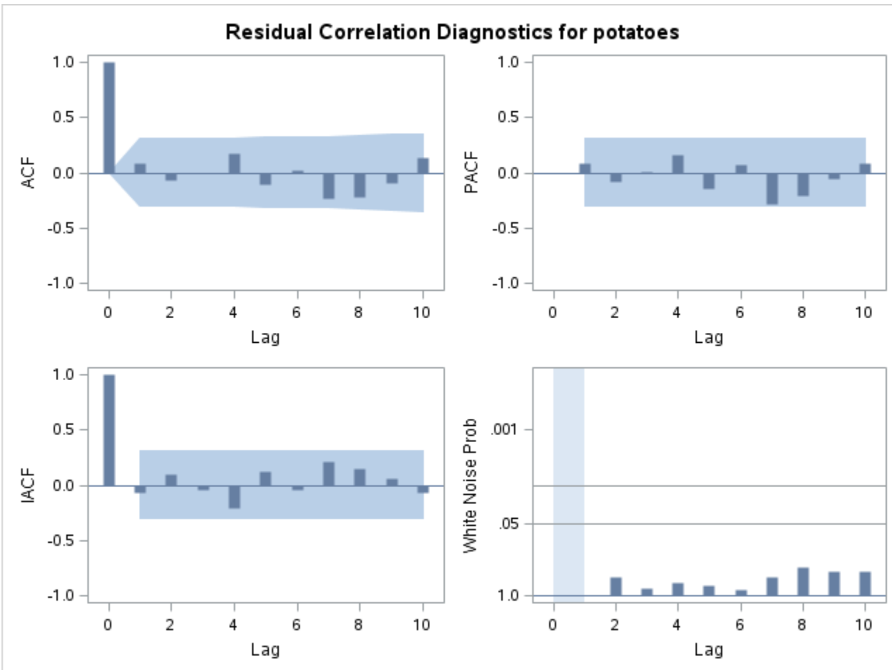

Figure 7-6: Residual correlation diagnostics for potato harvest, ARIMA(1,0,0)

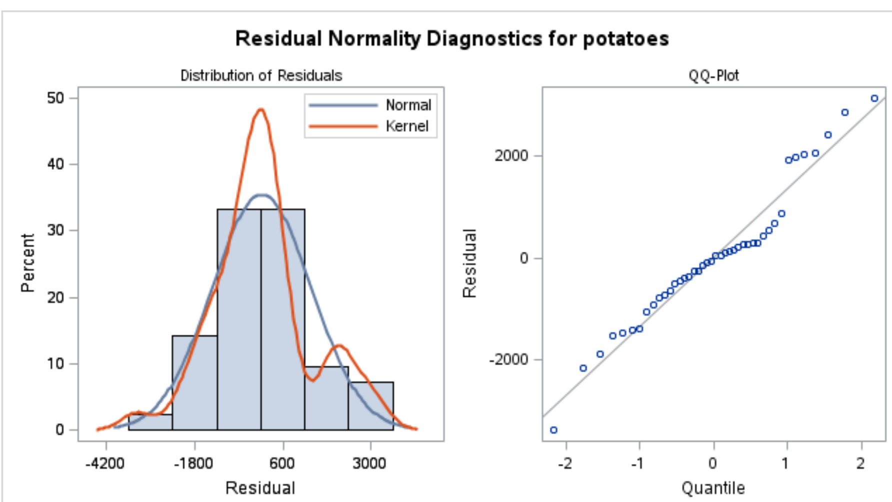

Figure 7-7: Residual normality diagnostics for potato harvest, ARIMA(1,0,0)

7.1.1.3. Hay cut

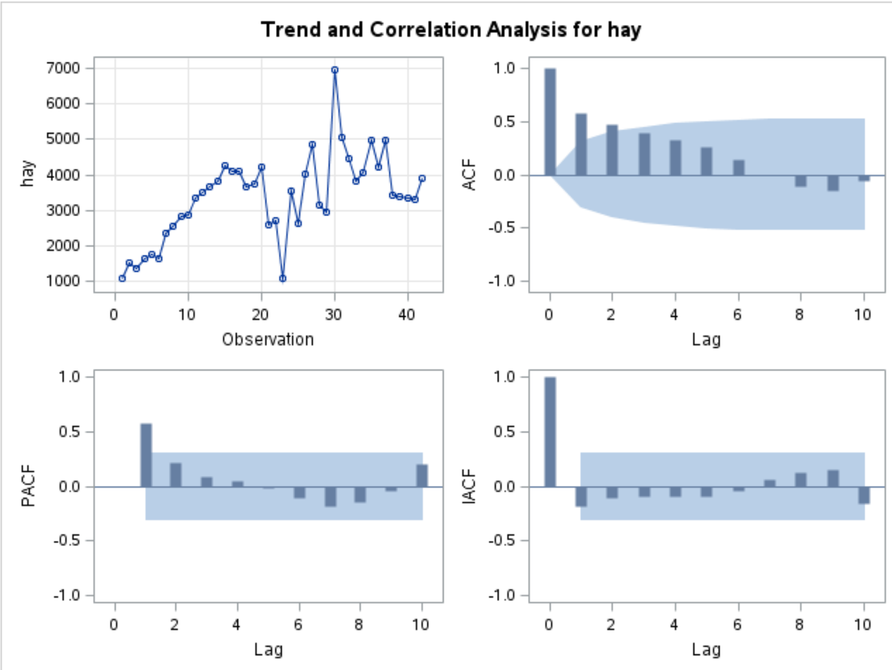

Figure 7-8: Trend and correlation analysis for hay cut

Table 7-13: Conditional least squares estimation, hay cut, ARIMA(0,1,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	61.76909	66.21151	0.93	0.3566	0
MA1,1	0.56036	0.13288	4.22	0.0001	1

Table 7-14: Autocorrelation check of residuals, hay cut, ARIMA(0,1,1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.61	5	0.9877	0.014	-0.035	-0.008	0.049	0.075	0.054
12	8.82	11	0.6382	-0.128	-0.238	-0.254	0.055	0.024	-0.099
18	14.02	17	0.6659	-0.008	-0.168	0.14	0.061	0.121	0.094
24	15.13	23	0.8899	0.051	0.041	-0.032	0.059	0.027	-0.052

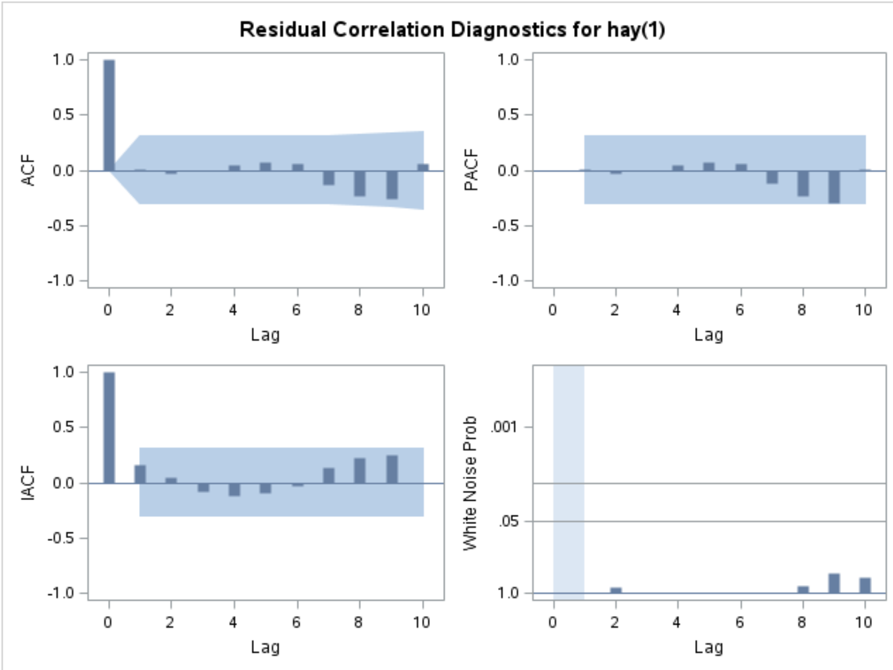

Figure 7-9: Residual correlation diagnostics for hay cut, ARIMA(0,1,1)

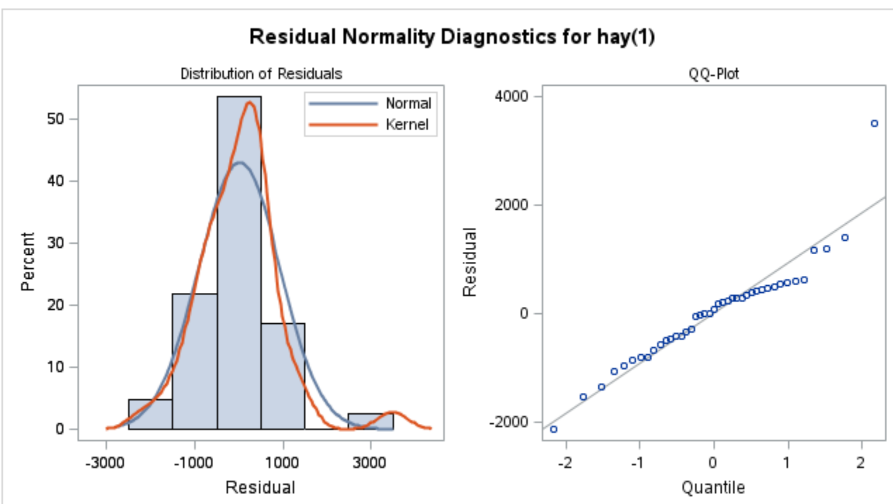

Figure 7-10: Residual normality diagnostics for hay cut, ARIMA(0,1,1)

7.1.1.4. *Oats harvest*

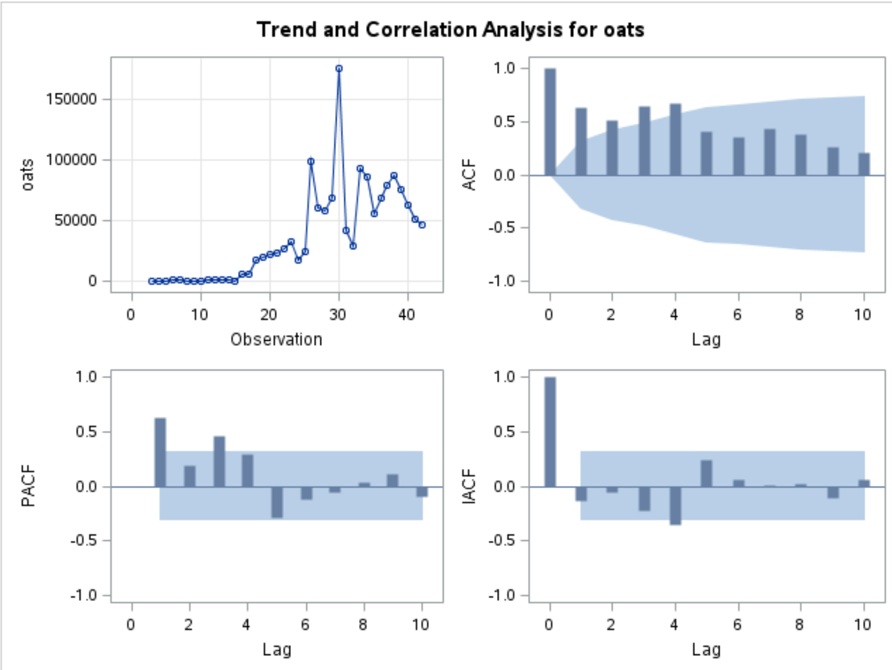

Figure 7-11: Trend and correlation analysis for oats harvest

Table 7-15: Conditional least squares estimation, oats harvest, ARIMA(3,1,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1238.7	1894.1	0.65	0.5175	0
MA1,1	0.72664	0.16539	4.39	0.0001	1
AR1,1	-0.17055	0.19871	-0.86	0.3967	2
AR1,2	0.20998	0.16962	1.24	0.2242	3
AR1,3	0.37337	0.18392	2.03	0.0502	4

Table 7-16: Autocorrelation check of residuals, oats harvest, ARIMA(3,1,1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.77	2	0.2501	0.015	0.048	0.012	0.008	-0.19	-0.14
12	7.95	8	0.4387	0.006	0.105	0.194	0.163	-0.09	-0.107
18	12.55	14	0.5623	-0.09	-0.146	-0.167	-0.112	0.017	0
24	12.87	20	0.8827	-0.014	-0.019	-0.028	-0.033	-0.027	-0.017

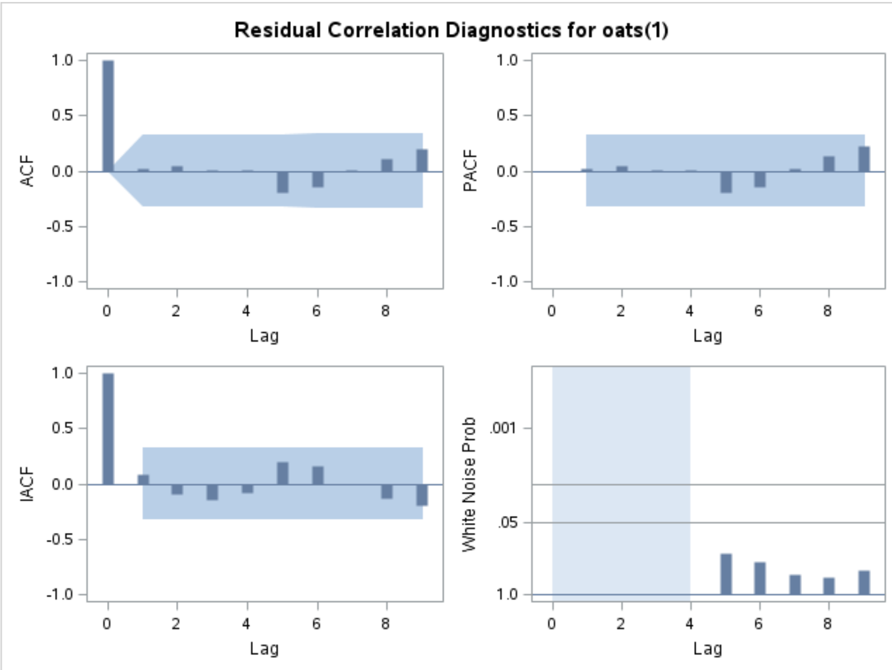

Figure 7-12: Residual correlation diagnostics for oats harvest, ARIMA(3,1,1)

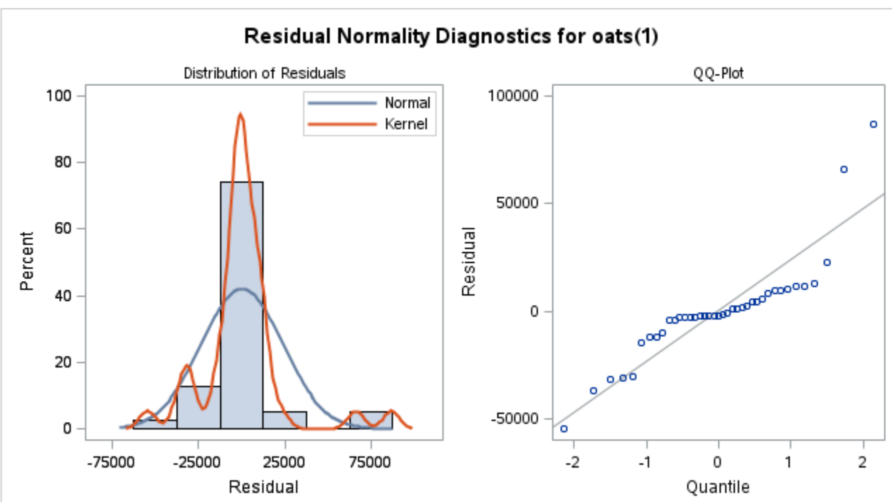

Figure 7-13: Residual normality diagnostics, oats harvest, ARIMA(3,1,1)

7.1.1.5. *Wheat harvest*

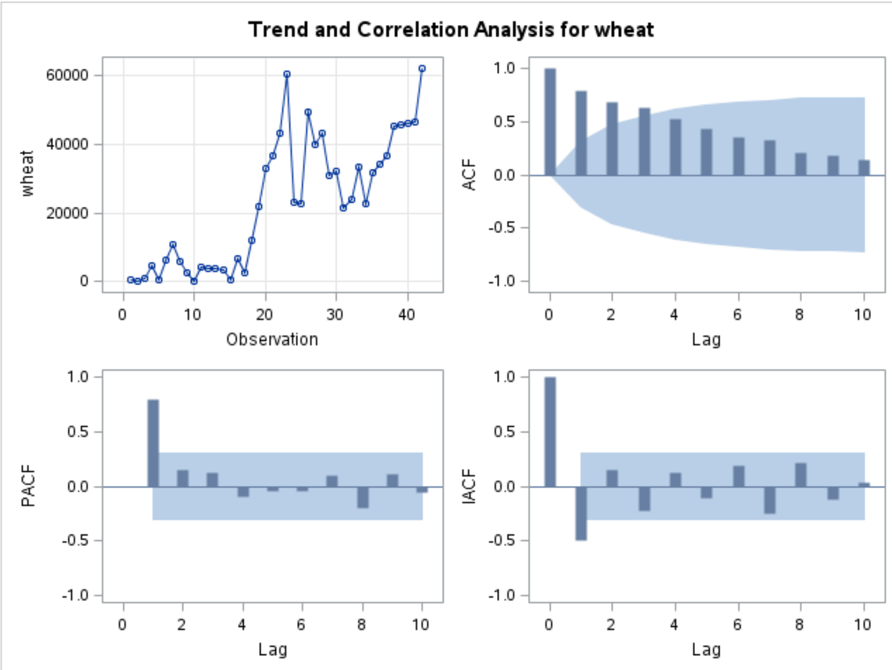

Figure 7-14: Trend and correlation analysis for wheat harvest

Table 7-17: Conditional lest squares estimation, wheat harvest

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	1390.7	944.02981	1.47	0.1487	0
MA1,1	0.37183	0.15301	2.43	0.0198	1

Table 7-18: Autocorrelation check of residuals, wheat harvest

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.06	5	0.8407	0.035	-0.133	0.122	-0.046	0.005	-0.091
12	14.1	11	0.2275	0.167	-0.262	-0.162	0.001	-0.302	-0.048
18	21.76	17	0.1943	-0.158	-0.037	0.087	0.199	-0.012	-0.187
24	27.29	23	0.2438	0.212	0.078	0.037	0.116	0.021	0.016

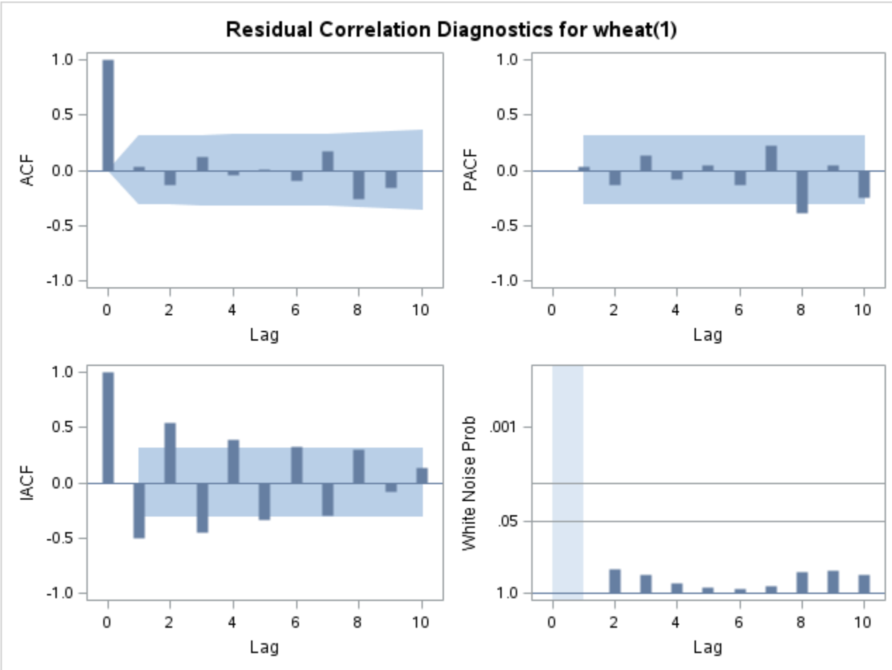

Figure 7-15: Residual correlation diagnostics for wheat harvest, ARIMA(0,1,1)

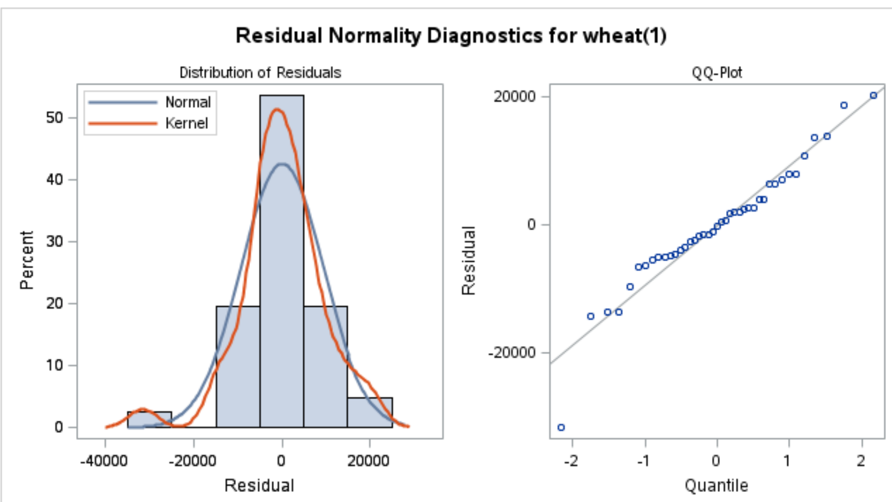

Figure 7-16: Residual normality diagnostics for wheat, ARIMA(0,1,1)

