Université de Montréal

Consanguinity, Epidemics and Early Life Survival in Colonial Quebec, 1720-1830

par Jessica Gagnon-Sénat

Département de démographie
Faculté des Arts et des Sciences

Mémoire présentée à la Faculté des études supérieurs et postdoctorales
en vue de l'obtention du grade de maîtrise ès sciences (M. Sc.) en démographie

Août, 2021

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé

Consanguinity, Epidemics and Early Life Survival in Colonial Quebec, 1720-1830

Présenté par

Jessica Gagnon-Sénat

A été évalué par un jury composé des personnes suivantes

Nadine Ouellette
Président-rapporteur

Alain Gagnon
Directeur de recherche

Lisa Y. Dillon
Codirecteur

Patrick Bergeron
Membre du jury

RÉSUMÉ

La consanguinité, soit l'union productive de conjoints partageant des allèles identiques provenant d'un ancêtre commun, s'est accumulée au fil du temps au Québec ancien. Parallèlement, le Québec a été victime de plusieurs épidémies. Le but de cette étude est d'évaluer la relation entre la mortalité des enfants et la consanguinité dans les périodes épidémiques du Québec ancien entre 1720 et 1830. D'une part, l'hypothèse émise est que les enfants ayant des gènes homologues sur plusieurs loci auraient un taux de mortalité significativement plus élevé comparativement aux enfants non consanguins, en raison du désavantage des homozygotes. D'autre part, les individus consanguins peuvent avoir une survie plus favorable en raison de l'effet d'enracinement, combien de générations une famille est établie dans la colonie, présent dans la mesure de la consanguinité. De plus, l'avantage social d'une famille étroitement liée peut favoriser la survie de l'enfant en accordant plus de soutien social aux parents et de surveillance de l'enfant. Les courbes de survie de Kaplan-Meier sont représentées graphiquement et des modèles de régression de Cox sont exécutés pour explorer et démêler partiellement les rôles des facteurs génétiques et environnementaux. Les immigrants, les naissances multiples et les individus sans généalogie du *Registre de la population du Québec ancien* (RPQA) et de l'*Infrastructure intégrée des microdonnées historiques de la Population du Québec* (IMPQ) sont exclus. Au total, 610 412 individus sont analysés dans les modèles de Cox. Les rapports de risque pour les épidémies augmentent avec l'âge et les rapports de risque pour la consanguinité éloignée ressemblent souvent au groupe référence, les non consanguins. De plus, les effets diffèrent selon le sexe et le groupe d'âge. Généralement, si les enfants avec une consanguinité proche, ceux identifiés comme consanguins avec seulement trois générations ascendantes, ne subissent pas de surmortalité dans un groupe d'âge précédent, les modèles de Cox signalent une survie défavorable de ces individus lors des épidémies. Des effets sous-jacents tels que des processus de sélection et des variables de contrôle relatives à l'enracinement peu robustes guident les résultats de l'interaction entre les épidémies et la consanguinité, de sorte que la prémisse reste à valider.

Mots-clés : consanguinité, épidémies, mortalité infantile, mortalité des enfants, Québec ancien, analyses de survie

ABSTRACT

Consanguinity, the productive union of spouses sharing identical alleles from a common ancestor, accumulated over time in Colonial Quebec. Concurrently, Quebec was the victim of several epidemics. The aim of this study is to evaluate the relationship between child mortality and consanguinity in epidemic periods of Colonial Quebec between 1720 and 1830. On the one hand, it is hypothesized that children with homologous genes on many loci would have a significantly higher mortality rate compared to non consanguineous children, due to homozygote disadvantage. On the other hand, consanguineous individuals may have a more favourable survival because of the effect of settlement, how many generations a family has been in the colony, present in the measure of consanguinity. Further, the social benefit of a closely bound family may favour child survival by providing more social support to the parents and child supervision. Kaplan-Meier survival curves are graphed, and Cox regression models are run to explore and partially disentangle the roles of genetic and environmental factors. Immigrants, multiple births and individuals lacking a genealogy from the *Registre de population du Québec ancien* (RPQA) and *Infrastructure intégrée des microdonnées historiques de la Population du Québec* (IMPQ) are excluded. Altogether, 610,412 individuals are analysed in the Cox models. Hazard ratios for epidemics increase with age and distant consanguinity hazard ratios often resemble the no consanguinity reference group. Further, the effects differ by sex and age group. Generally, if closely consanguineous children, those identified as consanguineous with only three ascending generations, do not undergo excess mortality in a previous age group, the Cox models signal an unfavourable survival of these individuals during epidemics. Underlying effects such as selection processes and unrobust control variables for settlement guide the results of the interaction between epidemics and consanguinity, so the premise, though convincing, remains to be validated.

Keywords: consanguinity, epidemics, infant mortality, child mortality, Colonial Quebec, survival analysis

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF TABLES IN ANNEXES

LIST OF FIGURES IN ANNEXES

LIST OF ABBREVIATIONS AND ACRONYMS

Consang.: Consanguinity

etc.: et cætera

ex.: for example

HR: Hazard ratio

IBD: identical by descent

i.e.: id est

IMPQ: Infrastructure intégrée des microdonnées historiques de la Population du Québec

PRDH: Programme de recherche en démographie historique

REF: Reference category in survival analysis models

RPQA: Registre de population du Québec ancien

SLSJ: Saguenay-Lac-Saint-Jean

y.o.: year(s) old

To everyone who helped me persevere,

ACKNOWLEDGEMENTS

First, I would like to thank my director, Alain Gagnon, for his insightful comments, support, and direction. This thesis would not have been as complete without his speciality in consanguinity and epidemiology within demographic studies, as well as his methodological expertise. Thank you to my co-director, Lisa Dillon, who has not only helped me build my initial idea for this thesis, but also trusted me with contracts of employment for the *Programme de recherche en démographie historique (PRDH)*.