7.1.1.6. Number of houses

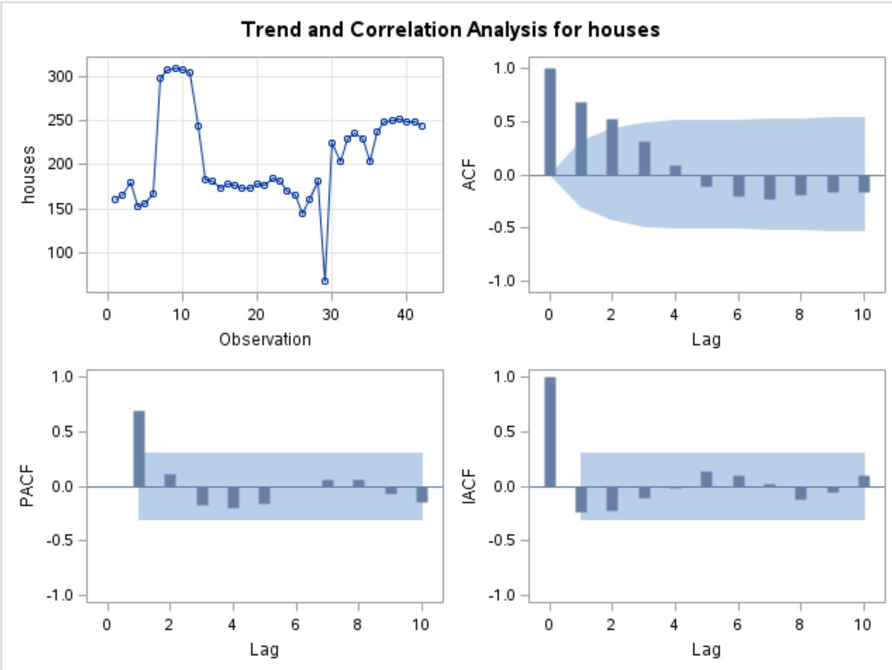

Figure 7-17: Trend and correlation analysis for the number of houses

Table 7-19: Conditional least squares estimation, number of houses, ARIMA(1,1,0)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	2.02913	4.964	0.41	0.6849	0
AR1,1	-0.26615	0.15443	-1.72	0.0927	1

Table 7-20: Autocorrelation check of residuals, number of houses, ARIMA(1,1,0)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.99	5	0.7012	0.013	0.048	-0.013	-0.098	-0.197	-0.1
12	3.44	11	0.9836	-0.087	-0.003	0.022	-0.019	0.007	-0.003
18	7.21	17	0.9807	0.038	0.027	-0.015	0.065	-0.087	-0.19
24	14.79	23	0.9019	-0.098	-0.002	-0.015	-0.109	0.239	0.029

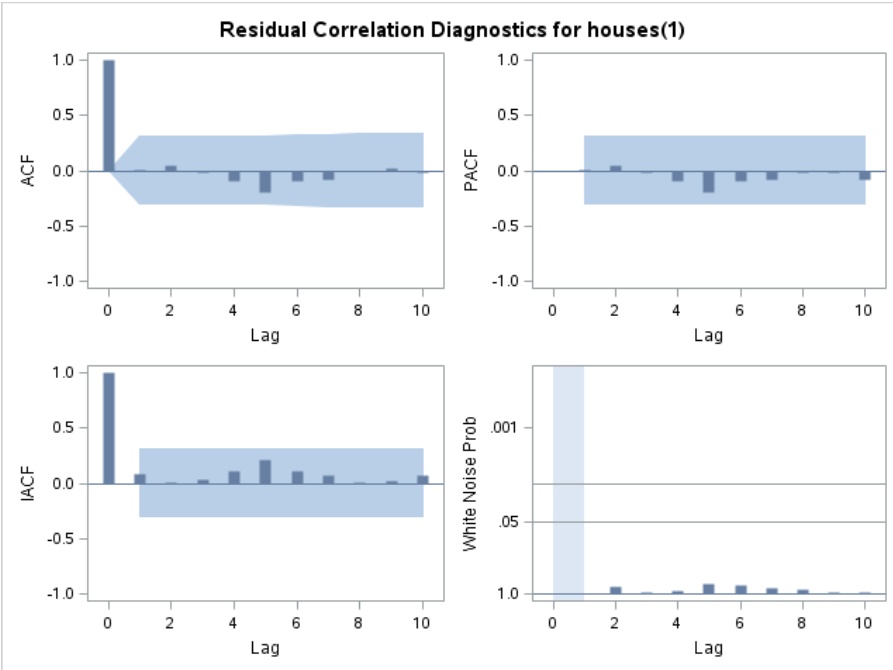

Figure 7-18: Residual correlation diagnostics for the number of houses, ARIMA(1,1,0)

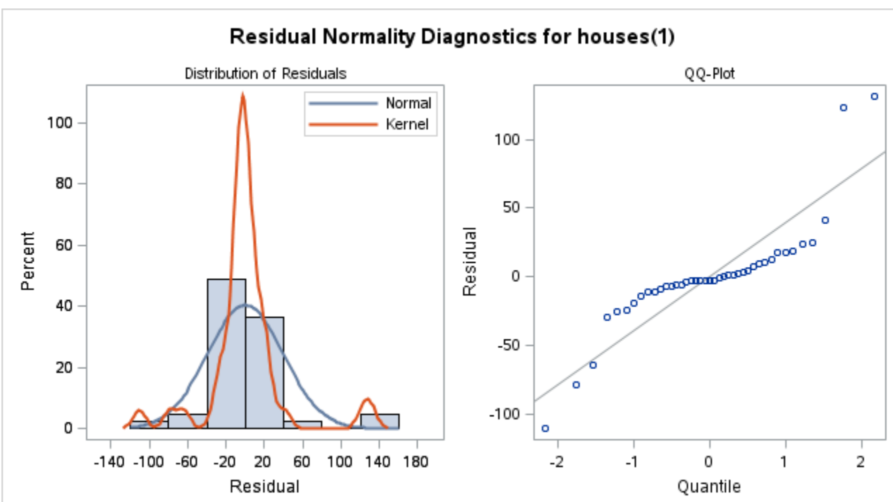

Figure 7-19: Residual normality diagnostics for the number of houses, ARIMA(1,1,0)

7.1.1.7. Number of ploughs

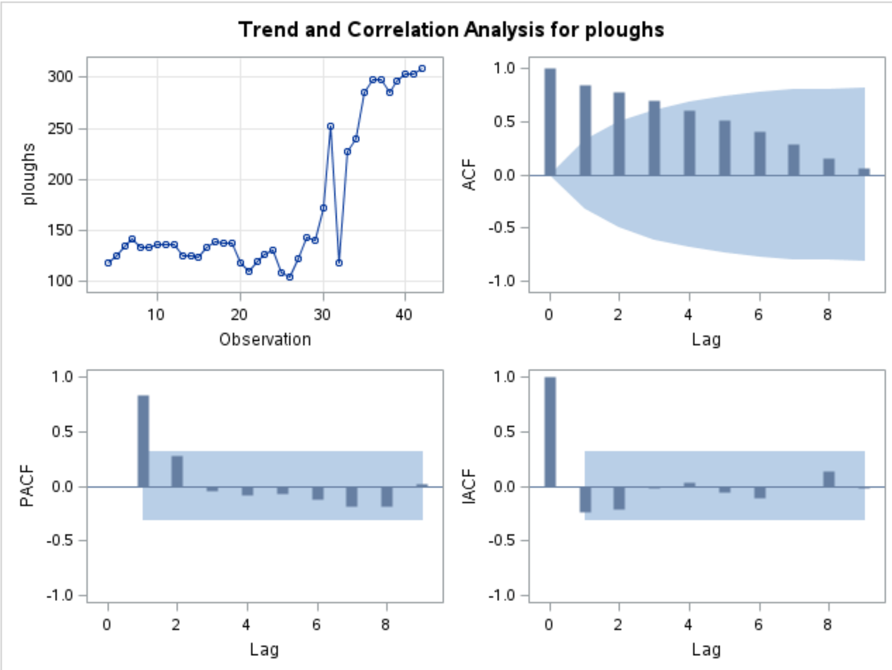

Figure 7-20: Trend and correlation analysis for the number of ploughs

Table 7-21: Conditional least squares estimation, number of ploughs, ARIMA(0,1,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	5.01317	2.6755	1.87	0.0691	0
MA1,1	0.45631	0.14815	3.08	0.0039	1

Table 7-22: Autocorrelation check of residuals, number of ploughs, ARIMA(0,1,1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.62	5	0.6048	-0.107	0.162	0.088	0.072	0.096	0.149
12	5.55	11	0.9017	0.081	-0.074	-0.067	-0.131	-0.025	0.046
18	7.85	17	0.9697	-0.061	-0.086	-0.116	-0.089	-0.023	-0.041
24	9.98	23	0.9914	-0.023	-0.083	-0.072	-0.091	-0.051	0.004

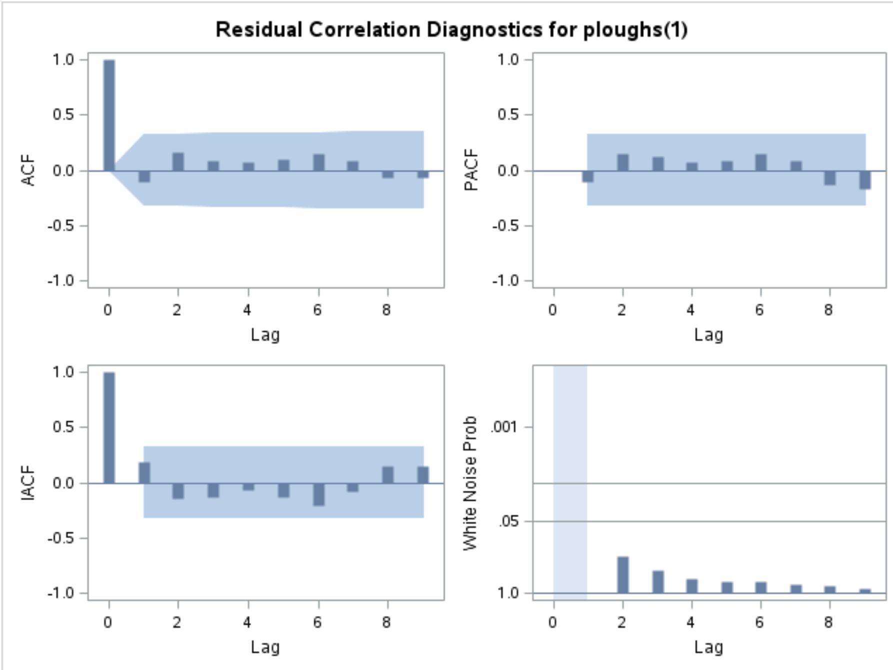

Figure 7-21: Residual correlation diagnostics for the number of ploughs, ARIMA(0,1,1)

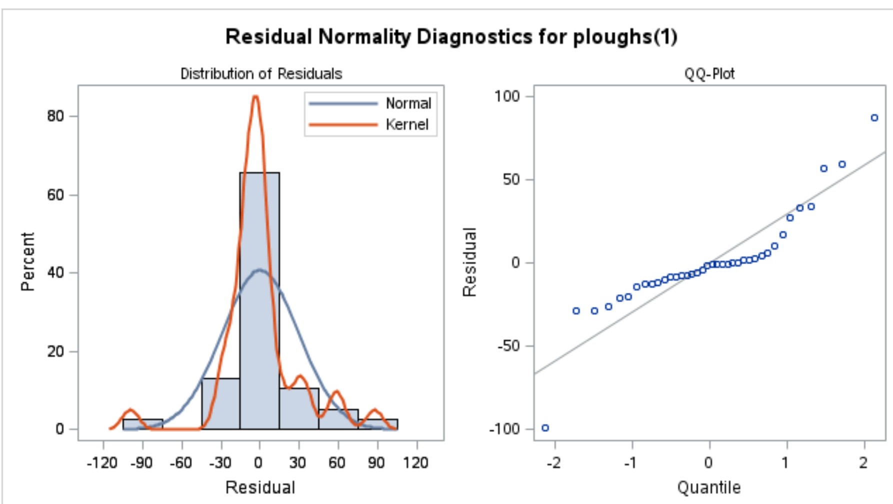

Figure 7-22: Residual normality diagnostics for the number of ploughs, ARIMA(0,1,1)

7.1.1.8. Number of stables

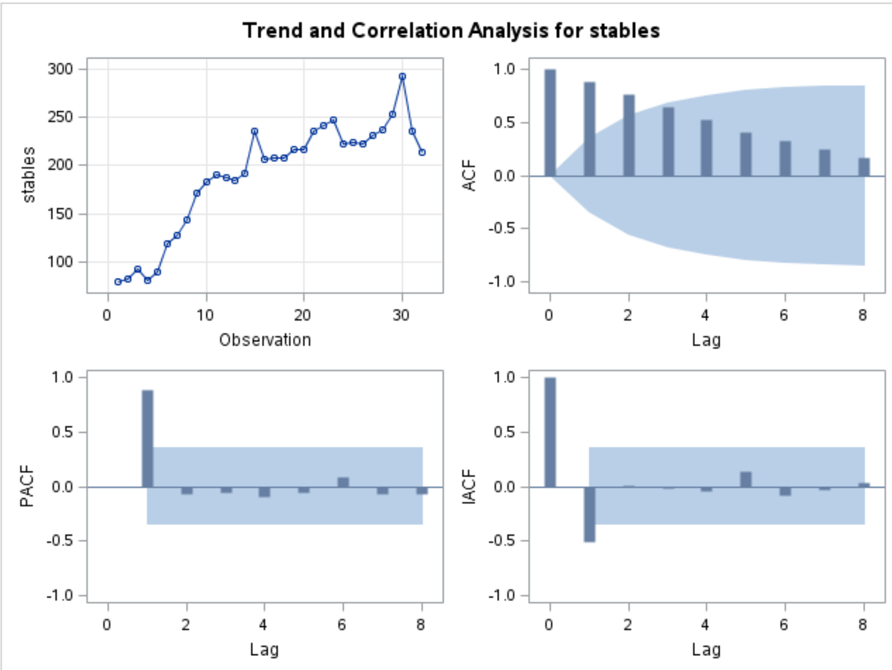

Figure 7-23: Trend and correlation analysis for the number of stables

Table 7-23: Autocorrelation check of residuals, number of stables, ARIMA(0,1,0)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.41	6	0.9653	-0.057	-0.138	-0.061	0.003	-0.096	0.057
12	5.08	12	0.9551	0.125	0.143	0.008	-0.16	-0.091	0.09
18	14.1	18	0.7225	-0.022	-0.044	0.274	-0.158	-0.177	-0.037
24	17.03	24	0.8476	0.057	0.073	0.031	-0.084	-0.092	0.033

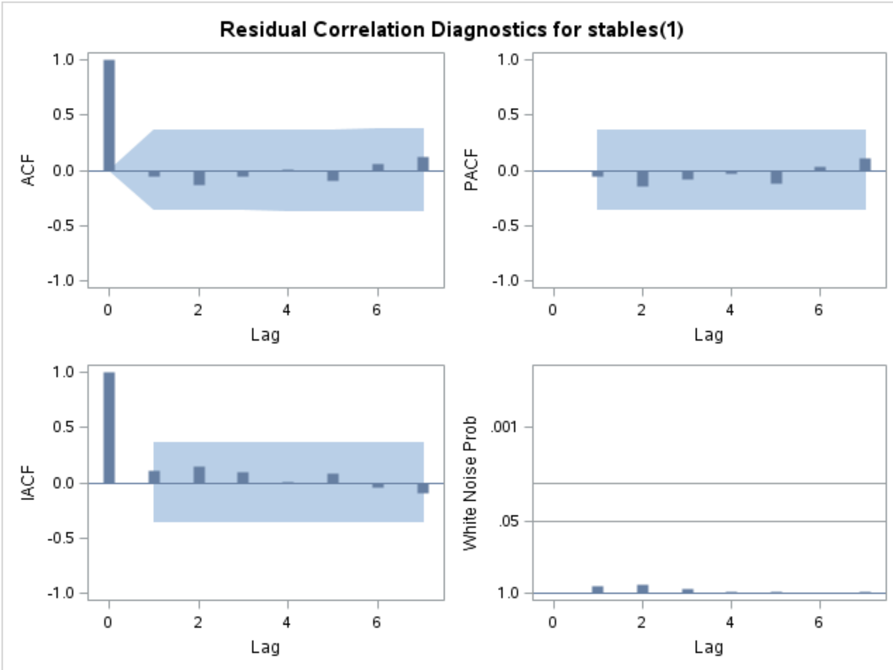

Figure 7-24: Residual correlation diagnostics for the number of stables, ARIMA(0,1,0)

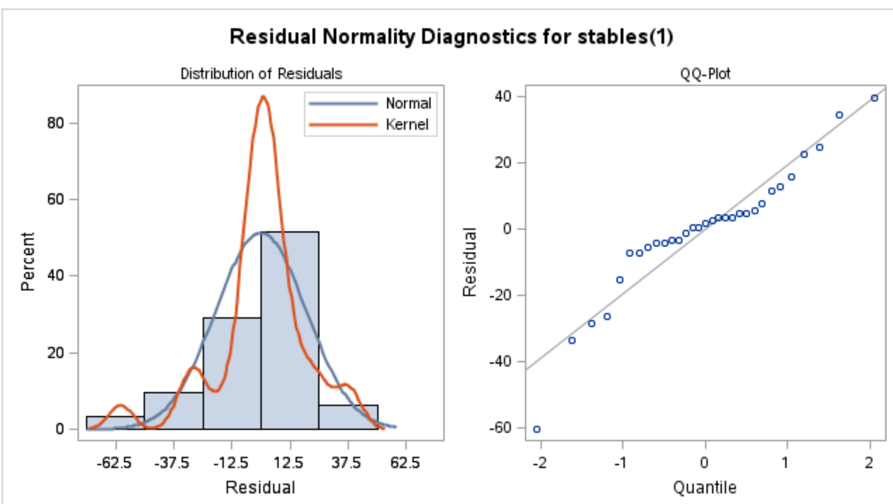

Figure 7-25: Residual normality diagnostics for the number of stables, ARIMA(0,1,0)

7.1.1.9. *Number of horses*

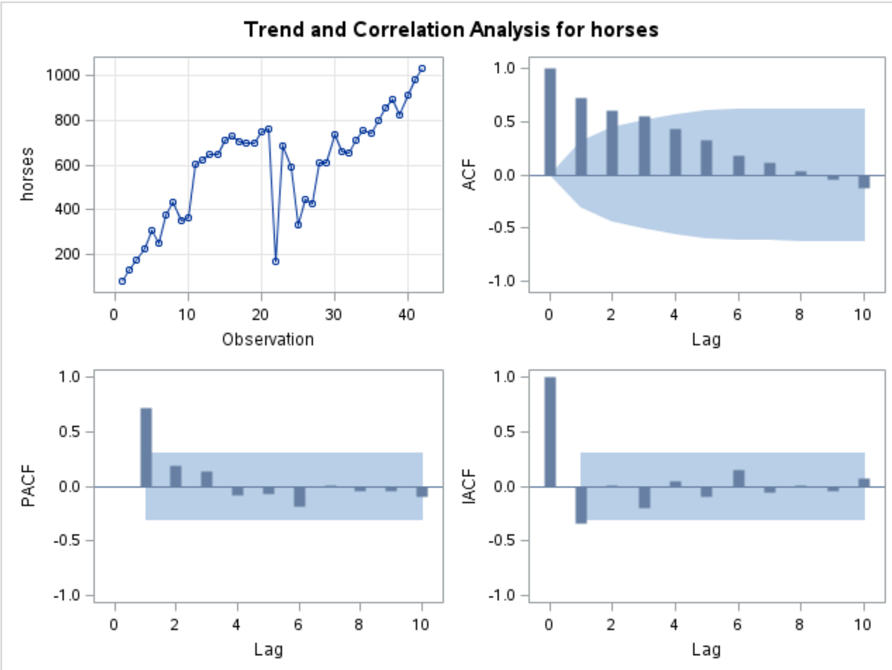

Figure 7-26: Trend and correlation analysis for the number of horses

Table 7-24: Conditional least squares estimation, the number of horses, ARIMA(0,1,1)

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	22.13117	9.89394	2.24	0.0311	0
MA1,1	0.52397	0.13727	3.82	0.0005	1

Table 7-25: Autocorrelation check of residuals, the number of horses, ARIMA(0,1,1)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	4.54	5	0.4743	-0.065	-0.039	0.257	0.099	0.078	-0.096
12	7.79	11	0.7322	0.016	-0.058	-0.039	-0.058	-0.216	0.021
18	10.94	17	0.8595	0.013	-0.169	-0.119	-0.02	-0.023	-0.057
24	12.94	23	0.9534	-0.065	-0.137	-0.018	-0.004	0.005	-0.02

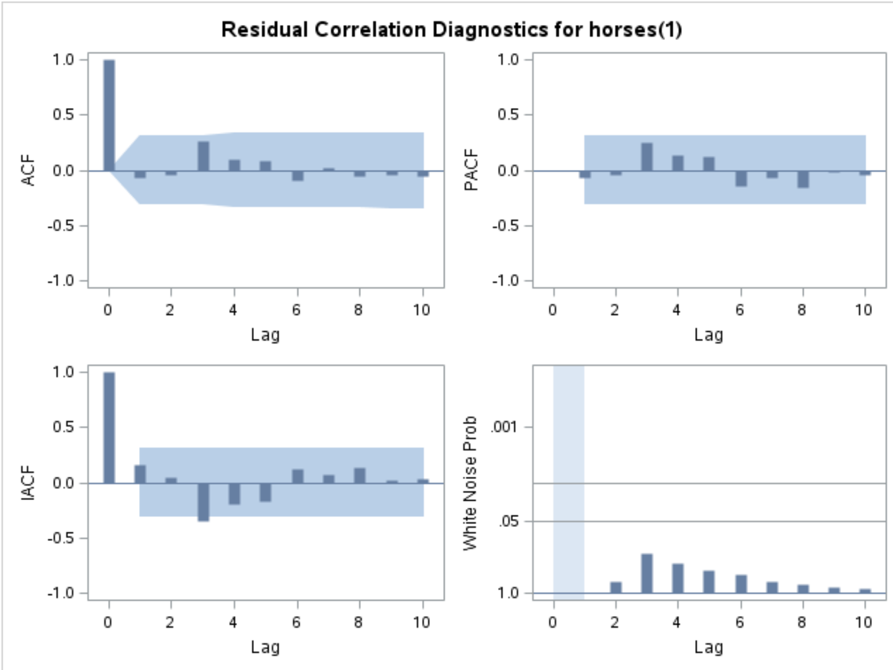

Figure 7-27: Residual correlation diagnostics for the number of horses, ARIMA(0,1,1)

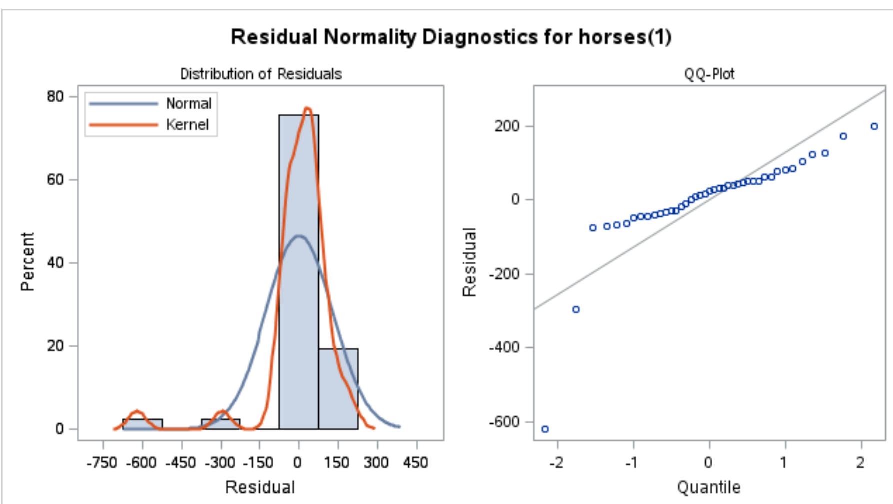

Figure 7-28: Residual normality diagnostics for the number of horses, ARIMA(0,1,1)

7.1.1.10. Number of oxen

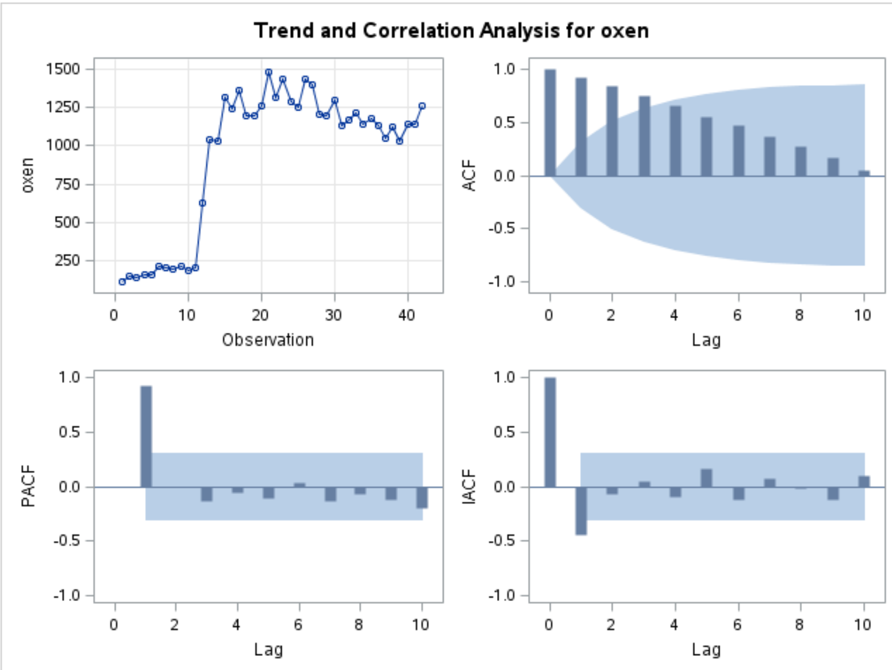

Figure 7-29: Trend and correlation analysis for the number of oxen

Table 7-26: Autocorrelation check of residuals, the number of oxen, ARIMA(0,1,0)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.32	6	0.8878	-0.024	0.136	0.037	0.026	-0.136	0.095
12	4.44	12	0.974	-0.105	0.087	0.059	-0.085	-0.03	-0.089
18	10.77	18	0.9039	0.006	0.001	-0.04	-0.262	0.094	-0.099
24	15.29	24	0.9121	-0.068	0.001	0.03	-0.126	0.022	-0.156

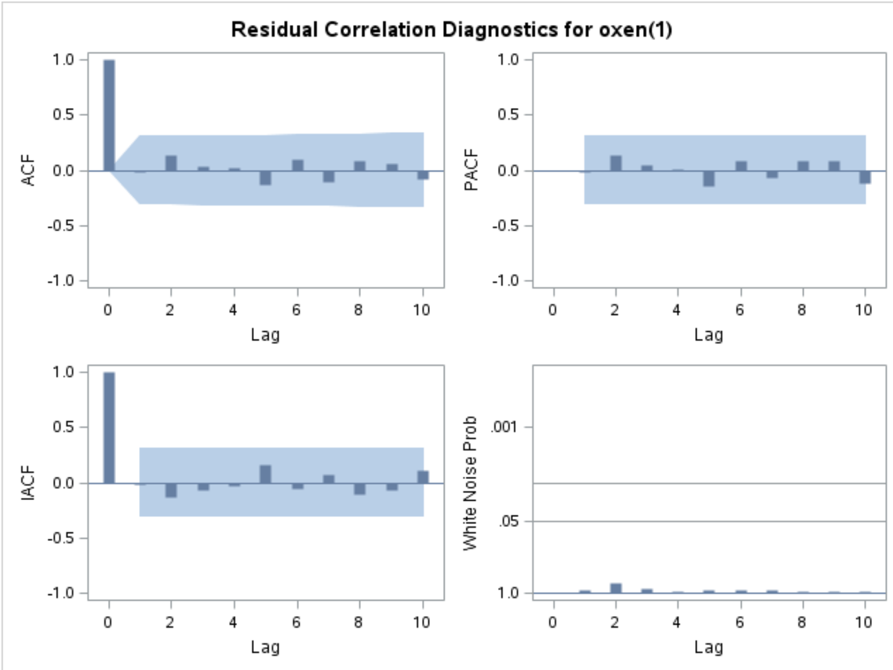

Figure 7-30: Residual correlation diagnostics for the number of oxen, ARIMA(0,1,0)

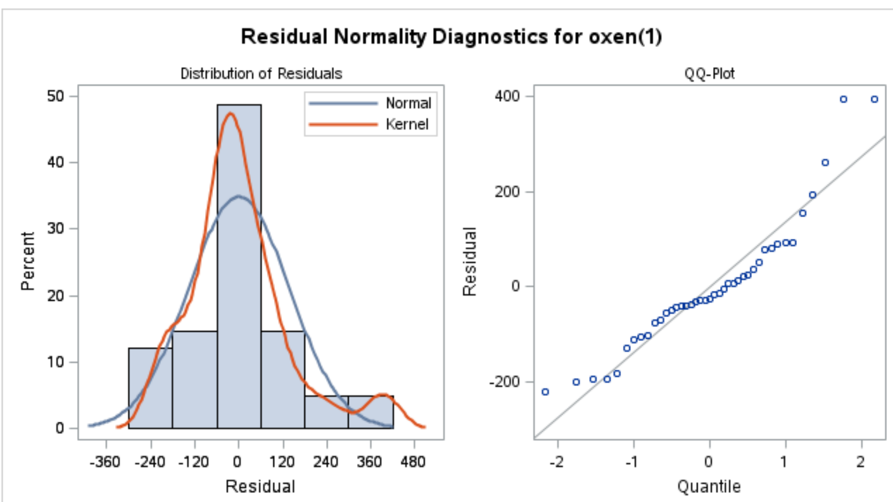

Figure 7-31: Residual normality diagnostics for the number of oxen, ARIMA(0,1,0)

7.1.1.11. Expenditures

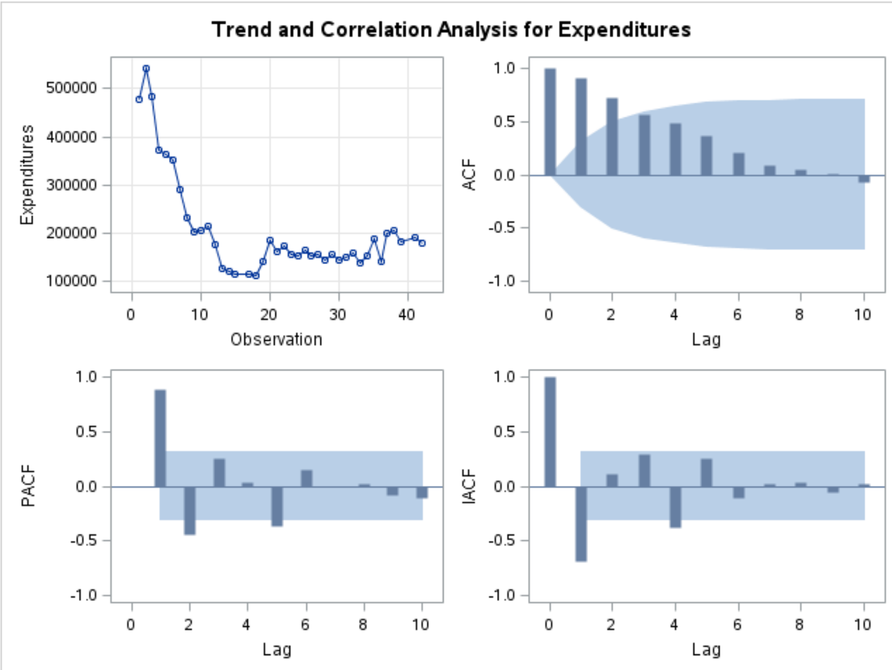

Figure 7-32: Trend and correlation analysis for expenditures

Table 7-27: Autocorrelation check of residuals, expenditures, ARIMA(0,1,0)

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	5.92	6	0.4322	0.104	-0.18	0.256	0.215	0.181	-0.02
12	11.81	12	0.4609	-0.209	0.123	0.304	0.042	-0.254	-0.109
18	15.32	18	0.6398	-0.05	-0.078	-0.015	-0.341	0.075	0.201
24	17.32	24	0.8345	-0.089	0.009	-0.222	-0.232	-0.004	-0.132

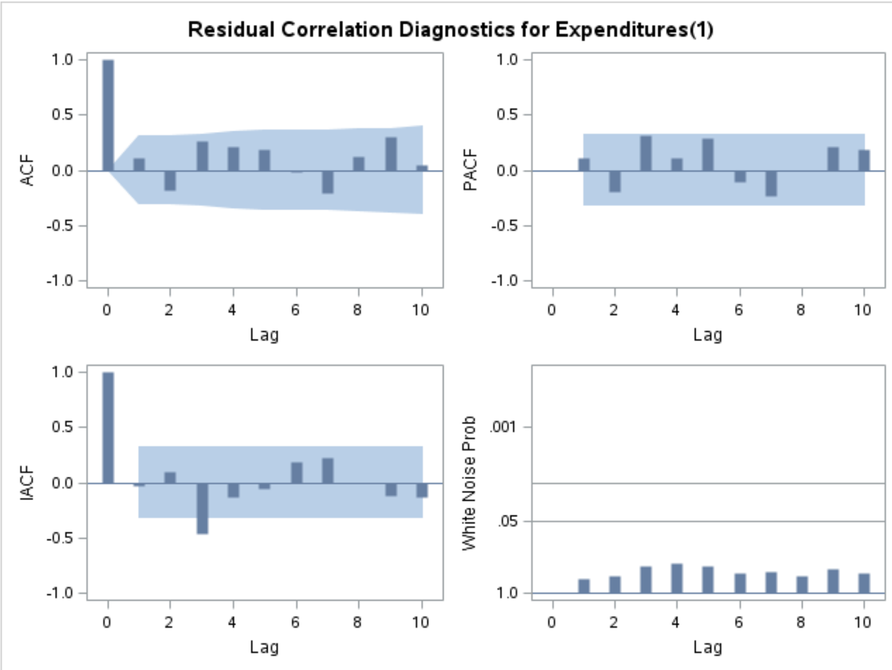

Figure 7-33: Residual correlation diagnostics for expenditures, ARIMA(0,1,0)

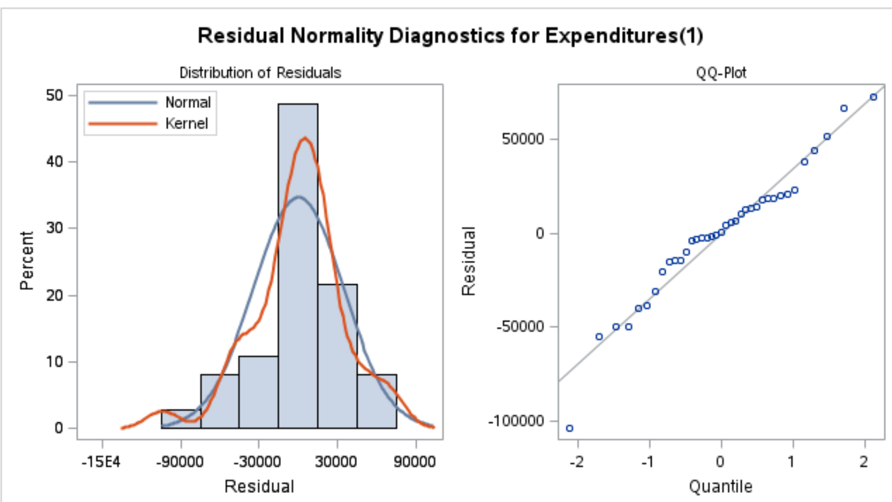

Figure 7-34: Residual normality diagnostics for expenditures, ARIMA(0,1,0)

7.1.2. Outlier analysis

7.1.2.1. Tuberculosis mortality

Tuberculosis mortality data series has outliers in years 1886, 1888, and 1893, corresponding to the first peak of the original data series, the drop in TB mortality following that peak, and the drop in TB mortality following the second peak in year 1891 (Figure 7-35).

TB mortality in year 1886 was 93.4 per 1,000, according to Ferguson. This number appears high when compared to data from other populations. For example, TB mortality in England and Wales in years 1891-1900 was 2 per 1,000 (McKeown & Record, 1962). In Russia, TB mortality in the outskirts of Moscow was 6.6 per 1,000 in years 1895-1904 (Prohorov.B.B., 2000). The rate of 93.4 deaths per 1,000 appears relatively low, when compared to TB mortality among the Inuit. In 1889, only three years after the peak of TB mortality in the Qu'Appelle Agency, half of the Inuit of Kuujuaq died from pulmonary diseases (Grygier, 1994) (p.56), corresponding to a rate of 500 per 1,000. In Coppermine, out of the 38 TB patients in October, 26 were dead by March (Grygier, 1994) (p.57). Since 25% of the population of Coppermine was infected with TB (Harper, 1992), mortality rate was 171 per 1,000. Other estimates put the population of Coppermine at around 100 (Usher, 1965), in which case mortality was around 260 per 1,000. In light of these rates, mortality rate of 93.4 per 1,000 does not seem impossible.

In addition to this, population of the Qu'Appelle Agency dropped from 1317 in 1886 to 1104 in 1887, corresponding to a total mortality rate of 161.7 per 1,000. Given that TB mortality data come from death certificates and were reported in Ferguson, while population data come from the census and were reported by the Department of Indian Affairs, the two sources of data are independent. In light of this, it would not be justified to drop observations corresponding to the first two peaks in TB mortality.

7.1.2.2. Years 1914 and 1915

The following data series have outliers in year 1914: the number of houses and stables, potato and oats harvest, and amount of hay cut. In 1915, outliers are present in oats harvest, number of stables and ploughs, and size of area farmed. All of these data series, with the exception of the number of houses, are related to agriculture. The number of houses has an outlier in 1914 because of the low number of houses in 1913, and the transformation of the data during ARIMA model fitting. Because of that, number of houses will be examined separately.

Since all of the outliers in years 1914 and 1915 are either agricultural indicators or are significantly correlated with them, environmental conditions may be an important place to start. Years 1914 and 1915 have the highest average monthly temperature in July and August, respectively (Figure 7-36), and were third and second wettest years among the five years examined (Table 7-28).

In addition to weather data, comparison with agricultural output in the Blackfoot Agency suggests that agricultural output profiles were not very different (Figure 7-37, Figure 7-38, Figure 7-39). Since the growing season was relatively warmer, and years 1914 and 1915 were relatively wet, and profiles of agricultural output around years 1914/15 were similar in Qu'Appelle and Blackfoot Agencies, it would seem that high agricultural output in years 1914/15 was likely real, and not attributable to error. Because of this, as well as the general recommendation not to drop observations unless there is a very good reason to do so, oats, potato, and hay harvest data points for years 1914 and 1915 should be retained.

7.1.2.3. Area farmed

In addition to the outlier in year 1915 discussed above along with a number of other outliers, area farmed data series has an outlier in year 1920 (Figure 7-40). The outlier in year 1920 corresponds to an outlier in the number of growing degree days, and an outlier in the number of ploughs in the same year. This suggests that year 1920 was very warm, perhaps because of that there was a significant agricultural expansion, as is evidenced by a growth in the number of ploughs and area farmed. The fact that there was a corresponding increase in the number of ploughs as well as a significant jump in the number of growing degree days suggests that the outlier in year 1920 should be retained for further analysis.

7.1.2.4. Potato harvest

In addition to outliers in this data series discussed above, potato harvest has outliers in years 1888, 1899, and 1902.

The 1888 outlier in the potato harvest data series coincides with an outlier in the expenditures data series. In year 1888 expenditures dropped significantly. Given that Indian Agents would cut expenditures in years in which harvest was good, the drop in expenditures supports the possibility of the outlier in the potato harvest data series being real, as opposed to it being a mistake in record keeping. In addition, Indian Agent reported good harvest in that year:

> "The crops this season are looking well, with every prospect of a bountiful harvest, the work was thoroughly done entirely with Indian labor, under the supervision of the farming instructors, and will compare favorably with any settlers crops in the district. The fencing on the whole could not be improved upon."(DIA., 1889)

For these reasons the 1888 data point in potato harvest data series should be retained for analysis.

Potato harvest data in years 1891- 1899 follow similar profiles in Qu'Appelle and Blackfoot Agencies (Figure 7-38), suggesting that the outlier might be real, as opposed to be caused by recording errors, as the two Agencies were located in different provinces.

Finally, the last outlier in potato harvest series is found in year 1902. Year 1902 was very warm – the number of heat days data series has an outlier in that year as well. There was also a small peak in the potato harvest data series in the Blackfoot Agency, suggesting that the 1902 data point should not be dropped from analysis.

7.1.2.5. Hay cut

In addition to the outlier in the hay cut data series discussed along with other outliers in years 1914-1915, data series has one outlier in year 1907, when reported amount of hay cut was lowest since 1886. The amount of hay cut also dropped in the Blackfoot Agency (Figure 7-39). Year 1907 appears to have been slightly colder than years immediately before and after (Figure 7-41). 1907 was also slightly drier than other years (Figure 7-42), with the exception of 1907, when the amount of hay cut also dropped (Figure 7-39). Because of the presence of environmental reasons, as well as similar data profiles for the Blackfoot Agency, it is likely that the 1907 data point is not a case of incorrect reporting and should be retained for further analysis.

7.1.2.6. Wheat harvest

Wheat harvest data series has outliers in years 1907, 1908, and 1910, corresponding to the first significant peak of the series, the subsequent drop in the amount of hay harvested, and a significant increase from year 1909 to 1910 (Figure 7-43).

The significant drop in wheat harvest could be attributed to frost in two subsequent years (DIA., 1909). Because of that, it appears that the three outliers were caused by natural events and should be retained for further analysis.

7.1.2.7. Oats harvest

Oats harvest data series has outliers in years 1910, 1914, 1915, 1916, 1918, 1925, and 1926. Outliers in years 1914 and 1915 were discussed above (Figure 7-37). The outlier in year 1910 is caused by a sharp increase in oats harvest after a particularly low harvest the year before. Like the outlier in wheat harvest in year 1910 discussed above, it appears that the low harvest in years immediately before could be attributed to frost.

Agricultural output in Qu'Appelle and Blackfoot Agencies demonstrates high variability from year 1910 onwards. Outliers in years 1916, 1918, 1925, and 1926 are part of this pattern. Some of these drops could be explained by natural factors, like frost. In other years, other factors played a role as well. Writing in year 1916, Agent Nichol said that:

> "Owing to unfavourable seasons, high wages outside, low prices for grain, and a ready market for wood, hay and pickets, the Indians lost considerable interest in farming for a time. They are, however, again realizing the benefits to be derived from farming, and the past two seasons show a marked increase in these operations." (DIA., 1917)

Similar patterns in agricultural output across the two Agencies and presence of other factors that affect planting decisions suggest that these outliers should be included for further analysis.

7.1.2.8. Number of horses

Number of horses data series has outliers in years 1895, 1906, and 1909 (Figure 7-44). Outlier in year 1895 is caused by a significant increase in the number of horses between years 1894 and 1895. The number of horses in year 1894 was lower than in years 1892 and 1893, which is why the data point for year 1895 appears as at outlier. Writing in 1895, Agent Lash said:

> "The winter was most favourable for stock, and they turned out in the spring in good condition, the increase being most satisfactory, especially in the government herd. From this band we supply all our own beef and work oxen, and have a surplus of thirty heifers for transfer to some other agency. The pedigree Galloway bulls have proved hardy, and are suitable in every respect for this country. At the round up of the government herd we branded ninety calves from a total of one hundred and three cows, and they are the finest lot of calves we ever had." (DIA., 1896)

This statement provides support for the number of horses in that year, suggesting that it should be retained for further analysis.

Potential outliers in years 1906 and 1909 were caused by significant drops in the number of horses from 761 in 1905 to 168 in 1906 in the recorded data. Agent Lash who oversaw the Agency in year 1906 did not mention a reduction in the number of horses. Quite to the contrary, he wrote: “There are a goodly number of good work horses on this reserve, and also quite a few Indian ponies.” (DIA., 1906b) Since there was no mention of the reduction in the number of horses in the Annual Report, while “a goodly number of good work horses” is mentioned several times in the report, it is likely that the outlier in year 1906 was caused by an error.

Potential outlier in year 1909 is caused by a significant drop in the number of horses from 592 in 1908, to 334 in year 1909 (Figure 7-44). Indian Agent did not mention a drop in the number of horses in 1909, but the number of horses in subsequent year is still below that in 1908, and does not reach similar level until 1912. It seems unlikely that a typo or another mistake was made three years in a row, suggesting that the 1909 data point might not be an outlier (Figure 7-44).

7.1.2.9. Number of oxen

The number of oxen data series has outliers in years 1896, 1897, and 1899 (Figure 7-45). The three outliers are in the beginning of a significant increase in the number of oxen, followed by a plateau, when the number of oxen stayed relatively constant. The number of oxen increased by 420 from 1895 and 1896, of which almost a third was due to an increase in just one herd (DIA., 1897.). The significant increase in the number of oxen is also consistent with an exponential increase in other indicators that took place at the end of 19th, and the beginning of 20th centuries. Because of that, the three potential outliers in the number of oxen data series should be retained for further analysis.