I am grateful to the *Département de démographie* for having awarded me the *Prix d'excellence Jacques-Henripin* and for the *Faculté des études supérieures et postdoctorales*'s scholarship of Excellence.

Fellow students of the Department, your support, encouragements and discussions always motivated me to persevere, and for that I thank you. Among these students, a special mention goes to Cathie-Anne Dupuis, my fellow colleague at PRDH and friend. She inspired me to continue in the most tenacious times; her motivation and support were consequential to the completion of this thesis. Also, thank you to my colleagues at the *Institut de la statistique du Québec* who, not only supported my decision to complete my thesis, but encouraged me to do so on a regular basis.

Thank you to my family and friends for your support, patience and belief in me. It means the World. But especially, thank you to my spouse, Jephthé, who was there for all the highs and lows of my master's degree and never stopped believing in my success. Thank you for pushing me and always having my best interests at heart.

This journey in my Graduate Studies would not have been such a learning and wholesome experience without all of you.

INTRODUCTION

High rates of consanguinity are still observed today in some countries in the Middle East and Africa (Bener et collab., 2007 : 262) even though the detrimental effect of consanguinity on health has been attested through increased infant mortality rates and rare diseases (Bener et Mohammad, 2017 : 316). Consanguinity is the result of the productive union of spouses sharing identical alleles, variant forms of a gene, from a common ancestor (Bittles, 2003 : 571 ; Bittles et Black, 2010b : 194 ; Scitable by Nature Education, 2014). Several studies have addressed the negative association between consanguinity and health or consanguinity and survival (Bener et collab., 2007 ; Bener et Mohammad, 2017 ; Bittles, 2003 ; Bittles et Black, 2010b ; Stoltenberg et collab., 1999). However, the effect of consanguinity on child survival during crises has not been addressed in literature.

The motivation for this study is Gauvin's (2016 : 114) advance that consanguineous individuals may be immunosuppressed as they have lower frequencies of white blood cells. On this basis, consanguineous children would have higher mortality during epidemic periods compared to their non consanguineous counterparts.

To further comprehend the impacts of consanguinity and epidemics on early life mortality, it is useful to observe these relations historically. Studying the mortality of consanguineous individuals in a period before modern medicine allows for a unique understanding of the phenomenon at a time prior to medical interventions that could mitigate its effects on health and mortality. Historical data sources are useful in this regard and available for Quebec. The *Registre de population du Québec ancien* (RPQA) and the *Infrastructure intégrée des microdonnées historiques de la Population du Québec* (IMPQ) are used in this research to explore the relationship between consanguinity and child mortality in Colonial Quebec from 1720 to 1830, and specifically during periods of epidemics, which may have compounded the effect of consanguinity on mortality. This period was chosen as there are not many consanguineous individuals in the colony prior to 1720. Further, the reconstitution of parish data allows the longitudinal study of infant and child mortality in Colonial Québec until 1830 without considerable loss of information. The RPQA-IMPQ permits the study of the semi-closed

population for many generations as migration out of the colony was limited and internal migrants remain in the data (Dillon et collab., 2017).

Child mortality in Colonial Quebec is studied with survival analysis by observing the age at death of consanguineous and non consanguineous children during epidemic and non epidemic periods. Consanguinity is measured genealogically using inbreeding coefficients - the probability of sharing identical alleles through a common ancestor - above 0.39%. Epidemic peaks are found by graphing the number of deaths per month and age group with the RPQA-IMPQ data. Past research permits the identification of many epidemics by name.

Chapter 1 presents the literature review which addresses the historical context of Colonial Quebec, its epidemics, and common concepts related to consanguinity such as its consequences. Protective effects against epidemics are discussed, such as the passive immunity transmitted by the mother at labour and during breastfeeding, as well as the known effects of reproductive variables on infant mortality in Colonial Quebec. Chapter 2 describes the objective of the thesis, conceptual framework and hypotheses. Chapter 3 explains the parish-based database, the variables analysed, data selection, as well as the methodology. Cox analyses are presented following the descriptive analyses in Chapter 4. Finally, Chapter 5 further discusses the results and limits of the study, as well as concludes the thesis.

This research shows the great potential of historical data in the study of epidemics and their effect on mortality at a time prior to vaccination. Without the detailed family reconstitutions available with the RPQA-IMPQ, consanguinity could not have been studied; In this regard, Gagnon (2000) described in his thesis that marital dispensations allowing for the union of kin do not capture all consanguineous marriages. Even without the need to control for medical intervention, other methodological difficulties encountered in the course of this research prevent drawing a simple, unambiguous conclusion about the association of consanguinity and epidemics in early childhood survival. There are signals of unfavourable survival for consanguineous persons during epidemics, yet this effect seems to depend on unobserved selection effects. Further, the measure of consanguinity is simultaneously an indicator of extended family settlement within the colony. Accordingly, individuals whose families have resided in Colonial Quebec for several generations may benefit from protective

effects in the form of childcare from social networks, an established land, or greater chances of passive immunity to epidemics in infancy, all of which favour survival.