7.1.2.10. Number of houses

The number of houses has potential outliers in years 1891, 1892, 1896, 1897, 1913, and 1914. The first four outliers create the bulge in the plot (Figure 7-47), and the last two correspond to the sharp drop in year 1913 and the subsequent increase in the number of houses.

The “bulge” in the number of houses between the years 1891 and 1896 exists in the corresponding data for the Blackfoot Agency (Figure 7-46). Unlike data series that can be expected to fluctuate on a seasonal and annual basis, such as agricultural output, the number of houses is expected to demonstrate less volatility. In light of that, the "bulge" in the plot of the data series is worth further examination. It could either be real, in that there was a significant construction boom across the two Agencies in year 1891, and a significant loss of houses in year 1896. The number of houses in the Qu'Appelle Agency before the "bulge" was 167, increases to 309 at the highest point of the "bulge", and returns to 183 in 1897. The number of houses is relatively stable before, during, and after the increase in 1891 and decline in 1896, suggesting that data points for years 1891, 1892, 1896, and 1897 could be real outliers.

Similarly, the sharp drop in the number of houses in year 1913 from 182 to 68, and subsequent increase to 224 suggests that the drop is not likely to occur naturally in normal circumstances. If data points for years 1891, 1892, 1896, 1897, 1913, and 1914 represent real outliers, and using linear extrapolation to replace them, alternative data series are presented in Figure 7-47.

7.1.2.11. The number of ploughs

The number of ploughs data series has outliers in years 1915-1920. The number of ploughs jumped from 172 in 1914 to 252 in 1915, and then dropped to 118 the following year, creating the first two outliers. Increase from 172 to 252 is significant, but appears to be possible, as the number of ploughs continues to increase thereafter (Figure 7-48). The drop in the number of ploughs in year 1916 appears less likely, as there is nothing in the report of the Indian Agent that would suggest that the number of ploughs declined. Quite on the contrary, he wrote that:

> "The Indians on the whole are well supplied with farm implements, Standing Buffalo more particularly in this respect. They take only fair care of their machinery and implements. A threshing outfit is owned by the four bands in common." (DIA., 1917)

Data points for years 1917 to 1920 are unlikely to be outliers because of the generally fluctuating nature of the number of ploughs data series (Figure 7-49).

7.1.2.12. The number of stables

The number of stables has outliers in years 1899, 1900, 1908, 1914, 1915, and 1916 (Figure 7-50). Profiles of the number of stables data series are similar in the Qu'Appelle and Blackfoot Agencies, suggesting that they were driven by the same set of factors. The outlier in the number of stables in year 1899 corresponds to an outlier in the number of oxen in the same year, suggesting that the increase in the number of stables in 1899 was real.

Similarly, there is a correspondence between changes in the number of the number of oxen and the number of stables in the Qu'Appelle Agency, suggesting that these outliers should be retained for further analysis (Figure 7-51).

7.1.2.13. Expenditures

Expenditures data series has outliers in years 1886 and 1888 (Figure 7-52). The outlier in year 1888 corresponds to a high potato harvest in the same year, suggesting that the drop in expenditures was driven by a good crop of potatoes, as discussed above.

The potential outlier in year 1886 represents the first peak in expenditures, after which expenditures began their long decline. Given the general profile of the expenditures data series and that expenditures were significantly lower in all subsequent years, the peak had to have been either in year 1885 or 1886, and with no evidence from other sources it would be prudent to retain the 1886 data point for further analysis.

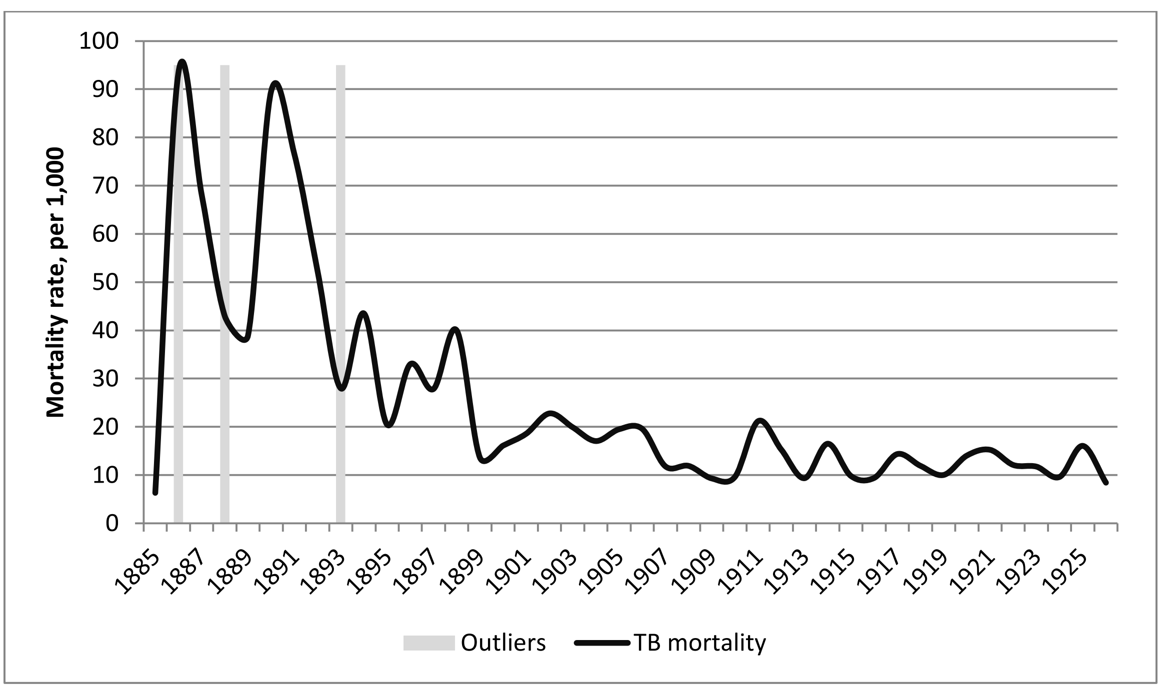

Figure 7-35: Tuberculosis mortality, Qu'Appelle Agency. Source: Ferguson (1928)

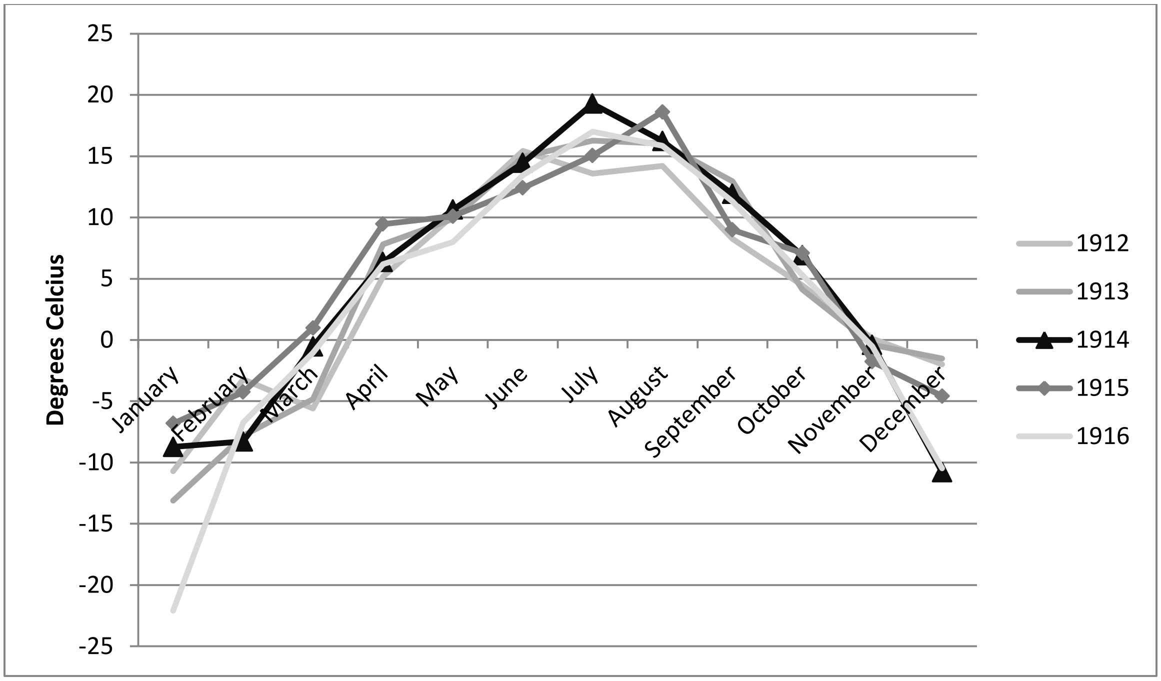

Figure 7-36: Average monthly temperatures, Regina. Source: (Environment Canada., 2013)

Table 7-28: Total precipitation in millimetres, by month, Regina. Source: (Environment Canada., 2013)

	1912	1913	1914	1915	1916
January	11.8	32.8	23.6	10.7	20.1
February	2.2	14.3	29.9	5.8	15.4
March	8.9	12.8	20.8	1.9	19.3
April	51.3	5.4	14.8	14.7	21.7
May	43.6	57.8	13.3	76.4	78.9
June	109.7	99.6	67.1	105.6	37.2
July	132.1	15.6	64.1	97.5	37.9
August	68.9	148.2	55.3	19.8	51.6
September	71.2	5.8	28.2	104.7	21.3
October	27.4	17	46.2	9.4	30.4
November	17.2	24.9	69.8	11.9	7
December	0	0	19.9	7.5	12.7
Yearly average	45.35833	36.18333	37.75	38.825	29.45833

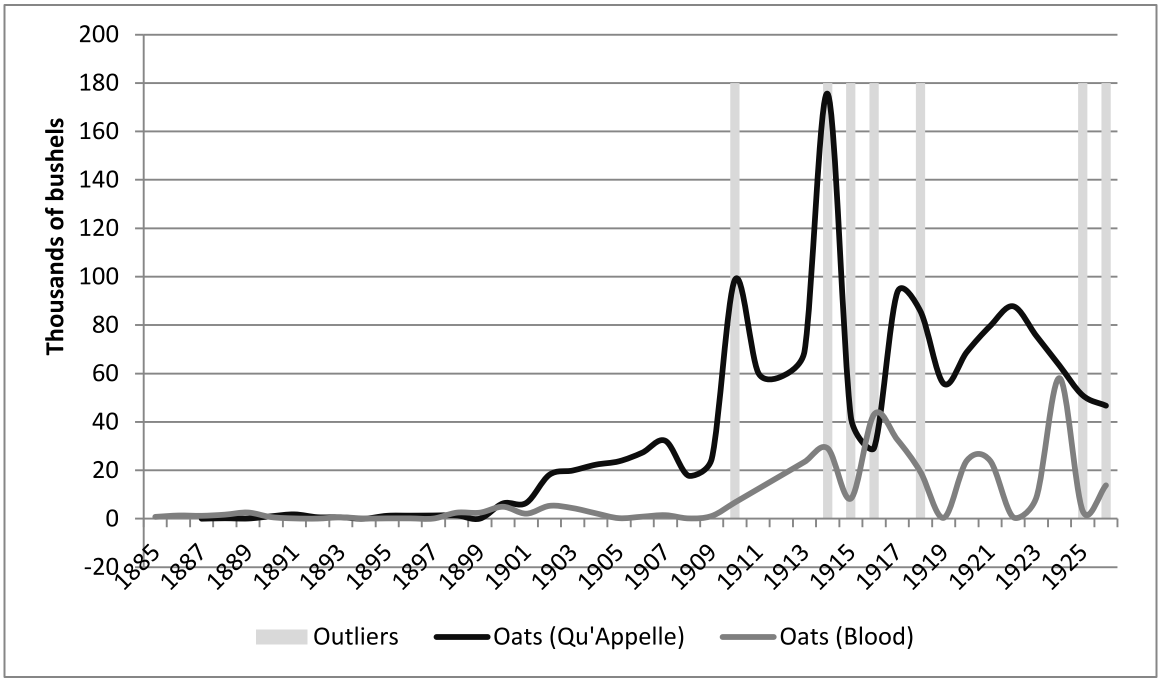

Figure 7-37: Oats harvest, Qu'Appelle and Blackfoot agencies. Source: DIA(1885-1926)

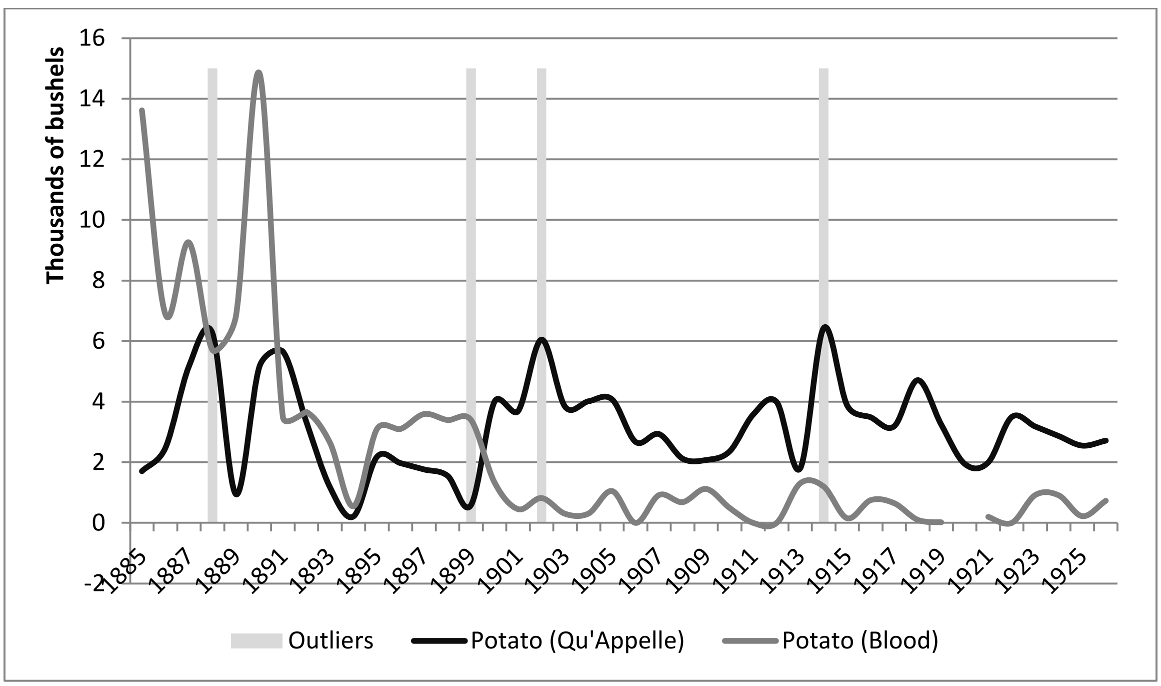

Figure 7-38: Potato harvest, Qu'Appelle and Blackfoot agencies. Source: DIA(1885-1926)

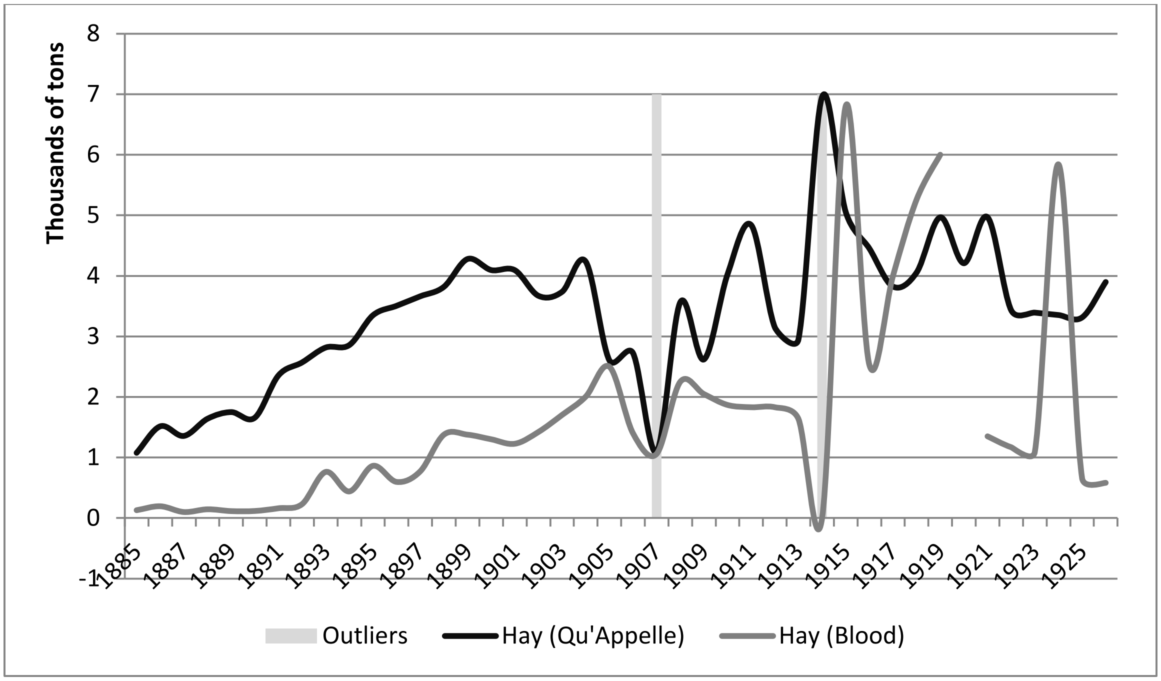

Figure 7-39: Hay harvest, Qu'Appelle and Blackfoot agencies. Source: DIA(1885-1926)

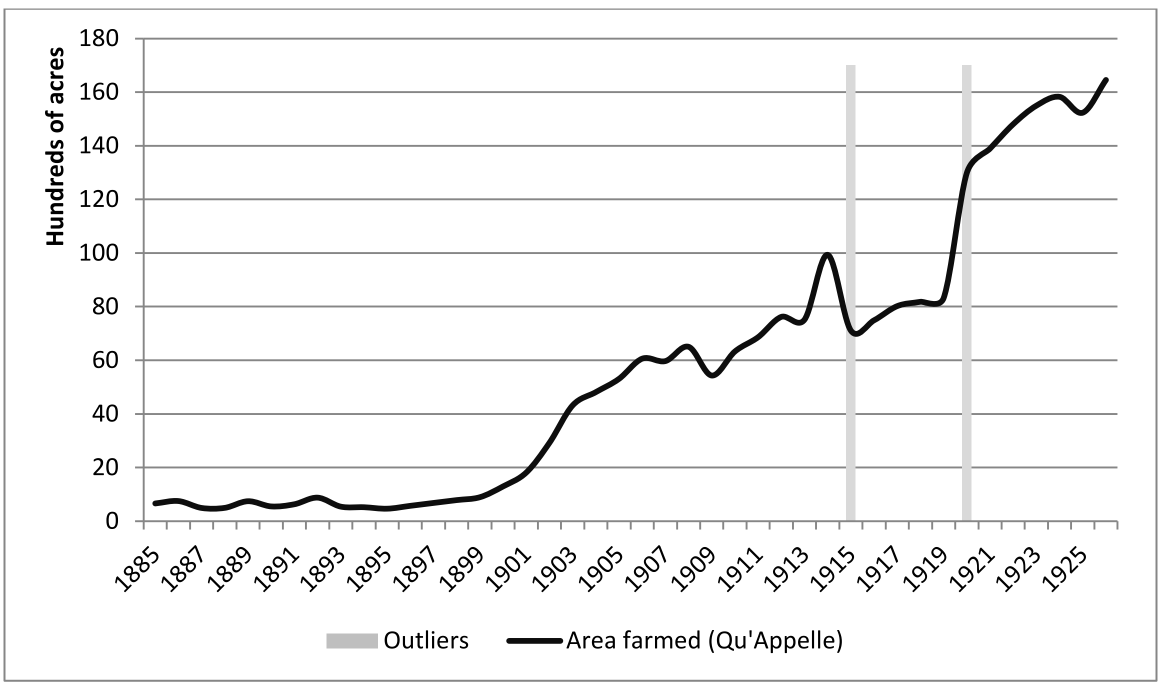

Figure 7-40: Area farmed, Qu'Appelle Agency. Source: DIA(1885-1926)

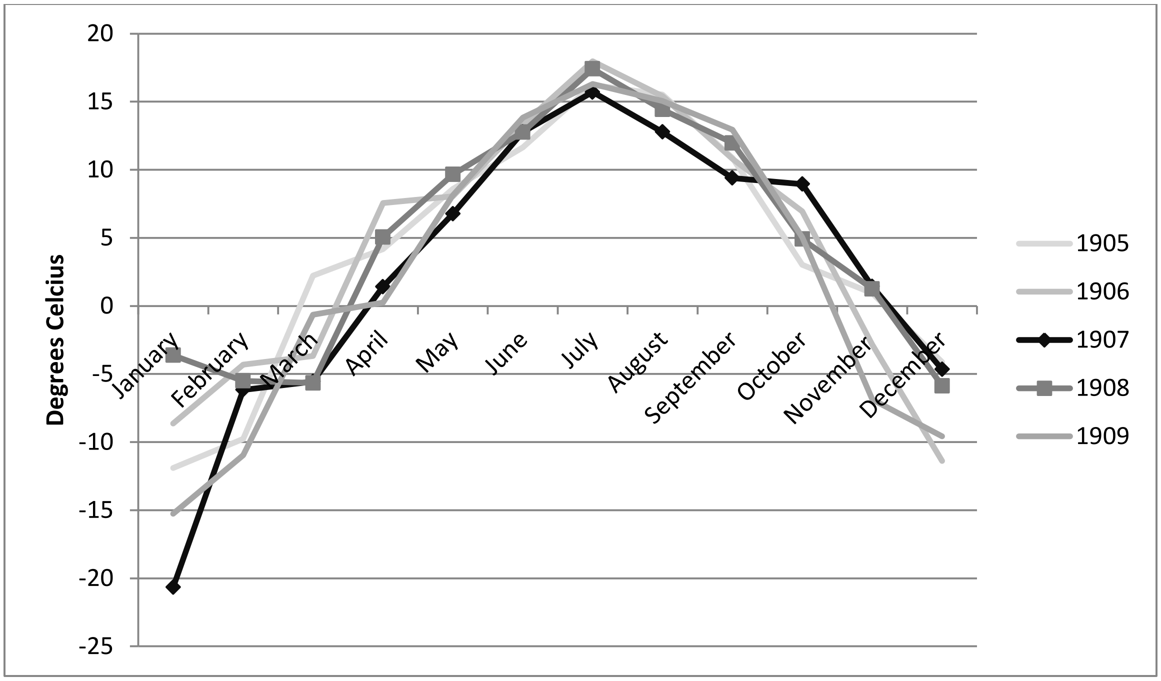

Figure 7-41: Average monthly temperatures, Regina. Source: (Environment Canada., 2013)