This explorative thesis sheds light on the range of data necessary to better control limits and underlying effects associated with studying consanguinity and epidemics concurrently. This work on the differential mortality of consanguineous children in and out of epidemics will signal the way for future studies in health, genetics, epidemiology, and, evidently, demography.

CHAPTER 1: LITERATURE REVIEW

Preliminary information about the region and its population must be advanced to appreciate the study of consanguinity in Colonial Quebec. This first chapter presents the literature review which describes the historical context of Colonial Quebec, its epidemics, and common concepts related to consanguinity. First, Colonial Quebec's foundation will be put forth, followed by the child mortality levels of the region to understand the various factors influencing this indicator. Then, epidemics that occurred in Colonial Quebec will be addressed, as well as specificities related to each epidemic. Next, the common concepts related to consanguinity which include indicators, causes, and theoretical consequences are described. Lastly, the findings will be reviewed to pave the way to the objectives of the project that are brought forth in Chapter 2.

1.1. Historical Context

With its first French settlement founded in 1608, Colonial Quebec is distinguished by its semi-closed character; after an initial period of immigration by the founders of the colony, immigration to Quebec and emigration from Quebec were limited (Charbonneau et collab., 2000 : 99 ; Dillon et collab., 2017 : 22). The initial period of immigration was to fulfill France's main interest for the colony: the fur trade (Charbonneau et collab., 2000 : 99-100). It coincided with the development of the three first cities: Quebec in 1608, Trois-Rivières in 1634 and Montreal in 1642 (Harris et Matthews, 1987). However, the "major immigrant wave" occurred in the second half of the seventeenth century, following the 1663 colonization policy which valued natural increase of the population (Charbonneau et collab., 2000 : 100, 102). The colony counted only 5,000 to 10,000 founders; however, due to a high rate of natural growth, the population increased to more than 70,000 by 1760 and over one million by 1861 (Charbonneau et collab., 2000 : 104 ; Dillon et collab., 2017 : 22). This led to what is commonly called a "founder effect", which occurs when a restricted group of founders gives birth to a (usually but not necessarily isolated) population with little to no influx from immigrations following the founding event (Mayr, 1942). Consequently, as the population relies on its own limited core to reproduce itself, consanguinity accumulates over time as descendants share more and more genes inherited from

the founders and are said “identical by descendance” (Bouchard et De Braekeleer, 1991a ; Gagnon, 2000 ; Gauvin, 2016 : 9 ; Mayr, 1942 ; Morgan, 2016 : 1). Quebec does not offer a typical example of a “founder population” – the typical example is rather that of an isolated island peopled by a handful of founders – but rather a special case where the founder effect has been regionalized. One region in particular, the Saguenay-Lac-Saint-Jean, has all the hallmarks of a founder effect, with the presence of rare, autosomal recessive disorders, while the gene pool of the rest of the province, and especially the southern part, shows no evidence of a founder effect (Bouchard et collab., 1995 ; Bouchard et De Braekeleer, 1991b ; Gagnon et Heyer, 2001).

The Catholic Church occupied an institutional role in Colonial Quebec. In the 1600s, religion occupied a political role as well as a social role (Trudel, 1968 : 276). Members of the Catholic Church also governed the schools and hospitals (Charbonneau et collab., 2000 : 101). This vast display of power in the beginning of the colony ensured supremacy of the Church even after their political authority waned in the eighteenth century (Trudel, 1968 : 276). Indeed, individuals of New France attributed a quarter of each year to religious obligations, including Sunday Mass and Catholic celebrations (Trudel, 1968 : 275). Society had to adhere to Catholic rules including baptising their newborns within days of the birth, learning their prayers, confessing annually, and following nuptial guidelines which disapproved of marriage with kin (Trudel, 1968 : 271-274). The members of the Church recorded all ceremonies and confessions, hereby institutionalising civil records which are contemporarily used as databases to study the historic population of Colonial Quebec (Dillon et collab., 2017 ; Trudel, 1968 : 275).

Since the end of the seventeenth century, the population grew mostly in consequence of the high rate of natural growth, however Colonial Quebec lost its semi-closed nature in the turn of the nineteenth century which allowed for growth due to migration to occur as well. In eighteenth century Colonial Quebec, 40% of women having lived complete fertile lives had at least 10 children (Charbonneau et collab., 2000 : 119). Moreover, the high fertility of the population and the semi-closed character of the colony contributed to the restricted marital market, which in turn, lead to dense family networks (Dillon, 2016). These dense networks favour an increase in relatedness - the number of genes shared between individuals - particularly in the least populated regions (Koellner et collab., 2018). In the beginning of the nineteenth century, in what was then Upper Canada and Lower Canada, the population grew on account of

many immigrants from United States (McInnis, 2000 : 376-378). Afterwards, British immigrants were the majority, the population reaching 20 000 immigrants in 1819 (McInnis, 2000 : 378-379). These immigrants had the potential to diversify the gene pool available in the region, which would have facilitated non consanguineous unions. However, people may have preferred to marry individuals who resemble themselves in terms of language and ethnicity. Mixed marriages were not as easy as one may think as anglophones and francophones did not share the same language, religion, and often geographic location (Gauvreau et collab., 2010 ; McInnis, 2000 : 376).