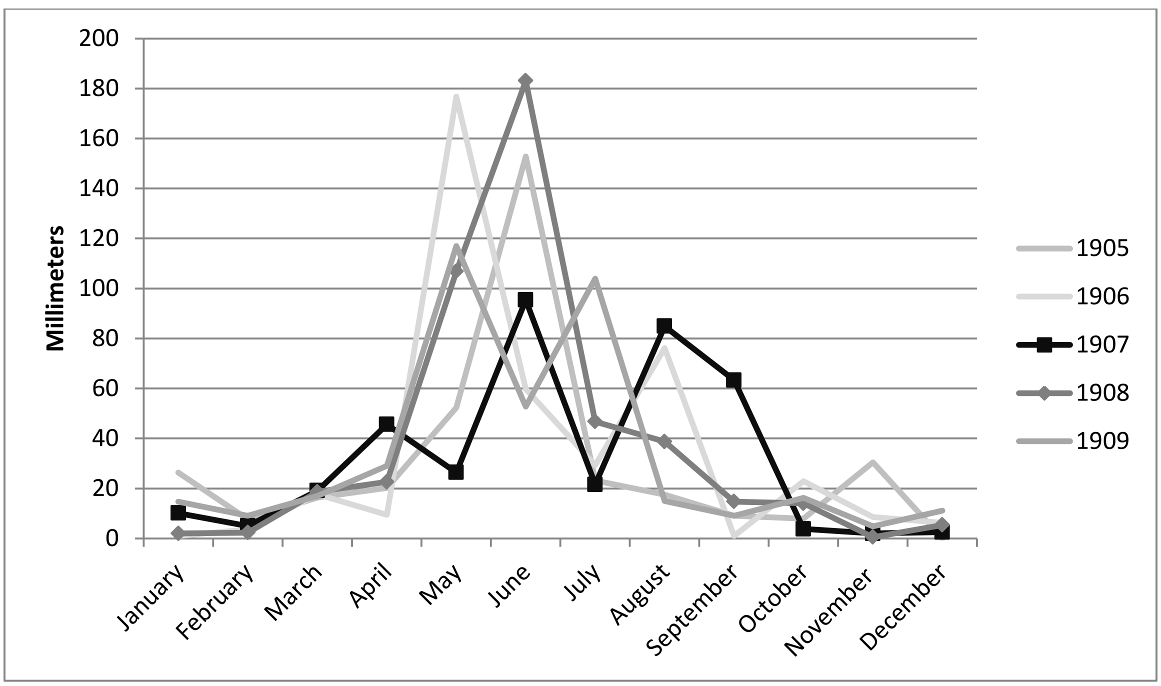

Figure 7-42: Total precipitation, by month, Regina. Source (Environment Canada., 2013)

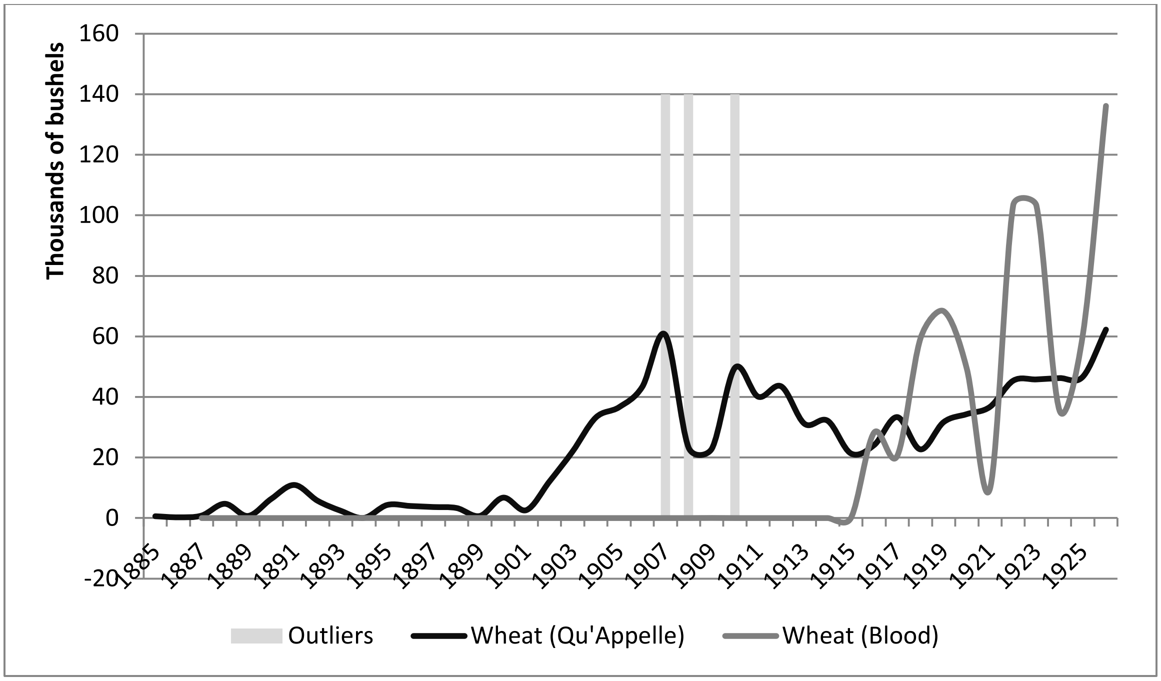

Figure 7-43: Wheat harvest, Qu'Appelle and Blackfoot Agencies. Source: DIA(1885-1926)

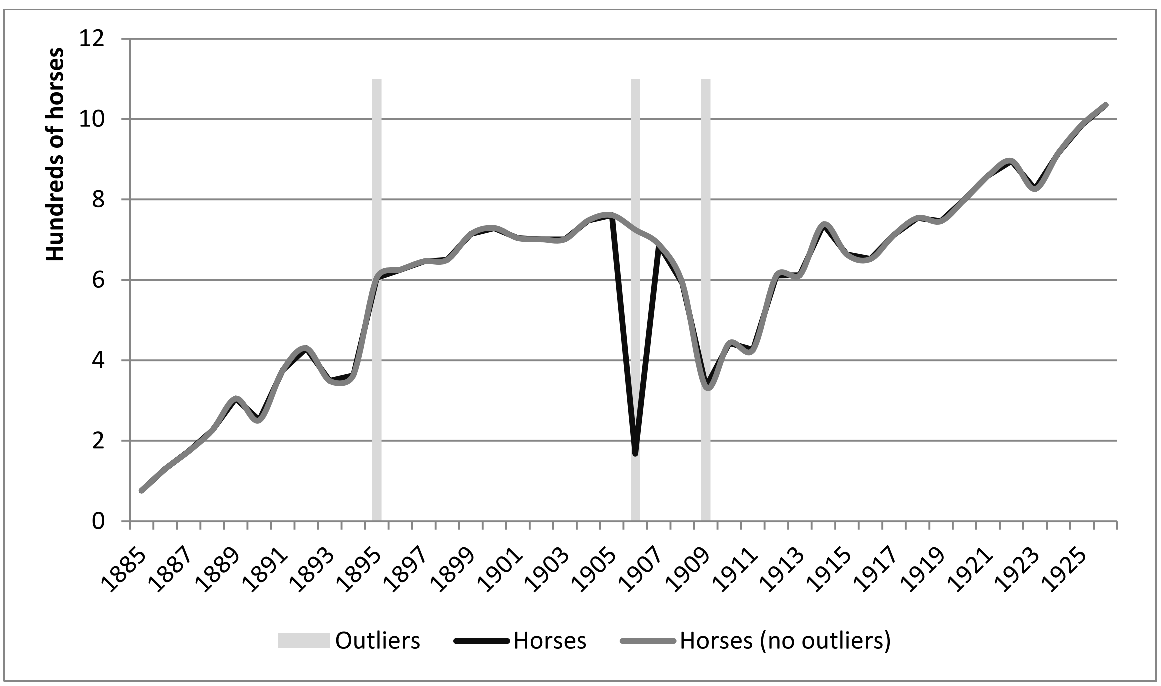

Figure 7-44: Number of horses with and without potential outliers, Qu'Appelle Agency. Source: DIA(1885-1926)

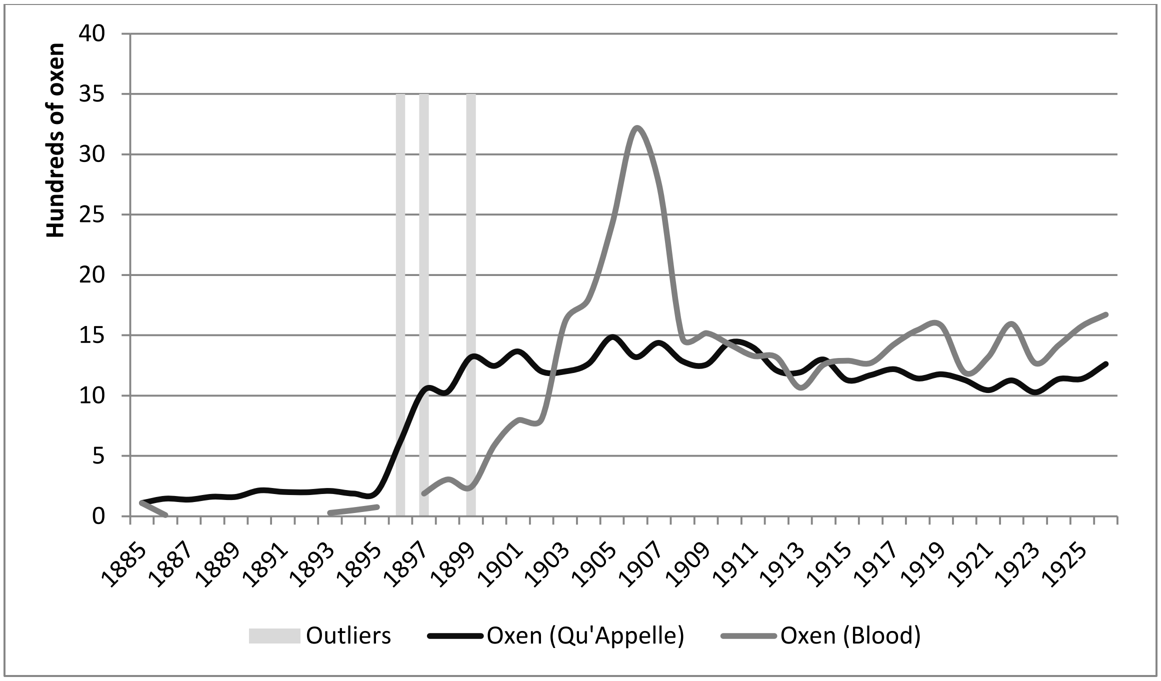

Figure 7-45: Number of oxen, Qu'Appelle and Blackfoot Agencies. Source: DIA(1885-1926)

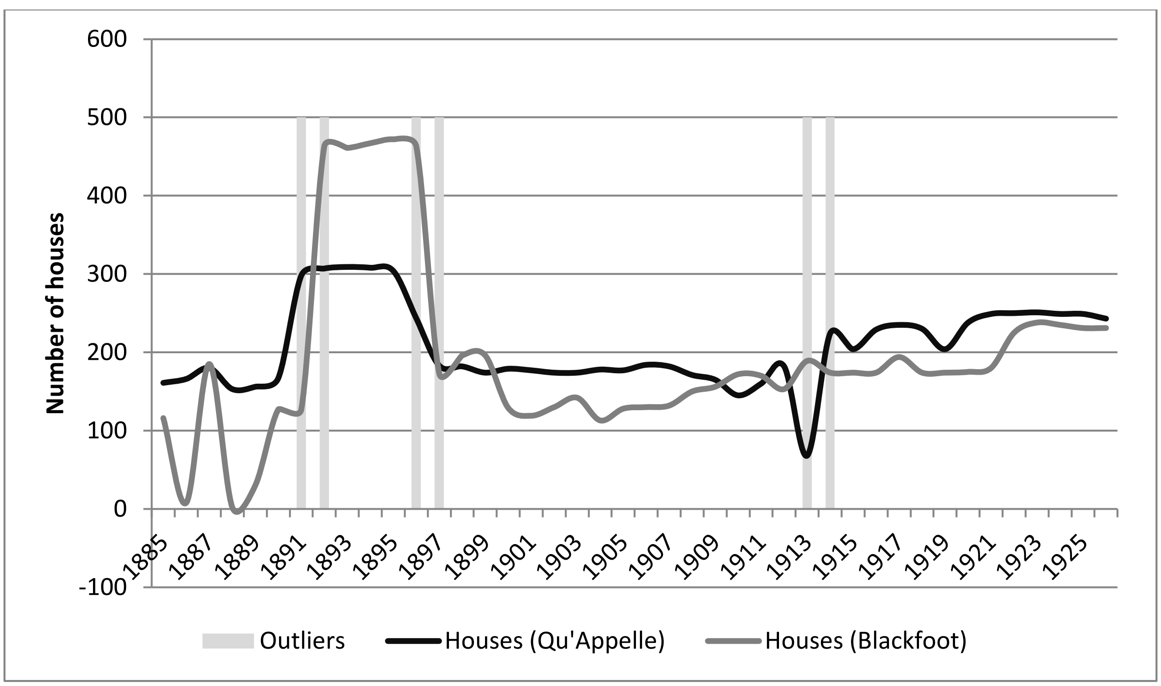

Figure 7-46: Number of houses, Qu'Appelle and Blackfoot agencies. Source: DIA(1885-1926)

Figure 7-47: Number of houses in the Qu'Appelle Agency, with and without outliers. Source: DIA(1885-1926)

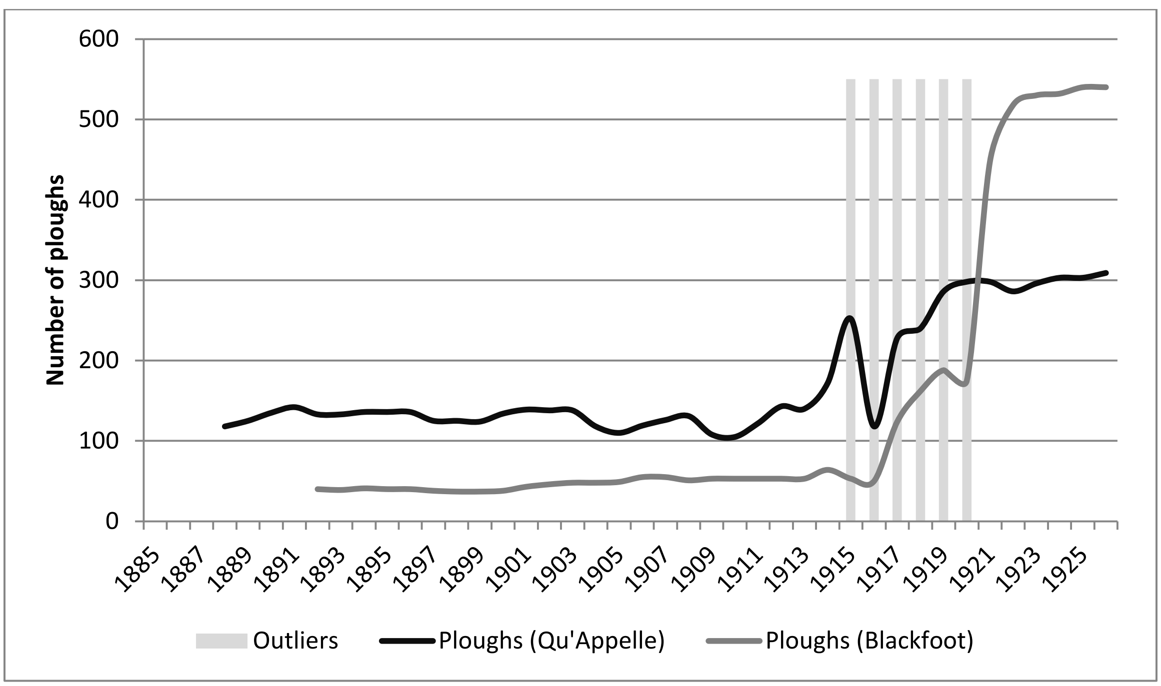

Figure 7-48: Number of ploughs, Qu'Appelle and Blackfoot Agencies. Source: DIA(1885-1926)

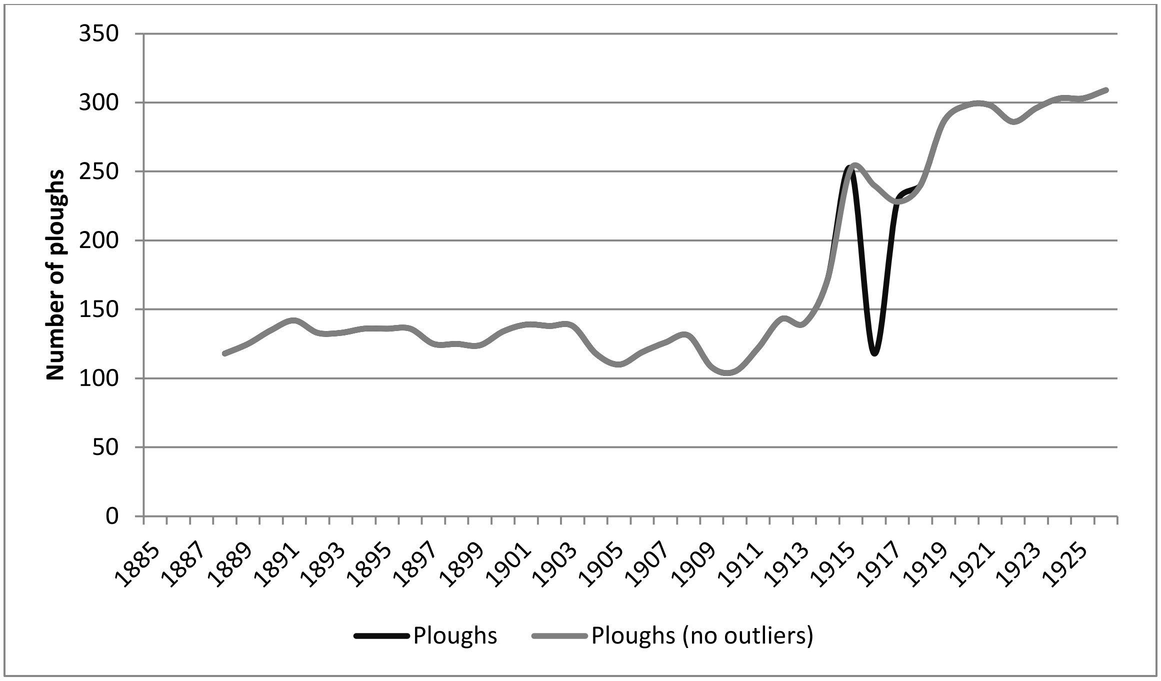

Figure 7-49: Number of ploughs with and without outliers, Qu'Appelle Agency. Source: DIA(1885-1926)

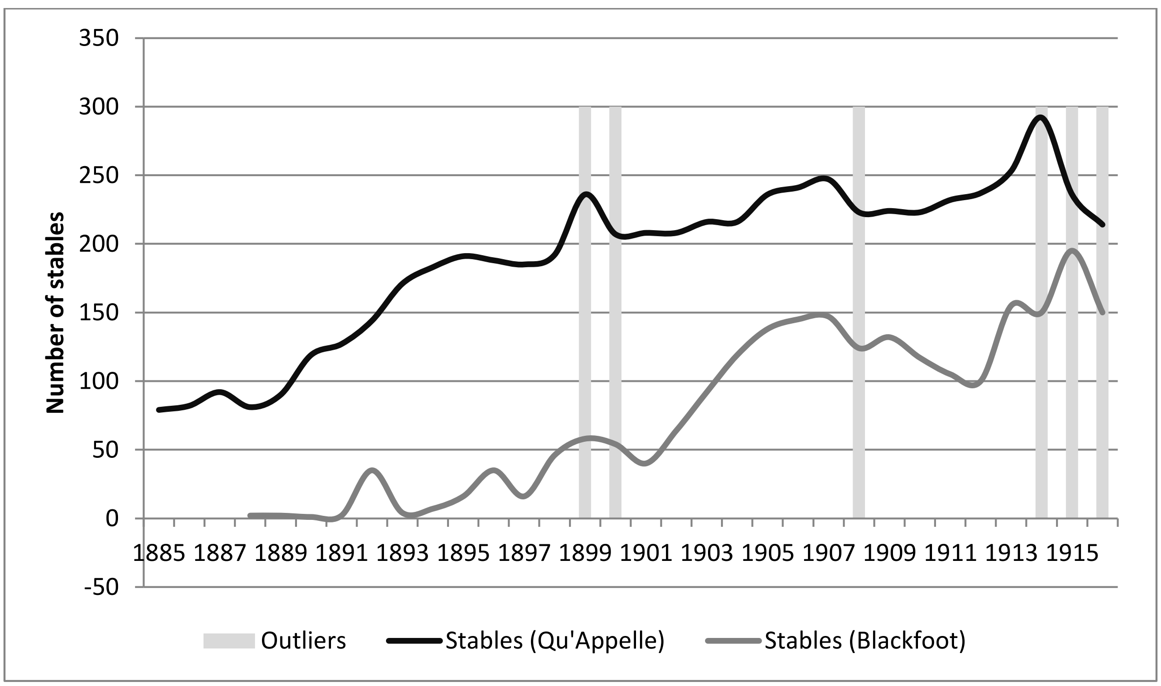

Figure 7-50: The number of stables, Qu'Appelle and Blackfoot Agencies. Source: DIA(1885-1926)

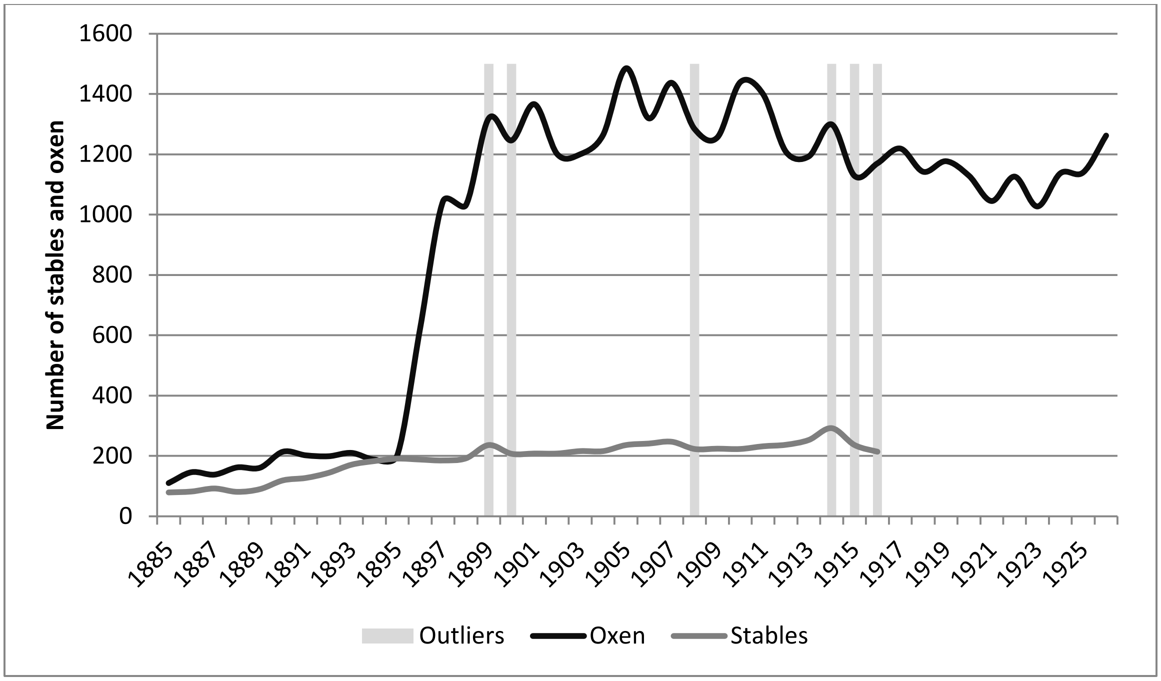

Figure 7-51: Number of stables and oxen, Qu'Appelle Agency. Source: DIA(1885-1926)

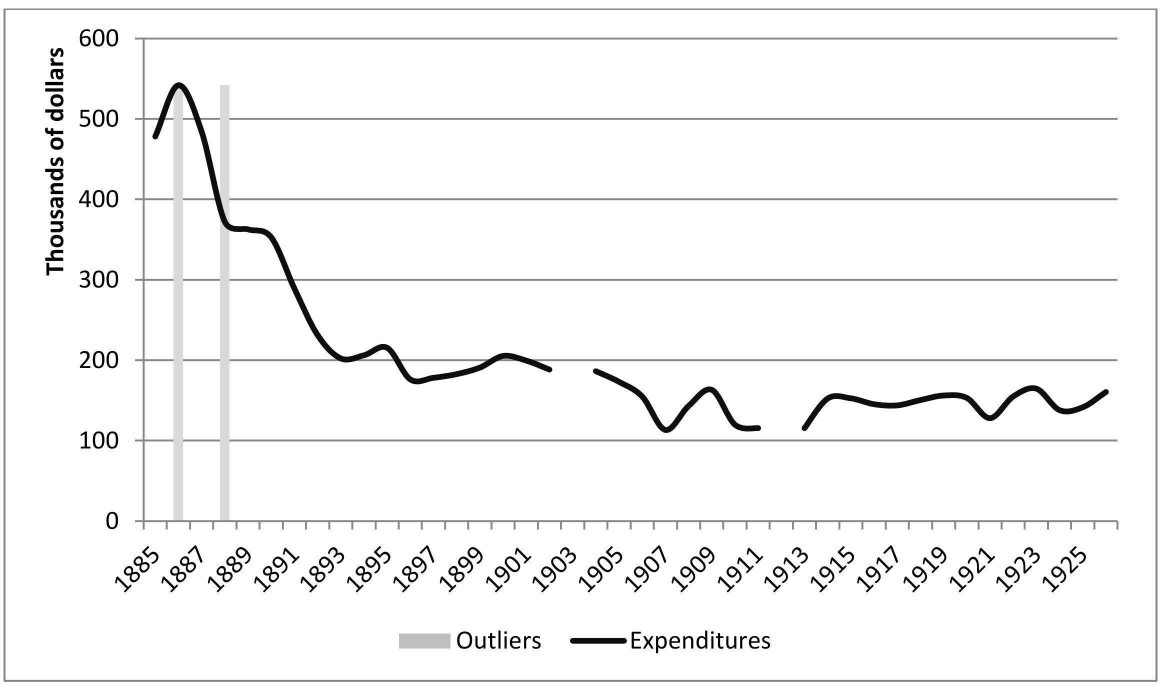

Figure 7-52: Expenditures, Manitoba. Source: DIA(1885-1926)

7.1.3. Multivariate regression results

Table 7-29: Model selection, multivariate regression of TB mortality on variables listed in "Variables in Model" column

Model Index	Number in Model	Adjusted R-Squared	R-Square	AIC	BIC	Variables in Model
1	4	0.2883	0.3613	170.0528	174.1469	Dummy L.Oxen Infect Exp1800
2	5	0.276	0.3688	171.5817	176.2741	L.Potato Dummy L.Oxen Infect Exp1800
3	5	0.2755	0.3684	171.6098	176.2926	Dummy L.Oxen Infect Flu Exp1800
4	5	0.2676	0.3615	172.0399	176.5764	Expenditures Dummy L.Oxen Infect Exp1900
5	5	0.2676	0.3615	172.0399	176.5764	Dummy L.Oxen Infect Exp1900 Exp1800
6	5	0.2676	0.3615	172.0399	176.5764	Expenditures Dummy L.Oxen Infect Exp1800
7	6	0.2619	0.3755	173.1565	178.4877	L.Potato Dummy L.Oxen Infect Flu Exp1800
8	4	0.2611	0.3369	171.554	175.2344	Expenditures Dummy L.Oxen Infect
9	4	0.2589	0.3349	171.6752	175.3226	L.Potato Dummy L.Oxen Infect
10	6	0.2551	0.3697	173.5224	178.7056	Expenditures L.Potato Dummy L.Oxen Infect Exp1900
11	6	0.2551	0.3697	173.5224	178.7056	L.Potato Dummy L.Oxen Infect Exp1900 Exp1800
12	6	0.2551	0.3697	173.5224	178.7056	Expenditures L.Potato Dummy L.Oxen Infect Exp1800
13	6	0.2543	0.369	173.5697	178.7339	Expenditures Dummy L.Oxen Infect Exp1900 Flu
14	6	0.2543	0.369	173.5697	178.7339	Dummy L.Oxen Infect Exp1900 Flu Exp1800
15	6	0.2543	0.369	173.5697	178.7339	Expenditures Dummy L.Oxen Infect Flu Exp1800
16	3	0.2535	0.3109	171.0925	174.0972	L.Oxen Infect Exp1800
17	5	0.2522	0.3481	172.8736	177.1292	Expenditures L.Potato Dummy L.Oxen Infect
18	3	0.2513	0.3089	171.2105	174.1903	Dummy L.Oxen Infect
19	5	0.2453	0.3421	173.24	177.3731	Expenditures Dummy L.Oxen Infect Flu
20	5	0.2449	0.3417	173.2653	177.3899	L.Potato Dummy L.Oxen Infect Flu
21	7	0.2409	0.3772	175.0486	180.9236	Expenditures L.Potato Dummy L.Oxen Infect Exp1900 Flu

22	7	0.2409	0.3772	175.0486	180.9236	L.Potato Dummy L.Oxen Infect Exp1900 Flu Exp1800
23	7	0.2409	0.3772	175.0486	180.9236	Expenditures L.Potato Dummy L.Oxen Infect Flu Exp1800
24	5	0.2395	0.337	173.5498	177.5799	L.Potato Dummy L.Oxen Infect Exp1900
25	4	0.2391	0.3171	172.7296	176.0914	L.Oxen Infect Flu Exp1800
26	4	0.2383	0.3164	172.7708	176.1215	Dummy L.Oxen Infect Flu
27	6	0.2358	0.3534	174.5482	179.321	Expenditures L.Potato Dummy L.Oxen Infect Flu
28	4	0.2355	0.3139	172.9161	176.2278	L.Potato L.Oxen Infect Exp1800
29	4	0.2333	0.3119	173.0323	176.3128	Expenditures L.Oxen Infect Exp1900
30	4	0.2333	0.3119	173.0323	176.3128	Expenditures L.Oxen Infect Exp1800
31	4	0.2333	0.3119	173.0323	176.3128	L.Oxen Infect Exp1900 Exp1800
32	4	0.2303	0.3092	173.1906	176.4287	Dummy L.Oxen Infect Exp1900
33	6	0.2257	0.3448	175.0729	179.6384	L.Potato Dummy L.Oxen Infect Exp1900 Flu
34	3	0.2211	0.281	172.792	175.4407	Expenditures L.Oxen Infect
35	5	0.2199	0.3199	174.5657	178.2612	L.Potato L.Oxen Infect Flu Exp1800
36	5	0.2173	0.3176	174.6988	178.3508	Expenditures L.Oxen Infect Exp1900 Flu
37	5	0.2173	0.3176	174.6988	178.3508	L.Oxen Infect Exp1900 Flu Exp1800
38	5	0.2173	0.3176	174.6988	178.3508	Expenditures L.Oxen Infect Flu Exp1800
39	5	0.2169	0.3173	174.7189	178.3644	Dummy L.Oxen Infect Exp1900 Flu
40	5	0.2137	0.3145	174.8817	178.4742	Expenditures L.Potato L.Oxen Infect Exp1900
41	5	0.2137	0.3145	174.8817	178.4742	L.Potato L.Oxen Infect Exp1900 Exp1800
42	5	0.2137	0.3145	174.8817	178.4742	Expenditures L.Potato L.Oxen Infect Exp1800
43	4	0.2036	0.2853	174.5528	177.4299	Expenditures L.Potato L.Oxen Infect
44	4	0.2027	0.2844	174.5991	177.464	Expenditures L.Oxen Infect Flu
45	6	0.1965	0.3201	176.5518	180.5422	Expenditures L.Potato L.Oxen Infect Exp1900 Flu
46	6	0.1965	0.3201	176.5518	180.5422	L.Potato L.Oxen Infect Exp1900 Flu Exp1800
47	6	0.1965	0.3201	176.5518	180.5422	Expenditures L.Potato L.Oxen

						Infect Flu Exp1800
48	5	0.1841	0.2887	176.36	179.4769	Expenditures L.Potato L.Oxen Infect Flu
49	4	0.1776	0.2619	175.838	178.3805	L.Potato Dummy L.Oxen Exp1800
50	3	0.1748	0.2383	175.0978	177.2741	Dummy L.Oxen Exp1800
51	4	0.1742	0.2589	176	178.5009	Dummy L.Oxen Flu Exp1800
52	3	0.1731	0.2367	175.1806	177.3401	L.Potato Dummy L.Oxen
53	5	0.1725	0.2786	176.9235	179.862	L.Potato Dummy L.Oxen Flu Exp1800
54	3	0.1661	0.2303	175.5189	177.6102	Infect Flu Exp1800
55	4	0.1658	0.2514	176.4056	178.8024	L.Potato Dummy L.Oxen Flu
56	4	0.1642	0.2499	176.4834	178.8602	Dummy Infect Flu Exp1800
57	5	0.1631	0.2704	177.3788	180.1743	L.Potato Dummy Infect Flu Exp1800
58	2	0.1629	0.2058	174.7698	176.6509	Infect Exp1800
59	4	0.1617	0.2477	176.6022	178.9488	Expenditures L.Potato Dummy L.Oxen
60	4	0.1591	0.2453	176.729	179.0432	L.Potato Dummy Infect Exp1800
61	3	0.1571	0.222	175.9475	177.9529	L.Potato Infect Exp1800
62	4	0.1569	0.2434	176.8305	179.1189	L.Potato Infect Flu Exp1800
63	3	0.1568	0.2216	175.965	177.9669	Dummy Infect Exp1800
64	5	0.1567	0.2648	177.6818	180.3827	Expenditures L.Potato Dummy L.Oxen Flu
65	4	0.1559	0.2425	176.8775	179.1539	Expenditures Dummy L.Oxen Flu
66	3	0.1554	0.2204	176.0283	178.0175	Expenditures Dummy L.Oxen
67	3	0.1554	0.2204	176.0297	178.0187	L.Potato L.Oxen Infect
68	5	0.1538	0.2623	177.8205	180.4782	Expenditures L.Potato Dummy L.Oxen Exp1900
69	5	0.1538	0.2623	177.8205	180.4782	L.Potato Dummy L.Oxen Exp1900 Exp1800
70	5	0.1538	0.2623	177.8205	180.4782	Expenditures L.Potato Dummy L.Oxen Exp1800
71	4	0.1514	0.2385	177.0902	179.3127	Expenditures Dummy L.Oxen Exp1900
72	4	0.1514	0.2385	177.0902	179.3127	Dummy L.Oxen Exp1900 Exp1800
73	4	0.1514	0.2385	177.0902	179.3127	Expenditures Dummy L.Oxen Exp1800
74	2	0.1512	0.1948	175.3225	177.1215	L.Oxen Exp1800
75	4	0.1506	0.2377	177.1287	179.3415	L.Potato Dummy L.Oxen Exp1900
76	5	0.1504	0.2593	177.9783	180.5871	Expenditures Dummy L.Oxen Exp1900 Flu
77	5	0.1504	0.2593	177.9783	180.5871	Expenditures Dummy L.Oxen Flu Exp1800

78	5	0.1504	0.2593	177.9783	180.5871	Dummy L.Oxen Exp1900 Flu Exp1800
79	2	0.1499	0.1935	175.3838	177.1737	L.Oxen Infect
80	3	0.1494	0.2148	176.3148	178.2469	L.Oxen Flu Exp1800
81	6	0.1475	0.2787	178.9199	182.0182	Expenditures L.Potato Dummy L.Oxen Exp1900 Flu
82	6	0.1475	0.2787	178.9199	182.0182	L.Potato Dummy L.Oxen Exp1900 Flu Exp1800
83	6	0.1475	0.2787	178.9199	182.0182	Expenditures L.Potato Dummy L.Oxen Flu Exp1800
84	3	0.1439	0.2097	176.572	178.4529	L.Potato L.Oxen Exp1800
85	2	0.1438	0.1877	175.6735	177.4205	Dummy L.Oxen
86	4	0.1436	0.2314	177.4584	179.5879	Expenditures Infect Exp1900 Flu
87	4	0.1436	0.2314	177.4584	179.5879	Infect Exp1900 Flu Exp1800
88	4	0.1436	0.2314	177.4584	179.5879	Expenditures Infect Flu Exp1800
89	3	0.1428	0.2088	176.6202	178.4915	Expenditures Infect Exp1900
90	3	0.1428	0.2088	176.6202	178.4915	Infect Exp1900 Exp1800
91	3	0.1428	0.2088	176.6202	178.4915	Expenditures Infect Exp1800
92	5	0.1419	0.2519	178.3778	180.8631	L.Potato Dummy L.Oxen Exp1900 Flu
93	3	0.1411	0.2072	176.7018	178.557	Dummy L.Oxen Flu
94	5	0.1397	0.25	178.4806	180.9343	Expenditures Dummy Infect Exp1900 Flu
95	5	0.1397	0.25	178.4806	180.9343	Dummy Infect Exp1900 Flu Exp1800
96	5	0.1397	0.25	178.4806	180.9343	Expenditures Dummy Infect Flu Exp1800
97	4	0.1383	0.2267	177.7056	179.7729	L.Potato L.Oxen Flu Exp1800
98	6	0.1382	0.2708	179.3552	182.2934	Expenditures L.Potato Dummy Infect Exp1900 Flu
99	6	0.1382	0.2708	179.3552	182.2934	L.Potato Dummy Infect Exp1900 Flu Exp1800
100	6	0.1382	0.2708	179.3552	182.2934	Expenditures L.Potato Dummy Infect Flu Exp1800
101	4	0.1377	0.2261	177.7319	179.7926	L.Potato L.Oxen Infect Flu
102	4	0.1371	0.2256	177.7618	179.815	L.Potato Dummy Infect Flu
103	4	0.136	0.2246	177.8115	179.8523	Expenditures L.Potato Dummy Infect
104	2	0.1355	0.1798	176.0575	177.7479	Expenditures Infect
105	5	0.1351	0.246	178.694	181.0822	Expenditures L.Potato Dummy Infect Flu
106	4	0.1344	0.2232	177.886	179.9081	Expenditures L.Potato Infect Exp1900

107	4	0.1344	0.2232	177.886	179.9081	L.Potato Infect Exp1900 Exp1800
108	4	0.1344	0.2232	177.886	179.9081	Expenditures L.Potato Infect Exp1800
109	5	0.1343	0.2453	178.7288	181.1063	Expenditures L.Potato Dummy Infect Exp1900
110	5	0.1343	0.2453	178.7288	181.1063	L.Potato Dummy Infect Exp1900 Exp1800
111	5	0.1343	0.2453	178.7288	181.1063	Expenditures L.Potato Dummy Infect Exp1800
112	4	0.1339	0.2227	177.9104	179.9264	Expenditures Dummy Infect Exp1900
113	4	0.1339	0.2227	177.9104	179.9264	Expenditures Dummy Infect Exp1800
114	4	0.1339	0.2227	177.9104	179.9264	Dummy Infect Exp1900 Exp1800
115	3	0.1334	0.2	177.0597	178.8441	L.Oxen Infect Flu
116	5	0.1325	0.2437	178.8152	181.1663	Expenditures L.Potato Infect Exp1900 Flu
117	5	0.1325	0.2437	178.8152	181.1663	Expenditures L.Potato Infect Flu Exp1800
118	5	0.1325	0.2437	178.8152	181.1663	L.Potato Infect Exp1900 Flu Exp1800
119	3	0.1322	0.1989	177.1137	178.8875	L.Potato Dummy Infect
120	4	0.1322	0.2212	177.9871	179.9839	L.Potato L.Oxen Infect Exp1900
121	4	0.1308	0.22	178.049	180.0304	Expenditures Dummy Infect Flu
122	3	0.1308	0.1977	177.1768	178.9382	Expenditures Infect Flu
123	3	0.1308	0.1977	177.1776	178.9388	Expenditures L.Oxen Exp1900
124	3	0.1308	0.1977	177.1776	178.9388	L.Oxen Exp1900 Exp1800
125	3	0.1308	0.1977	177.1776	178.9388	Expenditures L.Oxen Exp1800
126	3	0.1306	0.1975	177.1866	178.946	Expenditures L.Potato Infect
127	3	0.1303	0.1972	177.2004	178.9571	Expenditures Dummy Infect
128	3	0.1299	0.1968	177.2209	178.9735	Expenditures L.Oxen Flu
129	2	0.1298	0.1744	176.3208	177.9725	Expenditures L.Oxen
130	3	0.1296	0.1965	177.2339	178.984	L.Oxen Infect Exp1900
131	4	0.1293	0.2186	178.1181	180.0822	Expenditures L.Oxen Exp1900 Flu
132	4	0.1293	0.2186	178.1181	180.0822	L.Oxen Exp1900 Flu Exp1800
133	4	0.1293	0.2186	178.1181	180.0822	Expenditures L.Oxen Flu Exp1800
134	4	0.1236	0.2135	178.3804	180.2792	Expenditures L.Potato Infect Flu
135	3	0.1227	0.1902	177.5496	179.2377	Expenditures L.Potato L.Oxen
136	4	0.1208	0.211	178.5097	180.3764	Expenditures L.Potato L.Oxen Exp1900
137	4	0.1208	0.211	178.5097	180.3764	L.Potato L.Oxen Exp1900 Exp1800
138	4	0.1208	0.211	178.5097	180.3764	Expenditures L.Potato L.Oxen Exp1800

139	3	0.12	0.1877	177.6706	179.335	Dummy L.Oxen Exp1900
140	4	0.1184	0.2088	178.6175	180.4575	Expenditures L.Potato L.Oxen Flu
141	4	0.1168	0.2074	178.6898	180.5119	Dummy L.Oxen Exp1900 Flu
142	5	0.1152	0.2287	179.6018	181.7139	Expenditures L.Potato L.Oxen Exp1900 Flu
143	5	0.1152	0.2287	179.6018	181.7139	L.Potato L.Oxen Exp1900 Flu Exp1800
144	5	0.1152	0.2287	179.6018	181.7139	Expenditures L.Potato L.Oxen Flu Exp1800
145	5	0.1133	0.227	179.6901	181.7756	L.Potato Dummy Infect Exp1900 Flu
146	5	0.1128	0.2266	179.711	181.7902	L.Potato L.Oxen Infect Exp1900 Flu
147	3	0.1111	0.1794	178.0761	179.6614	L.Potato Dummy Exp1800
148	4	0.1109	0.2021	178.9555	180.712	L.Oxen Infect Exp1900 Flu
149	4	0.1076	0.1992	179.1036	180.8236	L.Potato Dummy Infect Exp1900
150	2	0.1072	0.153	177.3438	178.846	L.Potato Exp1800
151	1	0.1007	0.1238	176.7015	178.2308	Exp1800
152	3	0.0993	0.1685	178.6039	180.0868	Dummy Infect Flu
153	2	0.0969	0.1432	177.8053	179.2406	L.Potato Dummy
154	3	0.0946	0.1643	178.808	180.2515	Expenditures L.Potato Dummy
155	2	0.0936	0.1401	177.9499	179.3643	Dummy Exp1800
156	2	0.0884	0.1352	178.1788	179.5603	Expenditures L.Potato
157	2	0.0882	0.135	178.1873	179.5675	Dummy Infect
158	4	0.087	0.1807	180.0163	181.5137	L.Potato Dummy Flu Exp1800
159	2	0.0864	0.1333	178.2667	179.6355	L.Potato L.Oxen
160	4	0.0857	0.1795	180.074	181.5574	Expenditures L.Potato Dummy Exp1900
161	4	0.0857	0.1795	180.074	181.5574	L.Potato Dummy Exp1900 Exp1800
162	4	0.0857	0.1795	180.074	181.5574	Expenditures L.Potato Dummy Exp1800
163	3	0.085	0.1553	179.2338	180.5955	L.Potato Flu Exp1800
164	3	0.0843	0.1547	179.2621	180.6184	Expenditures L.Potato Exp1900
165	3	0.0843	0.1547	179.2621	180.6184	L.Potato Exp1900 Exp1800
166	3	0.0843	0.1547	179.2621	180.6184	Expenditures L.Potato Exp1800
167	2	0.0842	0.1312	178.3635	179.7184	L.Potato Infect
168	1	0.0822	0.1057	177.517	178.9659	Expenditures
169	2	0.0815	0.1286	178.4812	179.8192	Expenditures Exp1900
170	2	0.0815	0.1286	178.4812	179.8192	Exp1900 Exp1800
171	2	0.0815	0.1286	178.4812	179.8192	Expenditures Exp1800
172	3	0.0812	0.1519	179.3965	180.7271	L.Potato Infect Flu
173	2	0.0795	0.1267	178.568	179.8935	Flu Exp1800