1.1.1. Child mortality in Colonial Quebec

According to a study of legitimate child mortality in Colonial Quebec from 1680-1750, about 73% of children with an exact birth date recorded in the RPQA (*Registre de population du Québec ancien*) database survived until the age of five; most child deaths occurring in the first year of life (Amorevieta-Gentil, 2010 : 131 ; Beise, 2004 : 7 ; Charbonneau et collab., 2000 : 124 ; Gagnon et Mazan, 2009 ; Lalou, 1997 : 206). Further, mortality was more important during epidemics. Bruckner et collab. (2018 : 4) identified important and regular peaks in mortality in tune with smallpox epidemics for children under age seven. As for infant mortality, regardless of epidemics, apparent annual death probabilities were between 140‰ and 230‰ in the first half of the seventeenth century and dropped in the beginning of the second half of the seventeenth century, though infant mortality is underestimated for this period as many babies died prior to being registered in the parish data (Amorevieta-Gentil, 2010 : 131). Following the 1660's, apparent infant death probabilities fluctuated well into the nineteenth century, but remained higher than the rates observed during the seventeenth century; reaching new maximums of 350‰ in 1748-1749 (Amorevieta-Gentil, 2010 : 131).

Moreover, Amorevieta-Gentil illustrates factors influencing infant mortality in the context of natural fertility in Colonial Quebec (2010 : 74). She surveyed bio-genetic and environmental factors, including socioeconomic conditions and sociocultural practices, however scarce in data, environmental pathology, as well as political and climate conditions. Some bio-demographic factors include the age of the mother at birth, the rank of the birth, the birth interval between the child of interest and the previous birth as well as infant mortality of

siblings. Even the sex of the child affects mortality as boys are more at risk of death within their first year of life than girls (Amorevieta-Gentil, 2010 : 57 ; Lalou, 1990 : 245). Further, boys are also more susceptible to intrauterine death (Catalano et collab., 2005). In fact, there are selection processes visible in post-birth survival when events affecting a particular male cohort lead to a more significant number of intrauterine deaths for that cohort. A study of Swedish cohorts from 1751 to 1912 showed that "the "most culled" cohorts lived ≈ 3.7 more months than was expected from histor[ical data]" (Catalano et Bruckner, 2006 : 1641). In opposition, the "least culled" male cohort lived approximately 3 months less "than expected from the lifespan of females in that cohort as well [as] from historic trends unique to male cohort lifespan" (Catalano et Bruckner, 2006 : 1641). Thus, a period of exceptionally high mortality is often followed by a period of lower mortality within the same cohort. This effect is based on the typical trend of mortality in an area. However, not every region has the same mortality trends. In Colonial Quebec, differential mortality was also observed between the north and the south of the St-Lawrence River (Gagnon, 2012). The south west of the colony experienced more epidemics than, for instance, the north east (Mazan, 2011a ; Mazan et collab., 2009). Therefore, infant mortality was higher in regions more typically affected by mortality crises. Further, urban infant mortality is greater than rural infant mortality; there is also a superior quality of recorded data in urban areas which may very slightly affect the data, however, distinction in survival risk is undeniable (Amorevieta-Gentil, 2010 : 145-146). Additional environmental factors affecting infant mortality include nursing practices, famines and sanitary conditions (Amorevieta-Gentil, 2010 : 74 ; Boisvert et Mayer, 1994 ; Lalou, 1990 : 275 ; Maheu, 2001 ; Nault et collab., 1990 ; Thornton et Olson, 2011). In fact, the unfavourable sanitary conditions discussed are usually observed in urban areas and cause diseases and deaths.

Many articles researching child mortality or consanguineous trends in Colonial Quebec have examined similar variables affecting survival. Foremost, Nault et collab. (1990 : 285) studied the "effects of reproductive behaviour on infant mortality of French-Canadians during the seventeenth and eighteenth centuries" and concluded that three determining factors of infant mortality are mechanically linked: the proportion of siblings dying in the first year of age, birth intervals and sibship size. Indeed, if a mother bears children who die as infants, then she halts breastfeeding at the death of her child, therefore augmenting her risks of pregnancy as she no

longer benefits from lactational amenorrhea, her natural contraception, which facilitates short birth intervals between children, in turn promoting a large family size by the end of her fertile life (Nault et collab., 1990 : 281-285). Further, the shortened birth interval may put the mother at risk of complications if her body did not have enough time to recover from the previous birth (Amorevieta-Gentil, 2010). Infant mortality of siblings, short birth intervals, and high sibship size all bear negative relationships with survival but are correlated with one another (Amorevieta-Gentil, 2010 : 74 ; Gagnon et Mazan, 2009 : 1613 ; Nault et collab., 1990 : 281-285 ; Robert et collab., 2009 : 674). This issue of interdependence persists with the study of other factors related to child mortality such as the age of the mother at the birth, high birth rank, intergenesic intervals and family size (Amorevieta-Gentil, 2010 : 74 ; Boisvert et Mayer, 1994 : 698-705 ; Maheu, 2001 : 137). The mechanism at work in this case is that the older the mother is, the higher her family size and her newborns will be of higher birth order, all of which increase the risk of infant mortality, which further affects the intergenesic intervals between siblings. Even maternal death is dependant on these factors as bearing many children in a short time interval drains the body, which affects the survival of the mother as her health is put in peril, in turn, disturbing the health of the infant and its chance at survival, especially breastfed infants (Amorevieta-Gentil, 2010 : 74 ; Boisvert et Mayer, 1994 : 706 ; Gagnon et Mazan, 2009 : 1613). More specifically, Amorevieta-Gentil proposed four environmental factors affecting infant mortality in New-France: socioeconomic conditions, sociocultural practices such as breastfeeding, pathological environment as well as climatic and political conditions (2010: 74). The following table summarizes some variables used in the study of child mortality in Colonial Quebec. The variables retained in this study follow many categorisations of Amorevieta-Gentil's thesis (2010) on infant mortality in Colonial Quebec. The retained control variables for the study of consanguinity and epidemics on child survival are the period, the region and its urban or rural context (based on Gagnon's (2000) thesis), the birth interval between the index child and their previous sibling, the fate of the previous sibling (if sibling died prior to one year old), the mother's age at birth, the rank of the child, and the number of great grandparents found in the database as a proxy to extended family settlement. This is further discussed in Chapter 3.