174	3	0.0754	0.1465	179.649	180.9314	L.Potato L.Oxen Flu
175	2	0.0751	0.1225	178.7588	180.0569	Expenditures Dummy
176	4	0.0736	0.1686	180.6024	181.9584	Dummy Infect Exp1900 Flu
177	3	0.0722	0.1436	179.7875	181.0436	L.Potato Dummy Flu
178	3	0.0719	0.1433	179.8023	181.0556	L.Potato Dummy Exp1900
179	3	0.0709	0.1423	179.8454	181.0905	Expenditures Dummy Exp1800
180	3	0.0709	0.1423	179.8454	181.0905	Expenditures Dummy Exp1900
181	3	0.0709	0.1423	179.8454	181.0905	Dummy Exp1900 Exp1800
182	3	0.0706	0.1421	179.8569	181.0998	Dummy Flu Exp1800
183	4	0.0704	0.1657	180.7383	182.0618	Expenditures L.Potato Dummy Flu
184	3	0.0664	0.1383	180.0353	181.2443	Expenditures L.Potato Flu
185	3	0.0641	0.1361	180.1361	181.3261	Dummy Infect Exp1900
186	3	0.0626	0.1347	180.199	181.377	L.Potato L.Oxen Exp1900
187	2	0.0617	0.1098	179.3336	180.5498	Expenditures Flu
188	4	0.0611	0.1574	181.1357	182.3643	Expenditures L.Potato Exp1900 Flu
189	4	0.0611	0.1574	181.1357	182.3643	L.Potato Exp1900 Flu Exp1800
190	4	0.0611	0.1574	181.1357	182.3643	Expenditures L.Potato Flu Exp1800
191	1	0.0604	0.0845	178.4543	179.8114	L.Oxen
192	3	0.0604	0.1327	180.2938	181.454	L.Potato Infect Exp1900
193	5	0.0603	0.1808	182.0113	183.4106	Expenditures L.Potato Dummy Exp1900 Flu
194	5	0.0603	0.1808	182.0113	183.4106	L.Potato Dummy Exp1900 Flu Exp1800
195	5	0.0603	0.1808	182.0113	183.4106	Expenditures L.Potato Dummy Flu Exp1800
196	3	0.0598	0.1321	180.3204	181.4755	Expenditures Flu Exp1800
197	3	0.0598	0.1321	180.3204	181.4755	Expenditures Exp1900 Flu
198	3	0.0598	0.1321	180.3204	181.4755	Exp1900 Flu Exp1800
199	4	0.0555	0.1524	181.3744	182.5462	L.Potato Infect Exp1900 Flu
200	2	0.0539	0.1024	179.6642	180.8335	L.Oxen Flu
201	3	0.0524	0.1253	180.6334	181.7296	Expenditures Dummy Flu
202	2	0.0515	0.1001	179.7677	180.9223	Infect Flu
203	4	0.0515	0.1487	181.5456	182.6769	L.Potato L.Oxen Exp1900 Flu
204	1	0.0514	0.0758	178.8358	180.1556	L.Potato
205	1	0.0488	0.0732	178.945	180.2541	Infect
206	4	0.047	0.1447	181.7335	182.8204	Expenditures Dummy Exp1900 Flu
207	4	0.047	0.1447	181.7335	182.8204	Expenditures Dummy Flu Exp1800
208	4	0.047	0.1447	181.7335	182.8204	Dummy Exp1900 Flu Exp1800
209	4	0.0457	0.1436	181.7857	182.8603	L.Potato Dummy Exp1900 Flu
210	2	0.041	0.0902	180.2061	181.2989	L.Oxen Exp1900

211	1	0.0357	0.0605	179.4927	180.7486	Dummy
212	3	0.0351	0.1094	181.3548	182.3163	L.Oxen Exp1900 Flu
213	2	0.0299	0.0797	180.6666	181.6948	Infect Exp1900
214	3	0.0289	0.1036	181.6135	182.527	Infect Exp1900 Flu
215	2	0.0279	0.0778	180.7486	181.7653	L.Potato Exp1900
216	2	0.027	0.0769	180.7869	181.7982	L.Potato Flu
217	2	0.012	0.0627	181.3971	182.3235	Dummy Exp1900
218	2	0.0104	0.0611	181.465	182.382	Dummy Flu
219	3	0.0024	0.0792	182.6884	183.4044	L.Potato Exp1900 Flu
220	3	-0.0145	0.0636	183.3594	183.9539	Dummy Exp1900 Flu
221	1	-0.0174	0.0087	181.6369	182.6864	Exp1900
222	1	-0.0248	0.0015	181.9274	182.9492	Flu
223	2	-0.0427	0.0108	183.5527	184.1843	Exp1900 Flu

Table 7-30: Collinearity diagnostics, Model 1 (Table 7-29)

Number	Eigenvalue	Condition	Proportion of Variation			
		Index	Intercept	L.Oxen	Infect	Exp1800
1	1.6396	1	0.17582	0.03114	0.17688	0.09668
2	0.99118	1.28615	0.06802	0.86583	0.000884	0.01429
3	0.87948	1.36538	0.03592	0.00453	0.13888	0.84013
4	0.48973	1.82974	0.72024	0.09849	0.68336	0.0489

Table 7-31: Parameter estimates, Model 1 (Table 7-29)

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Tolerance	95% Confidence Limits	
Intercept	1	-7.04394	2.60838	-2.7	0.0106	0	.	-12.3392	-1.74866
Dummy	1	4.94951	2.97741	1.66	0.1054	0.25057	0.80319	-1.09496	10.99397
L.Oxen	1	0.02691	0.00973	2.77	0.009	0.3967	0.8877	0.00717	0.04665
Infect	1	7.91942	3.05031	2.6	0.0137	0.35744	0.96274	1.72696	14.11188
Exp1800	1	0.00012	7.06E-05	1.7	0.0988	0.24914	0.8452	-2.4E-05	0.000263

D. Alternative scenario formulations

7.1.4. Oats harvest

Alternative definition of oats harvest produces a significantly improved model fit. This definition is based on the following formula :

$$Oats = \propto_O + \beta_{O1}Wheat + \beta_{O2}GDD + \beta_{O3}Hay + \beta_{O4}Precip + \beta_{O5}Area + \beta_{O6}Ploughs + \varepsilon$$

Unlike the definition used to formulate base case scenario, alternative definition includes growing degree days (GDD), area farmed (Area), and the number of ploughs. Area farmed in particular can improve model fit in years 1885-1899. Area farmed, however, is not a statistically significant predictor of oats harvest when actual oats harvest data are made stationary by fitting an ARIMA model as was done here. This is an example of how the choice of alternative statistical methods may have unexpected impacts later on.

7.1.5. Number of oxen

Alternative definition of the number of oxen produces a significantly improved model fit in years 1900-1926. Unlike the definition used in the base case scenario, alternative definition includes lagged number of oxen, precipitation, hay, and the number of horses. Like the oats harvest, these variables lost significance when ARIMA model was fit to the actual number of oxen.

$$Oxen = \propto_{Ox} + \beta_{Ox1}OxenL1 + \beta_{Ox2}Hay + \beta_{Ox3}Horses + \beta_{Ox4}Precip + \beta_{Ox5}Stables + \varepsilon$$

E. Sensitivity analyses

Tables below present partial rank correlation coefficient p-values for correlations between a selected number of variables and model outputs. Each table corresponds to a model output, with variables listed in the top row of each table. Columns represent time periods: years from 1885 – 1926. P-values significant at 10% are highlighted in bold.

Table 7-32: PRCC p-values for correlations between selected variables and the number of active TB cases

	pInf	pPrg	Suscept.	prevL	prevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.04	-0.13	-0.17	0.04	-0.19	-0.07	-0.01	-0.23	0.02
1886	-0.06	-0.05	-0.26	0.15	-0.17	0.16	0.28	-0.27	0.05
1887	0.04	-0.06	0.00	-0.05	-0.02	-0.11	0.03	-0.06	0.04
1888	-0.13	-0.01	-0.26	0.11	-0.28	-0.09	0.24	**-0.29**	0.08
1889	-0.04	-0.06	-0.26	0.17	-0.18	-0.14	0.10	-0.18	0.02
1890	-0.16	0.03	**-0.33**	0.23	-0.28	-0.06	**0.31**	**-0.34**	0.12
1891	0.02	-0.08	-0.08	0.06	-0.12	-0.07	0.06	-0.15	0.05
1892	0.00	-0.12	-0.02	0.02	-0.16	-0.21	0.10	-0.13	-0.07
1893	-0.19	-0.06	**-0.32**	0.19	-0.27	0.00	0.16	**-0.34**	0.15
1894	-0.12	-0.10	**-0.30**	0.08	-0.28	-0.04	0.22	-0.26	0.07
1895	-0.21	-0.11	-0.19	0.15	**-0.38**	-0.05	**0.30**	-0.21	0.17
1896	-0.07	0.10	-0.25	-0.03	-0.14	0.09	0.24	-0.24	0.11
1897	-0.13	-0.08	-0.27	0.12	-0.24	0.06	0.12	-0.26	0.11
1898	0.11	-0.11	-0.04	0.03	-0.22	-0.09	0.04	-0.14	0.11
1899	-0.06	0.03	**-0.30**	0.03	-0.24	-0.08	0.16	**-0.30**	0.00
1900	-0.10	-0.01	-0.25	0.14	-0.23	-0.01	0.18	-0.26	0.12
1901	-0.09	0.05	**-0.31**	0.16	-0.22	-0.02	0.20	-0.21	0.04
1902	-0.14	0.06	-0.29	0.14	-0.28	-0.06	0.24	-0.14	0.18
1903	0.07	-0.13	-0.14	0.02	-0.16	-0.15	0.00	-0.16	0.02
1904	-0.12	0.00	-0.12	0.06	-0.16	-0.01	0.24	-0.16	0.05

1905	0.03	0.00	-0.24	-0.02	-0.22	0.01	0.19	**-0.33**	0.11
1906	-0.08	0.03	-0.21	0.13	-0.23	-0.10	0.19	-0.14	0.11
1907	-0.06	-0.03	-0.26	0.02	-0.27	0.00	0.18	-0.28	0.06
1908	-0.14	0.13	0.05	0.08	-0.04	-0.17	**0.36**	0.16	0.06
1909	0.00	0.03	-0.27	0.04	-0.17	-0.06	0.18	-0.22	0.01
1910	-0.05	0.08	-0.25	0.07	-0.16	0.00	0.25	-0.15	0.06
1911	-0.08	-0.05	-0.27	0.05	-0.22	-0.02	0.16	**-0.31**	0.00
1912	0.08	-0.13	-0.10	0.01	-0.24	-0.10	0.03	-0.13	0.02
1913	-0.09	0.03	-0.23	0.20	-0.18	-0.08	0.26	-0.16	0.12
1914	-0.06	-0.18	-0.08	0.06	-0.22	-0.19	0.17	-0.14	0.04
1915	-0.12	0.02	**-0.30**	0.15	-0.23	-0.04	0.25	-0.24	0.11
1916	-0.10	-0.07	-0.13	0.03	-0.20	-0.05	0.10	-0.25	0.08
1917	-0.21	-0.04	-0.27	0.05	-0.19	-0.17	0.11	-0.18	-0.05
1918	-0.05	0.06	**-0.31**	0.09	-0.19	-0.04	0.25	-0.24	0.05
1919	0.03	0.02	-0.26	0.07	-0.20	-0.13	-0.01	-0.25	-0.01
1920	-0.04	-0.01	-0.15	0.07	-0.18	-0.07	0.22	-0.22	0.03
1921	-0.23	0.02	-0.15	0.01	-0.20	0.00	0.25	**-0.29**	0.06
1922	-0.04	-0.07	-0.20	0.12	-0.28	-0.08	0.11	-0.28	0.01
1923	-0.08	-0.11	**-0.34**	0.10	-0.25	-0.09	0.18	**-0.31**	-0.02
1924	-0.08	-0.03	**-0.30**	0.08	-0.25	-0.04	0.16	**-0.30**	0.07
1925	-0.10	-0.03	-0.25	0.13	-0.25	-0.11	0.19	-0.24	0.10
1926	-0.20	-0.04	**-0.31**	0.19	**-0.33**	-0.06	0.24	**-0.31**	0.09

Table 7-33: PRCC p-values for correlations between selected variables and the number of latent TB cases

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.10	-0.07	-0.13	-0.02	-0.14	-0.10	-0.03	-0.22	-0.02

1886	0.11	-0.07	-0.13	-0.02	-0.13	-0.10	-0.05	-0.21	0.00
1887	0.07	-0.09	-0.13	-0.01	-0.19	-0.11	-0.01	-0.22	-0.02
1888	0.11	-0.07	-0.12	-0.03	-0.13	-0.09	-0.06	-0.21	0.00
1889	0.10	-0.07	-0.12	-0.03	-0.14	-0.10	-0.05	-0.21	0.00
1890	0.11	-0.08	-0.13	-0.02	-0.13	-0.10	-0.04	-0.21	0.00
1891	0.11	-0.07	-0.13	-0.02	-0.13	-0.10	-0.04	-0.23	0.00
1892	0.12	-0.07	-0.10	0.02	-0.17	-0.06	-0.02	-0.25	0.08
1893	0.10	-0.07	-0.12	-0.01	-0.15	-0.10	-0.04	-0.19	0.01
1894	0.10	-0.06	-0.12	-0.01	-0.13	-0.10	-0.04	-0.22	0.00
1895	0.13	-0.03	-0.15	-0.03	-0.12	-0.11	-0.06	-0.21	0.01
1896	0.10	-0.07	-0.13	-0.01	-0.14	-0.09	-0.03	-0.20	0.02
1897	0.11	-0.08	-0.12	-0.02	-0.14	-0.10	-0.05	-0.22	0.00
1898	0.10	-0.06	-0.14	-0.02	-0.13	-0.10	-0.06	-0.22	0.00
1899	0.10	-0.08	-0.14	0.00	-0.15	-0.10	-0.02	-0.23	0.01
1900	0.11	-0.08	-0.13	0.00	-0.16	-0.09	-0.03	-0.23	0.02
1901	0.11	-0.06	-0.14	-0.02	-0.14	-0.09	-0.04	-0.22	0.01
1902	0.11	-0.07	-0.12	-0.01	-0.13	-0.09	-0.05	-0.23	0.00
1903	0.10	-0.06	-0.11	-0.01	-0.16	-0.09	-0.02	-0.24	0.00
1904	0.10	-0.07	-0.13	-0.03	-0.14	-0.10	-0.05	-0.21	0.01
1905	0.12	-0.05	-0.12	-0.04	-0.10	-0.10	-0.06	-0.21	0.00
1906	0.11	-0.07	-0.14	-0.02	-0.14	-0.11	-0.05	-0.22	-0.01
1907	0.10	-0.03	-0.11	-0.02	-0.10	-0.10	-0.07	-0.22	0.00
1908	0.03	-0.06	-0.20	0.04	-0.21	0.10	0.10	**-0.42**	0.10
1909	0.10	-0.07	-0.14	0.00	-0.15	-0.10	-0.03	-0.22	0.01
1910	0.10	-0.08	-0.12	0.00	-0.15	-0.10	-0.03	-0.22	0.02
1911	0.10	-0.06	-0.11	-0.01	-0.14	-0.11	-0.04	-0.24	0.00
1912	0.12	-0.06	-0.13	-0.02	-0.12	-0.09	-0.04	-0.21	0.02

1913	0.09	-0.06	-0.13	-0.04	-0.15	-0.10	-0.05	-0.20	0.00
1914	0.09	-0.06	-0.14	-0.03	-0.11	-0.09	-0.05	-0.19	0.00
1915	0.10	-0.06	-0.12	-0.03	-0.14	-0.10	-0.05	-0.22	0.02
1916	0.11	-0.07	-0.14	-0.01	-0.15	-0.09	-0.02	-0.18	0.02
1917	0.16	-0.06	-0.07	-0.03	-0.15	-0.08	0.00	-0.25	0.02
1918	0.10	-0.07	-0.13	0.00	-0.14	-0.10	-0.03	-0.22	0.01
1919	0.07	-0.09	-0.11	-0.01	-0.14	-0.10	-0.03	-0.22	0.01
1920	0.11	-0.07	-0.13	-0.03	-0.14	-0.10	-0.05	-0.22	0.00
1921	0.12	-0.07	-0.13	-0.03	-0.14	-0.11	-0.04	-0.21	-0.01
1922	0.11	-0.05	-0.15	0.00	-0.12	-0.10	-0.05	-0.21	0.00
1923	0.10	-0.04	-0.10	-0.04	-0.11	-0.09	-0.05	-0.20	-0.01
1924	0.08	-0.09	-0.11	0.02	-0.14	-0.10	0.00	-0.21	0.03
1925	0.12	-0.05	-0.14	-0.01	-0.13	-0.10	-0.05	-0.21	0.01
1926	0.11	-0.07	-0.13	-0.02	-0.14	-0.10	-0.04	-0.21	0.00

Table 7-34: PRCC p-values for correlations between selected variables and the area farmed

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	-0.14	-0.10	0.23	0.02	0.02	0.10	0.11	0.21	0.01
1886	-0.17	0.05	0.10	-0.01	0.06	0.16	0.10	0.18	-0.01
1887	-0.14	0.05	0.13	0.04	0.13	0.12	0.09	0.21	0.01
1888	0.06	-0.07	**0.35**	-0.06	0.20	0.20	-0.09	0.19	0.07
1889	-0.13	0.06	0.12	0.00	0.10	0.16	0.05	0.21	0.01
1890	-0.07	0.10	0.12	-0.03	0.16	0.15	0.05	0.19	0.03
1891	-0.15	-0.11	0.24	0.07	0.10	**0.30**	-0.01	0.07	0.08
1892	-0.05	-0.04	0.24	0.02	0.07	0.23	0.10	0.07	-0.04
1893	-0.08	-0.06	-0.03	-0.09	0.07	0.18	0.05	0.05	-0.05
1894	-0.13	0.06	0.07	0.00	0.10	0.15	0.03	0.20	0.01

1895	-0.14	-0.02	0.18	0.05	0.14	0.09	0.02	0.21	0.01
1896	-0.18	-0.04	0.11	0.05	0.08	0.22	0.03	0.13	0.07
1897	-0.14	0.00	0.17	0.03	0.12	0.14	0.07	0.20	0.04
1898	-0.19	0.01	0.06	0.01	0.14	0.05	0.04	0.22	-0.03
1899	-0.10	0.05	0.09	0.04	0.08	0.11	0.07	0.18	0.01
1900	-0.17	0.03	0.16	0.06	0.08	0.17	0.07	0.17	0.09
1901	-0.18	0.05	0.08	-0.01	0.09	0.12	0.08	0.17	-0.03
1902	-0.12	0.05	0.11	0.03	0.12	0.11	0.04	0.23	-0.01
1903	-0.18	0.15	**-0.29**	0.06	-0.02	0.05	-0.04	0.15	-0.06
1904	-0.17	0.01	0.19	0.04	0.08	0.10	0.04	0.22	0.04
1905	0.04	-0.06	**0.35**	-0.08	0.13	0.25	-0.09	0.13	0.10
1906	-0.13	0.08	0.13	0.02	0.11	0.15	0.05	0.20	0.00
1907	-0.13	0.06	0.12	0.03	0.13	0.14	0.08	0.20	0.02
1908	-0.13	0.03	0.10	0.06	0.10	0.12	0.12	0.18	0.05
1909	-0.14	0.04	0.15	0.02	0.09	0.18	0.09	0.19	0.00
1910	**-0.30**	0.10	0.01	0.12	0.06	0.11	0.22	0.02	-0.06
1911	-0.11	0.02	0.20	0.11	0.16	0.19	-0.02	0.20	0.07
1912	-0.13	-0.01	0.04	0.05	0.06	0.10	-0.01	0.16	0.01
1913	-0.11	0.04	0.12	-0.01	0.16	0.13	0.08	0.19	0.03
1914	-0.19	0.05	0.04	0.00	0.07	0.14	0.09	0.15	-0.07
1915	-0.04	0.07	0.21	-0.13	0.18	0.25	0.06	0.20	0.11
1916	-0.09	0.04	0.14	-0.02	0.16	0.16	0.09	0.19	-0.05
1917	-0.11	0.05	0.12	0.03	0.14	0.15	0.04	0.20	0.01
1918	0.01	-0.06	0.06	0.16	-0.09	0.18	0.14	-0.03	0.09
1919	0.02	-0.03	0.26	0.00	0.23	0.09	0.01	**0.30**	0.03
1920	-0.13	0.00	0.06	-0.01	0.10	0.11	0.10	0.20	0.03
1921	-0.04	0.01	0.10	-0.05	0.18	0.16	0.03	0.19	-0.01

1922	-0.15	0.05	0.04	0.01	0.12	0.14	0.05	0.25	-0.02
1923	0.06	0.01	**0.41**	-0.05	0.15	0.05	-0.11	**0.30**	0.22
1924	**-0.31**	0.05	-0.01	0.06	0.05	0.15	0.23	0.11	-0.05
1925	-0.15	0.02	0.10	0.07	0.04	0.12	0.10	0.18	0.03
1926	-0.17	-0.01	0.18	0.03	0.03	0.09	0.09	0.20	0.00

Table 7-35: PRCC p-values for correlations between selected variables and the oats harvest