Table 1. Variables of Interest in Studies about Child Mortality in Colonial Quebec

Variable	Source
Sex of child	Amorevieta-Gentil, 2010 : 74
Intergenesic interval /Birth interval	Boisvert et Mayer, 1994 : 704-705 Amorevieta-Gentil, 2010 : 74 Maheu, 2001 : 137 Robert et collab., 2009 : 674 Nault et collab., 1990
Birth Rank	Boisvert et Mayer, 1994 : 703-704 Nault et collab., 1990 Gagnon et Mazan, 2009 : 1613 Amorevieta-Gentil, 2010 : 74
Death of siblings	Nault et collab., 1990 : 281-282 Amorevieta-Gentil, 2010 : 74, 211
Mean # siblings surviving until a certain age Age at death of siblings	Gagnon et Mazan, 2009 : 1613
Family size	Nault et collab., 1990 : 280-281 Robert et collab., 2009 : 674
Parents' age at birth of index child (usually mother's)	Amorevieta-Gentil, 2010 : 74 Beise, 2004 : 25 Boisvert et Mayer, 1994 : 701-702 Nault et collab., 1990 Maheu, 2001 : 137
Difference in age between spouses	Boisvert et Mayer, 1994 Maheu, 2001 : 137
Maternal and Paternal deaths	Gagnon et Mazan, 2009 : 1613 Amorevieta-Gentil, 2010 : 74 Boisvert et Mayer, 1994 : 706
Mean length of fecund life for women	Boisvert et Mayer, 1994 : 699
Kinship level of parents	Maheu, 2001 : 137
Social Status/Education/Profession	Boisvert et Mayer, 1994 : 706-707 Amorevieta-Gentil, 2010 : 74
Period	Boisvert et Mayer, 1994 : 708 Maheu, 2001 : 137
Region of Residence (East/West) or Urban/Rural	Gagnon et Mazan, 2009 : 1613 Amorevieta-Gentil, 2010 : 74
Epidemic	Boisvert et Mayer, 1994 : 709 Amorevieta-Gentil, 2010 : 74

The table shows that many variables affect child mortality, and they may all operate concurrently. Therefore, many control variables are necessary in the study of mortality, especially at young ages when reproductive variables play a greater role.

The levels of infant mortality in Colonial Quebec were high in the eighteenth century compared to contemporary standards, but the risk of infant death was even higher during periods of crises, notably the Conquest of Quebec (Amorevieta-Gentil, 2010 : 55, 131). In fact, Amorevieta-Gentil uses infant mortality peaks as indicators of infectious outbreaks in Colonial Quebec (2010 : 84, 131).

1.2. Concepts of Epidemics

1.2.1. Epidemics in Colonial Quebec

"An epidemic occurs when an infectious disease spreads rapidly throughout a community at a particular time" (Cadotte, 2013 : 1). With rapid population growth, Colonial Quebec eventually presented an ideal environment for the hasty spreading of infectious diseases, as the risk of spreading a virus increases with the size and density of the population (Cadotte, 2013 : 1 ; Mazan, 2011a : 31). Desjardins (1996 : 60) estimates a population of 18,159 on the territory in 1702, which was enough to allow for the transmission of infection considering the smallpox epidemic of that time. Colonial Quebec was the victim of several epidemics between settlement and 1830; among other outbreaks, there was typhus in 1687, smallpox in 1702-1703, measles in 1714-1715, smallpox again in 1732-1733 and typhus again in 1746-1750 (Amorevieta-Gentil, 2010 ; Bernard, 1994 ; Charbonneau et collab., 1993 ; Desjardins, 1996 ; Gagnon et Mazan, 2009). Some epidemics such as smallpox and measles only affect those who have not had contact with the infection before. With those airborne diseases, an infected individual either generates antibodies that will help fight the disease during a subsequent outbreak or dies. This leads to the reduction of the population at risk of infection, i.e., the so-called pool of "susceptible" (Mazan, 2011a : 6). Data suggests that other epidemics have affected residents of the St. Lawrence Valley, but they are not well documented (Amorevieta-Gentil, 2010 ; Mazan, 2011a : 29). Furthermore, not all sources agree on the time frame of an epidemic, nor on the specific disease that spread within a specific time frame. For instance, for

1687 to 1688, some sources claim a spread of typhus in Colonial Quebec (Amorevieta-Gentil, 2010 : 131 ; Bernard, 1994) while others suggest the infection was smallpox or purpuric fevers (Trudel, 1968 : 241). There is a source which claims both types of infection: typhus and smallpox or purpuric fevers (Desjardins, 1996 : 50-51). Next, some known epidemics from Colonial Quebec are presented and briefly described.

An unsanitary environment is often the precursor of the typhus epidemic, ergo why it was commonly transported by boat in Colonial Quebec (Batten et collab. (dir.). 2017 : 2100-2101 ; Dechambre, 1885 : 574 ; Pâquet, 1999 : 278, 282). Typhus has many symptoms preceding the rash including headaches, fever and chills (Batten et collab. (dir.). 2017 : 2101). The fever can last two weeks and be accompanied by low blood pressure, confusion or seizures, and delirious episodes, that typically occur in the second week of infection (Batten et collab. (dir.). 2017 : 2101 ; Dechambre, 1885 : 583). Consequences of the infection include damaged organs, coma and death (Batten et collab. (dir.). 2017 : 2101).

New France is thought to have had typhus outbreaks in 1659, 1687-1688, 1717, 1742-1744, 1748-1750, and 1755-1756 (Amorevieta-Gentil, 2010 : 131 ; Bernard, 1994 : 19, 24 ; Trudel, 1968 : 241). Among these, the outbreaks considered epidemics are the ones in 1687-1688 and the late 1740's (Amorevieta-Gentil, 2010 : 131 ; Bernard, 1994 : 19 ; Cadotte, 2013 : 4). In 1746, a severe typhus outbreak hit the east of Colonial Quebec, killing a third of the Mi'kmaq population (Cadotte, 2013 : 4). Other researchers time the epidemic in 1748-1750 or declare two typhus epidemics, one in 1743-1746 and the other in 1750 (Amorevieta-Gentil, 2010 : 131 ; Trudel, 1968 : 241). Lack of research on this infection prevents additional information to be gathered, such as death rates.

Next, smallpox "is caused by the variola virus and is found only in humans" (Batten et collab. (dir.). 2017 : 194). This is a viral infection that affects only the susceptible (Bernard, 1994 : 26). Some symptoms of smallpox resemble the flu such as headaches, fevers, body aches and vomiting, but these symptoms are accompanied by painful rashes (Batten et collab. (dir.). 2017 : 194). Ultimately, the sickness may last 30 days (Bernard, 1994 : 6). The highly contagious and potentially fatal disease can infect a susceptible person through direct contact with the fluids of smallpox blisters, whether it be on a person or object. Further, the inhalation of infected air

due to the coughing victims of smallpox may also spread the disease (Batten et collab. (dir.). 2017 : 194).

There were several smallpox outbreaks in New France: 1702-1703, 1732-1733, 1755-1757, 1777, 1784, and 1798 (Amorevieta-Gentil, 2010 : 133 ; Bernard, 1994 : 2, 23-24 ; Bruckner et collab., 2018 ; Desjardins, 1996 : 64 ; Spaulding et Foster-Sanchez, 2019 : 2). In 1702-1703, no one was spared the smallpox epidemic. The contagion killed 6-6.5% of the population, roughly 1300 individuals (Desjardins, 1996 : 63-64), and mostly children and women of childbearing ages (Desjardins, 1996 : 59-60). In 1733, the epidemic killed mostly individuals under 30 years old, ergo those who did not experience the 1702-1733 epidemic (Bruckner et collab., 2018 : 3). The Canadian Encyclopedia cites Montcalm, a military commander of New France, to describe the disastrous smallpox epidemic of 1755-1757 as he testified in 1757 that 2,500 cases of smallpox affected Quebec City, and estimated that 20% of those infected died (Spaulding et Foster-Sanchez, 2019 : 2). Starting with the 1777 smallpox epidemic, a seven year cycle of smallpox is identified by Bruckner et collab. (2018) until the end of the eighteenth century, but it may have continued in the following century.

Measles infections are typically of shorter duration than smallpox, typically 8 to 12 days, and include symptoms such as fever, cough and rashes (Perry et Halsey, 2004 : S4). The groups most at risk of contracting the disease are young children (<5 years old) and adults (>20 years old) (Perry et Halsey, 2004 : S4). Historically, children rarely died of measles per se. Rather, the virus reduced the efficiency of the immune system causing children to die of otherwise minor illnesses such as diarrhea and pneumonia (Mazan, 2011a : 15 ; Perry et Halsey, 2004 : S5-S8). Therefore, a risk factor associated with the disease is immune deficiency (Perry et Halsey, 2004 : S11), a weakness which potentially afflicted consanguineous persons in Colonial Quebec (Gauvin, 2016 : 114).

The 1714-1715 measles epidemic of Colonial Quebec is the most studied epidemic of New France and was the subject of a PhD thesis (Mazan, 2011a). However, this specific epidemic will not be studied in the context of this thesis as not many children living in the colony prior to 1720 are identified as consanguineous. There was another measles outbreak in 1729-1730, but it has not been as heavily researched (Amorevieta-Gentil, 2010 : 131). The outbreak of 1714-1715, an important epidemic in historical demographic standards, spread to all regions

of Colonial Quebec in a few months and killed many (Mazan, 2011a : 44). For instance, the uncorrected probability of death of infants rose from 215‰ in 1713 to 300‰ in 1714 (Amorevieta-Gentil, 2010 : 131). Even the death rates of young children aged 1 to 4 years old increased significantly within the 2nd quarter of 1714 and the 1st quarter of 1715 going from approximately 88 deaths per thousand to 111 deaths per thousand (Mazan, 2011a : 48). Furthermore, it must be noted that these colony-wide statistics conceal important regional effects; the mortality of young children was more intense in Quebec City, the rural East and Montreal during the measles epidemic of 1714-1715 (Mazan, 2011a : 49, 2011b : 49 ; Mazan et collab., 2009 : 314).