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.07	0.13	0.27	0.07	0.01	0.15	0.19	0.11	-0.03
1886	-0.12	0.27	-0.11	0.17	-0.07	0.17	0.28	-0.18	0.16
1887	-0.12	0.13	0.23	0.12	0.01	0.10	0.26	0.08	0.03
1888	0.01	0.17	0.12	0.00	-0.08	0.19	0.22	-0.08	0.02
1889	0.03	0.18	0.22	0.08	-0.08	0.16	0.27	0.06	0.12
1890	0.04	0.16	0.03	0.05	-0.15	0.10	0.24	-0.06	0.05
1891	-0.03	0.11	0.01	0.20	-0.20	0.17	**0.35**	-0.19	0.06
1892	-0.10	0.13	0.13	0.06	0.02	0.22	0.23	-0.03	0.05
1893	0.03	0.20	-0.08	0.05	-0.27	0.16	**0.37**	-0.24	-0.01
1894	-0.14	0.21	-0.03	0.07	0.10	0.21	0.18	0.11	0.09
1895	-0.01	0.26	0.04	0.15	-0.21	0.09	**0.39**	-0.15	0.09
1896	0.00	0.16	0.03	0.09	-0.26	0.06	**0.31**	-0.21	0.06
1897	0.09	0.18	0.19	0.00	-0.04	0.11	0.26	-0.03	-0.02
1898	-0.11	0.15	0.09	0.10	-0.01	0.13	0.20	0.09	-0.02
1899	-0.03	0.09	0.21	0.05	0.01	0.16	0.27	0.03	-0.02
1900	-0.12	0.25	0.04	0.10	-0.18	0.05	**0.42**	-0.11	0.09
1901	-0.01	0.18	-0.02	0.20	**-0.33**	-0.06	0.28	-0.26	0.06

1902	-0.08	0.15	0.15	0.07	-0.01	0.11	0.25	0.10	0.04
1903	0.20	-0.12	0.18	-0.10	-0.12	-0.01	-0.04	-0.13	0.06
1904	-0.01	0.13	-0.11	0.19	**-0.33**	0.04	**0.37**	-0.29	0.13
1905	-0.04	0.22	0.06	0.07	-0.20	0.13	**0.31**	-0.06	0.09
1906	-0.05	0.18	0.26	0.06	0.03	0.16	0.22	0.14	0.07
1907	-0.11	0.13	-0.12	0.10	**-0.29**	-0.04	**0.33**	-0.25	0.09
1908	-0.11	0.18	0.01	0.07	-0.17	0.12	**0.42**	-0.17	0.04
1909	0.05	0.16	0.16	0.07	-0.13	0.23	**0.31**	-0.03	0.07
1910	-0.10	0.16	-0.04	0.13	-0.21	0.16	**0.32**	-0.18	0.02
1911	0.02	**0.29**	0.20	0.07	0.00	0.17	0.25	-0.01	0.19
1912	-0.14	0.19	-0.03	0.21	-0.12	0.17	**0.31**	-0.24	0.06
1913	0.00	0.20	0.02	0.06	-0.13	0.13	**0.30**	-0.17	0.07
1914	0.01	0.13	-0.11	0.13	-0.25	0.04	**0.34**	-0.28	-0.10
1915	-0.08	0.19	-0.14	0.12	-0.27	0.01	**0.43**	-0.28	0.03
1916	-0.02	0.12	0.16	-0.02	-0.05	0.14	**0.29**	-0.08	-0.07
1917	-0.04	0.14	-0.01	0.12	-0.13	0.18	**0.31**	-0.21	0.07
1918	0.03	0.19	-0.04	0.14	-0.19	0.02	**0.38**	-0.23	0.05
1919	-0.04	0.13	-0.04	0.08	-0.10	0.11	**0.30**	-0.20	0.04
1920	-0.12	0.13	-0.04	-0.01	-0.09	0.25	**0.31**	-0.14	0.02
1921	0.01	0.13	-0.09	0.14	-0.17	0.12	0.28	-0.26	0.03
1922	0.01	0.15	0.02	0.12	-0.20	0.08	**0.29**	-0.13	0.06
1923	0.11	0.20	0.01	0.07	-0.24	0.01	0.24	-0.19	0.08
1924	0.04	0.14	-0.14	0.09	**-0.29**	0.00	0.24	**-0.34**	0.03
1925	-0.13	0.16	0.16	0.12	0.02	0.13	0.22	0.18	0.09

1926	-0.03	0.18	-0.14	0.15	-0.26	0.00	**0.43**	-0.26	0.03

Table 7-36: PRCC p-values for correlations between selected variables and the wheat harvest

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.06	0.15	0.25	0.08	-0.01	0.14	0.24	0.13	-0.02
1886	-0.11	**0.29**	-0.09	0.18	-0.06	0.16	**0.29**	-0.15	0.18
1887	-0.11	0.14	0.24	0.12	-0.01	0.09	0.28	0.10	0.06
1888	0.02	0.20	0.12	0.02	-0.08	0.17	0.24	-0.06	0.04
1889	0.04	0.20	0.23	0.07	-0.07	0.15	0.28	0.08	0.14
1890	0.04	0.16	0.03	0.04	-0.16	0.09	0.24	-0.05	0.06
1891	-0.01	0.13	0.03	0.18	-0.16	0.16	**0.35**	-0.13	0.06
1892	-0.06	0.17	0.08	0.08	-0.05	0.20	0.26	-0.07	0.06
1893	0.03	0.21	-0.10	0.07	-0.28	0.15	**0.38**	-0.24	0.01
1894	-0.14	0.26	-0.06	0.03	0.10	0.23	0.19	0.07	0.16
1895	-0.01	0.27	-0.02	0.18	-0.26	0.05	**0.39**	-0.21	0.11
1896	0.04	0.20	0.03	0.08	-0.23	0.04	0.28	-0.16	0.08
1897	0.09	0.19	0.17	0.01	-0.06	0.09	0.28	-0.01	0.00
1898	-0.08	0.19	0.01	0.14	-0.09	0.09	**0.30**	0.02	0.02
1899	-0.02	0.11	0.19	0.06	0.00	0.15	**0.29**	0.02	-0.01
1900	-0.02	0.25	-0.03	0.10	-0.21	0.06	**0.38**	-0.17	0.11
1901	0.01	0.20	-0.07	0.15	**-0.31**	0.04	**0.33**	-0.27	0.08
1902	-0.05	0.16	0.15	0.07	-0.01	0.09	0.26	0.13	0.04
1903	0.13	-0.01	0.14	-0.08	**-0.31**	-0.18	0.10	-0.05	-0.04
1904	0.00	0.14	-0.10	0.17	**-0.30**	0.04	**0.36**	-0.25	0.16
1905	-0.03	0.23	0.06	0.07	-0.19	0.10	**0.32**	-0.04	0.11

1906	-0.05	0.19	0.29	0.07	0.02	0.15	0.23	0.15	0.08
1907	-0.09	0.17	-0.16	0.13	**-0.29**	-0.03	**0.38**	-0.27	0.07
1908	-0.05	0.22	-0.02	0.07	-0.19	0.13	**0.37**	-0.20	0.10
1909	0.06	0.16	0.17	0.05	-0.12	0.22	**0.29**	-0.02	0.07
1910	-0.11	0.22	-0.11	0.17	-0.28	0.10	**0.37**	-0.25	0.08
1911	0.03	0.27	0.15	0.09	-0.04	0.14	0.27	-0.01	0.19
1912	-0.15	0.21	-0.03	0.21	-0.10	0.15	**0.31**	-0.21	0.06
1913	0.00	0.20	0.02	0.10	-0.14	0.10	**0.31**	-0.16	0.09
1914	-0.06	0.15	-0.11	0.13	-0.20	0.05	**0.32**	-0.25	-0.04
1915	-0.06	0.21	-0.11	0.10	-0.25	0.00	**0.42**	-0.26	0.03
1916	-0.01	0.18	0.10	0.03	-0.12	0.14	**0.36**	-0.08	0.01
1917	-0.06	0.18	-0.11	0.19	-0.19	0.11	**0.43**	-0.25	0.12
1918	0.04	0.20	-0.01	0.13	-0.18	0.00	**0.38**	-0.19	0.05
1919	-0.08	0.12	-0.06	0.14	-0.16	0.13	**0.32**	-0.17	0.08
1920	-0.09	0.14	-0.01	-0.01	-0.07	0.25	**0.30**	-0.10	0.05
1921	0.02	0.17	-0.07	0.15	-0.16	0.10	0.28	-0.24	0.06
1922	0.03	0.17	-0.03	0.15	-0.26	0.05	**0.32**	-0.16	0.07
1923	0.06	**0.31**	-0.05	0.09	-0.17	0.01	**0.34**	-0.23	0.09
1924	-0.01	0.20	-0.14	0.09	-0.27	-0.02	0.27	**-0.31**	0.03
1925	-0.12	0.20	0.16	0.14	0.00	0.15	**0.30**	0.12	0.04
1926	0.00	0.20	-0.09	0.13	-0.24	0.03	**0.41**	-0.22	0.06

Table 7-37: PRCC p-values for correlations between selected variables and the number of oxen

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	-0.25	0.02	**-0.32**	0.12	-0.14	0.13	0.00	-0.04	0.11

1886	-0.11	-0.03	-0.24	-0.04	-0.08	-0.09	0.04	-0.07	-0.04
1887	-0.22	0.09	-0.17	0.00	0.12	0.19	0.02	0.10	-0.08
1888	-0.07	-0.01	-0.19	0.01	-0.13	-0.05	0.06	-0.18	-0.01
1889	-0.20	-0.03	**-0.31**	-0.06	-0.04	-0.01	0.00	-0.21	-0.11
1890	-0.05	-0.03	-0.16	0.10	-0.17	0.04	0.06	-0.25	0.10
1891	-0.16	-0.02	**-0.49**	0.11	-0.02	0.15	-0.14	-0.07	-0.01
1892	-0.17	0.10	-0.14	-0.09	0.17	0.21	-0.11	0.17	-0.05
1893	-0.13	-0.10	-0.23	0.06	-0.08	0.22	-0.03	-0.12	0.08
1894	**-0.30**	0.06	**-0.41**	0.05	0.00	0.20	0.21	-0.06	-0.12
1895	-0.23	-0.05	-0.08	-0.13	0.11	0.15	-0.11	0.10	-0.05
1896	-0.03	-0.03	**-0.32**	0.11	-0.18	0.04	-0.10	-0.16	0.12
1897	-0.21	0.05	**-0.42**	0.07	-0.11	-0.01	0.02	-0.07	-0.05
1898	**-0.33**	0.05	**-0.38**	-0.05	0.06	0.18	0.04	0.09	-0.13
1899	-0.02	0.03	-0.12	0.02	-0.03	-0.02	0.03	-0.15	0.01
1900	-0.21	-0.01	**-0.32**	0.09	-0.16	-0.04	0.03	-0.14	0.06
1901	-0.15	0.04	-0.17	0.13	-0.15	-0.02	0.05	-0.14	0.19
1902	-0.08	-0.01	**-0.36**	0.05	-0.14	-0.03	-0.05	-0.27	0.01
1903	-0.18	-0.04	**-0.46**	0.03	-0.13	0.12	-0.07	-0.22	-0.11
1904	-0.09	-0.05	-0.28	0.10	-0.17	0.00	0.04	-0.21	0.03
1905	-0.09	-0.04	**-0.47**	0.08	-0.04	-0.14	0.01	0.04	-0.06
1906	-0.14	0.05	**-0.39**	0.07	-0.09	-0.12	-0.02	0.00	-0.08
1907	-0.26	-0.09	**-0.43**	-0.02	-0.11	0.05	0.03	-0.12	-0.19
1908	-0.24	0.13	-0.20	-0.10	0.16	0.09	0.04	0.19	-0.06
1909	**-0.34**	0.04	**-0.39**	-0.02	-0.01	0.11	0.10	0.00	-0.01
1910	-0.07	0.09	0.05	-0.11	0.22	0.27	-0.05	-0.01	-0.03

1911	-0.02	0.00	-0.05	-0.01	-0.08	0.02	0.06	-0.11	0.09
1912	-0.20	0.03	**-0.44**	0.07	-0.14	0.01	0.09	-0.01	-0.03
1913	-0.28	0.11	-0.26	0.05	0.04	0.04	0.12	0.04	-0.03
1914	-0.17	0.07	**-0.41**	0.08	-0.09	0.00	0.04	-0.06	0.03
1915	-0.10	-0.08	-0.26	0.03	-0.13	0.06	-0.07	-0.21	0.17
1916	-0.27	-0.04	**-0.31**	0.04	-0.02	-0.04	0.15	-0.04	-0.03
1917	-0.19	-0.18	**-0.30**	-0.07	-0.14	0.14	0.02	-0.18	-0.08
1918	-0.23	0.03	-0.06	-0.01	-0.02	0.10	-0.03	0.04	0.01
1919	-0.07	0.00	**-0.34**	0.04	0.04	-0.14	-0.08	0.06	0.00
1920	-0.14	-0.05	**-0.31**	0.06	-0.18	-0.08	-0.04	-0.22	0.04
1921	-0.14	-0.03	**-0.29**	0.16	0.03	0.04	0.10	-0.07	0.11
1922	**-0.30**	0.08	**-0.39**	0.17	-0.09	-0.04	0.18	-0.03	-0.06
1923	0.09	0.03	-0.13	-0.18	0.14	0.06	-0.26	0.09	-0.06
1924	0.06	0.00	**-0.36**	0.00	-0.12	0.00	-0.15	**-0.29**	0.11
1925	-0.29	-0.04	-0.27	0.09	-0.05	-0.01	0.01	0.02	-0.02
1926	-0.14	-0.03	-0.26	0.08	0.05	-0.17	-0.07	0.17	0.00

Table 7-38: PRCC p-values for correlations between selected variables and population size

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.05	0.01	0.06	-0.08	0.04	0.00	0.02	0.00	0.05
1886	0.12	-0.01	-0.06	-0.01	-0.06	-0.09	-0.02	-0.17	0.05
1887	-0.09	0.07	0.14	0.04	0.12	0.06	0.05	0.23	0.01
1888	0.13	0.03	-0.15	-0.03	-0.07	-0.09	-0.04	-0.23	0.09
1889	0.18	-0.10	-0.19	-0.12	-0.13	0.02	-0.09	-0.21	0.02
1890	0.16	-0.15	-0.17	0.02	-0.16	-0.02	-0.09	-0.25	0.05

1891	0.12	-0.04	-0.13	0.02	-0.14	-0.13	-0.03	-0.16	0.06
1892	0.13	-0.06	0.28	-0.11	0.12	0.09	-0.14	0.09	0.04
1893	0.10	-0.08	-0.04	-0.19	-0.05	0.10	-0.04	-0.14	-0.02
1894	0.07	0.00	-0.15	-0.02	-0.05	-0.02	0.04	-0.17	0.05
1895	0.17	0.01	0.00	0.03	-0.01	0.05	-0.09	-0.16	0.09
1896	0.21	-0.01	-0.08	0.02	-0.09	-0.06	-0.11	-0.22	0.08
1897	0.18	-0.06	-0.09	-0.01	-0.09	-0.11	-0.07	-0.19	0.09
1898	0.02	-0.05	**0.34**	-0.04	0.17	0.07	-0.17	**0.31**	-0.01
1899	0.14	0.00	-0.07	0.00	-0.10	-0.06	0.01	-0.15	0.05
1900	0.13	-0.08	-0.10	-0.04	-0.15	-0.05	-0.01	-0.15	0.09
1901	0.05	-0.01	-0.06	0.04	-0.06	-0.08	0.07	-0.12	0.13
1902	0.19	-0.07	-0.16	0.00	-0.08	-0.02	-0.11	-0.25	0.09
1903	0.11	-0.05	-0.11	0.00	-0.08	-0.08	-0.03	-0.22	0.11
1904	0.10	-0.04	-0.14	0.04	-0.10	-0.12	0.00	-0.16	0.13
1905	0.09	-0.07	-0.13	0.04	-0.16	-0.06	0.02	-0.20	0.10
1906	0.13	0.05	0.07	0.01	0.03	-0.09	0.01	0.03	0.11
1907	0.09	0.03	-0.07	0.04	-0.09	-0.08	0.01	-0.14	0.09
1908	0.15	0.08	0.05	-0.01	0.04	-0.07	-0.02	-0.09	0.15
1909	0.07	0.00	-0.06	0.01	-0.06	-0.07	0.01	-0.11	0.11
1910	0.25	0.00	0.02	-0.09	0.07	0.07	-0.11	-0.13	0.07
1911	0.09	-0.03	-0.13	0.03	-0.14	-0.08	0.03	-0.20	0.08
1912	0.11	-0.06	0.09	-0.14	0.01	0.11	0.03	-0.04	0.06
1913	**0.38**	-0.17	0.19	-0.09	0.06	0.07	-0.24	0.07	0.07
1914	0.13	-0.03	-0.07	0.00	-0.11	-0.12	0.06	-0.08	0.11
1915	0.15	-0.02	0.07	-0.17	0.11	-0.01	-0.11	-0.02	0.06

1916	0.01	-0.06	0.02	-0.02	-0.03	-0.05	0.09	-0.06	0.04
1917	0.04	-0.02	0.04	-0.02	0.00	-0.02	0.09	-0.08	0.02
1918	0.23	-0.11	0.01	-0.11	-0.01	0.17	-0.09	-0.19	0.06
1919	0.10	0.04	0.08	-0.11	0.09	0.04	-0.08	-0.04	0.02
1920	0.08	-0.07	-0.23	0.01	-0.11	-0.02	0.08	-0.24	-0.02
1921	0.01	-0.01	-0.11	0.05	-0.10	-0.11	0.09	-0.11	0.13
1922	0.14	-0.06	-0.16	-0.02	-0.19	-0.10	-0.01	-0.21	0.04
1923	0.23	0.03	0.17	-0.12	0.13	0.12	-0.13	-0.02	0.11
1924	0.04	0.05	0.00	-0.04	0.02	-0.03	-0.02	-0.09	0.10
1925	0.20	-0.06	0.12	-0.13	0.01	0.00	-0.13	0.05	0.02
1926	0.12	-0.11	-0.11	-0.01	-0.11	-0.10	-0.05	-0.22	0.09

Table 7-39: PRCC p-values for correlations between selected variables and TB mortality

	pInf	pPrg	tbsens	TBprevL	TBprevA	Tbcmin	Tbcmean	Tbcmax	cFood
1885	0.05	**0.35**	0.02	0.00	-0.03	-0.12	-0.06	-0.17	-0.06
1886	-0.03	**0.48**	-0.04	-0.22	0.10	**-0.36**	**-0.32**	0.00	0.02
1887	0.07	**0.35**	-0.07	0.10	0.08	**-0.38**	-0.04	0.12	0.26
1888	-0.03	0.07	0.00	0.24	0.03	-0.03	-0.08	0.02	0.13
1889	0.05	0.21	0.15	0.02	0.01	-0.09	-0.13	-0.02	-0.03
1890	0.01	0.15	0.04	0.10	0.13	-0.17	-0.01	-0.08	0.03
1891	0.09	0.21	0.02	0.17	-0.27	**-0.41**	-0.13	-0.11	0.25
1892	0.06	0.28	-0.08	-0.02	0.22	-0.04	-0.20	-0.02	0.28
1893	0.11	0.06	-0.07	0.06	-0.03	-0.23	-0.05	0.13	0.01
1894	**0.34**	**0.33**	0.04	0.05	0.01	-0.22	-0.19	0.09	0.19
1895	0.08	0.03	0.03	0.22	0.04	-0.16	-0.13	0.15	0.10

1896	-0.17	0.10	-0.17	**0.31**	-0.19	-0.22	0.16	-0.04	0.24
1897	-0.11	0.19	-0.02	0.28	0.05	-0.14	-0.15	-0.08	0.09
1898	0.07	0.20	-0.04	0.20	-0.13	-0.04	-0.13	**-0.31**	0.14
1899	-0.07	-0.01	-0.13	0.21	-0.05	-0.02	-0.11	0.00	0.03
1900	0.00	0.25	-0.03	0.18	-0.11	-0.07	-0.14	0.01	0.03
1901	-0.09	0.00	-0.10	-0.08	0.03	-0.11	-0.15	-0.04	0.03
1902	-0.05	0.09	-0.18	**0.34**	-0.24	-0.15	0.13	-0.02	0.02
1903	-0.06	0.12	-0.19	0.19	-0.05	-0.03	-0.22	-0.01	0.22
1904	0.04	0.15	0.08	0.25	-0.11	**-0.38**	-0.23	0.01	0.12
1905	0.01	0.17	-0.14	0.15	-0.26	0.00	-0.05	-0.08	-0.09
1906	-0.04	**0.36**	-0.16	0.16	-0.07	-0.16	-0.15	-0.22	0.15
1907	-0.29	-0.02	-0.20	0.26	**-0.30**	-0.19	-0.09	-0.08	0.10
1908	0.03	0.01	0.14	0.08	-0.11	-0.15	-0.10	0.27	0.20
1909	-0.28	-0.04	-0.14	0.16	-0.17	-0.08	0.08	0.00	-0.02
1910	-0.26	**0.40**	-0.15	0.18	-0.12	-0.12	-0.17	-0.09	0.08
1911	0.21	0.20	-0.07	0.11	0.11	-0.26	-0.09	0.20	**0.33**
1912	-0.03	0.17	-0.02	0.05	0.09	-0.23	-0.14	-0.06	-0.02
1913	0.16	**0.29**	0.07	-0.02	-0.10	-0.09	-0.11	-0.02	0.13
1914	0.05	0.22	0.01	0.22	-0.09	-0.22	-0.15	0.06	0.15
1915	0.11	0.25	-0.11	0.15	0.00	**-0.30**	-0.06	0.02	0.14
1916	0.03	**0.31**	-0.14	0.07	-0.01	-0.23	-0.13	0.13	0.02
1917	-0.08	0.19	-0.23	0.15	-0.04	-0.11	-0.15	0.10	0.28
1918	-0.26	0.03	-0.10	0.22	-0.20	-0.07	0.06	0.10	**0.32**
1919	0.06	-0.09	0.00	-0.05	-0.14	0.06	-0.04	0.26	**0.31**
1920	0.06	0.22	-0.07	0.09	0.16	0.09	-0.01	-0.10	0.07

1921	0.03	0.18	0.04	0.11	-0.01	-0.12	-0.13	0.05	0.01
1922	0.05	0.18	-0.06	0.04	-0.02	-0.19	-0.13	0.06	-0.05
1923	0.15	0.10	-0.06	-0.03	-0.11	-0.26	-0.17	0.03	-0.07
1924	0.00	0.03	-0.17	0.18	-0.05	**-0.29**	-0.21	-0.07	0.14
1925	0.12	0.15	-0.16	0.10	-0.12	-0.26	-0.12	-0.09	-0.11
1926	0.03	0.20	-0.14	0.11	0.10	-0.03	-0.20	-0.02	-0.05