There are other epidemics which occurred in Colonial Quebec which were short lived or simply not as infamous or prevalent in research as typhus, smallpox, and measles. For example, tuberculosis, which resulted from unsanitary conditions, was brought to New France by Europeans in the seventeenth century. It returned to the territory periodically, often in tune with immigrant waves (Bailey et collab., 2019 : 2). Unlike some other highly infectious diseases, tuberculosis requires extended exposure to a coughing patient in order to spread (Bailey et collab., 2019 : 4). Another example concerns yellow fever, which is transmitted by mosquitos; fortunately, Quebec's cold winter was able to kill the spread of the deadly infection (Cadotte, 2013 : 5). Influenza was another common infection which infected the Colonial Quebec population, for instance, in 1700 (Amorevieta-Gentil, 2010 : 131 ; Desjardins, 1996 : 50 ; Trudel, 1968 : 241).

Given the various disagreements in the literature concerning the timing of epidemics in Colonial Quebec, the next table presents dates attributed to epidemics in the region. The last column of the table indicates any conflicting information between authors, whether it be the years of the epidemics or the epidemic itself.

Table 2. Epidemics in Colonial Quebec from 1634-1798 by Source

Epidemic, year	Source	Author's claims (if different)
Pandemic 1634	Delâge, 2006 : 115	Probably smallpox
Smallpox 1639	Delâge, 2006 : 115	
Smallpox 1641	Delâge, 2006 : 115	
Typhus 1659	Cadotte, 2013 : 4	
	Amorevieta-Gentil, 2010 : 131	
Ship fevers 1666	Canada. Dept. of Agriculture, 1873 : 160-162 Desjardins, 1996 : 50	
Smallpox 1669-1670	Canada. Dept. of Agriculture, 1873 : 160-162 Desjardins, 1996 : 50	
	Amorevieta-Gentil, 2010 : 131	
Unspecified disease 1684	Canada. Dept. of Agriculture, 1873 : 160-162 Desjardins, 1996 : 50	
Typhus 1685	Desjardins, 1996 : 50	
	Trudel, 1968 : 241	
Typhus 1687-1688	Bernard, 1994	
	Amorevieta-Gentil, 2010 : 131	
	Trudel, 1968 : 241	Smallpox or purpuric fever
	Desjardins, 1996 : 50	Smallpox or purpuric fever or typhus
Purpuric fevers 1697-1698	Canada. Dept. of Agriculture, 1873 : 160-162 Desjardins, 1996 : 50	
Typhus 1699	Desjardins, 1996 : 51	
	Amorevieta-Gentil, 2010 : 131	Fevers
Influenza 1700	Desjardins, 1996 : 50	
	Trudel, 1968 : 241	
	Amorevieta-Gentil, 2010 : 131	
Smallpox 1702-1703	Bernard, 1994	
	Canada. Dept. of Agriculture, 1873 : 160-162	
	Desjardins, 1996 : 50	
	Trudel, 1968 : 241	
	Amorevieta-Gentil, 2010 : 131	
Yellow fever 1710	Trudel, 1968 : 241	Fevers
	Cadotte, 2013 : 5	
	Amorevieta-Gentil, 2010 : 131	1709-1711

Table 2 *Continued*

Measles 1714-1715	Mazan, 2011b, 2011a, 2012	
	(Mazan et collab., 2009)	
	Amorevieta-Gentil, 2010 : 131	
Fevers and Typhus 1717-1718	Trudel, 1968 : 241	Fevers 1718
	Amorevieta-Gentil, 2010 : 131	Fevers and typhus, 1717
Measles 1729-1730	Amorevieta-Gentil, 2010 : 131	Measles or smallpox
Smallpox 1733-1734	Trudel, 1968 : 241	1734 only
	Amorevieta-Gentil, 2010 : 131	
	Dépatie, 1988	
Typhus 1742-1744	Trudel, 1968 : 241	1743-1746
	Amorevieta-Gentil, 2010 : 131	Dearth of food, typhus and fevers
Typhus 1746-1750	Trudel, 1968 : 241	1750 only
	Amorevieta-Gentil, 2010 : 131	1748-1750
	Cadotte, 2013: 4	1746 for Mi'kmaq population
Smallpox 1755-1757	Trudel, 1968 : 241	1755 only
	Spaulding et Foster-Sanchez, 2019 : 2	
	(Amorevieta-Gentil, 2010 : 131)	Smallpox and typhus 1755-1756
Typhus 1756-1757	(Trudel, 1968 : 241)	
	(Amorevieta-Gentil, 2010 : 131)	Smallpox and typhus 1755-1756
Smallpox, Fevers and Dearth of food 1757-1758	(Amorevieta-Gentil, 2010 : 131)	
Typhus 1759	(Trudel, 1968 : 241)	
Smallpox 1769	(Amorevieta-Gentil, 2010 : 131)	
Smallpox 1775-1777	(Spaulding et Foster-Sanchez, 2019 : 2)	1775
	(Amorevieta-Gentil, 2010 : 131)	1775-1776
	(Bruckner et collab., 2018 : 4-5)	1777
Smallpox 1784	(Bruckner et collab., 2018 : 4-5)	
Smallpox 1798	(Bruckner et collab., 2018 : 4-5)	

Epidemics from 1800-1830 in Colonial Quebec have not yet received much attention from researchers. However, it is important to understand that the effects of an epidemic may last longer than the epidemic itself. Mazan (2012) studied the lingering effects of the 1714-1715 measles epidemic for 25 months after the acute phase of the epidemic. He found that children exposed to measles during toddlerhood had higher mortality after having survived the measles epidemic (Mazan, 2012). Furthermore, Clements and Hussey (2004) explain that morbidity and

mortality are intensified the year after an epidemic because the immune system remains compromised following an infection. This trend, however, may be countered by the opposite effect of selection. Palloni (1990 : 201) discusses the lower post-crisis levels of age-specific mortality due to the excess death of frail individuals. The most frail beings are more likely to die during a crisis, and their elimination selects a more robust post-crisis population (Palloni, 1990 : 201). This is the same selection concept discussed previously for the intrauterine deaths of boys, a selection mechanism which explains how the frailty of particular male cohorts is less apparent when they were especially selected pre-birth (Catalano et Bruckner, 2006). For the age groups where selection is at its strongest, the discrepancy between the crisis mortality and the post-crisis mortality will be significant for some time after the excess mortality (Palloni, 1990 : 201). This selection may also take place because of other factors such as consanguinity. or in the context of parental mortality (Pavard et collab., 2005 ; Willführ et Gagnon, 2013 : 200). An inverse selection process is also possible, although much less likely: if a crisis eliminates the least frail, then the remaining population could exhibit higher levels of mortality in the subsequent post-crisis period (Palloni, 1990 : 201-202). One way that the frail infants are partially protected from certain crisis episodes (ex. epidemics, famines) is through breastfeeding and labour (Niewiesk, 2014 : 1 ; Palloni, 1990 : 200).

1.2.2. Passive Immunity

Maternal antibodies are transmitted to the child from the mother in the third trimester with the placenta, allowing for antibodies in the bloodstream of the child, and postnatal transmission of antibodies through breastmilk, allowing for temporary antibodies in the gastro-intestinal tract of the child (Niewiesk, 2014 : 1). Both types of maternal antibodies are temporary. The maternal antibodies in the child's bloodstream wane with time. In fact, most children lose this passive immunity within 6 to 12 months, ergo they become more susceptible to disease and viruses after that age (Niewiesk, 2014 : 2). The maternal antibodies obtained with breastmilk also protect the child from disease, and its temporary effect depends on the mother's nursing practice and the age the child is weaned off breast milk. "Protection is better against some diseases (e.g., measles, rubella, tetanus) than others (e.g., polio, pertussis)" (Wodi et Morelli, 2015 : 2). In fact, measles is one of the most documented viruses for maternal antibodies.

It is uncertain if passive immunity would be different for consanguineous persons. A doctoral thesis studying descendants of Colonial Quebec demonstrated innovative results concerning immunity and consanguinity; it hinted at a lower frequency of white blood cells among consanguineous individuals, which may reduce the effectiveness of the immune system and increase susceptibility to infections (Gauvin, 2016 : 114). Gauvin (2016 : 111) found this link between immunity and consanguinity by studying 727 French Canadians with known genotypic information and identifying their stretches of homozygous genotypes. Therefore, the inbred population of Colonial Quebec may have a higher mortality than the non consanguineous population during epidemics due to their potentially weaker immune systems.

1.3. Concepts of Consanguinity

A father and a mother equally contribute to the genetic material of their offspring (Bateson et Mendel, 1902). In fact, human cells have two sets of chromosomes, each with the genetic material of a parent (Bateson et Mendel, 1902). A locus is a specific location on a chromosome and alleles, forms of genetic material, are found at these locations (Rédei, 2008a : 477). Theoretically, inbreeding coefficients, denoted F, designate "the probability that two alleles at a [randomly chosen] locus in an individual are identical by descent from a common ancestor" (Rédei, 2008b : 420). Furthermore, inbreeding coefficients are identical in value to parental coefficients of kinship, denoted φ, which describe the relatedness of spouses due to common ancestry (Jacquard, 1974 ; Malécot et Blaringhem, 1948). To avoid confusion, coefficients of relationship, another measure of relatedness between two individuals, denoted r, are twice that of kinship coefficients (Hamilton, 1964 : 3). These coefficients vary between zero and one, where the value of zero implies no known genetic relationship between the parents of the subject. The more the parents have common ancestors – and through which it is also possible to draw several genealogical paths forming loops – the higher the inbreeding coefficient is. If there is no other kinship relationship between the parents, in other words, if there is just one genealogical path and no other loops, the inbreeding coefficient of an individual based on their parents' relatedness is the following: F=0.25 if the parents are siblings, F=0.125 if the parents are half siblings, F=0.0625 if the parents are first cousins, F=0.03125 if the parents are second cousins and F=0.00390625 if the parents are third cousins (Gagnon, 2000). However, it is

unrealistic to assume no other kinship relationship between the parents or just one genealogical path, especially for French Canadians of Colonial Quebec. The Colonial Quebec population has more distant consanguinity than close inbreeding (Bouchard et De Braekeleer, 1991a ; Gauvin, 2016 : 41). The distinction between close and distant consanguinity is in the number of ascending generations being observed. An individual with a high inbreeding coefficient has ancestors who have lived in the colony for a (usually) substantial amount of time, especially when all ascending generations are considered. Inbreeding due to distant common ancestors will be referred to as distant consanguinity and inbreeding due to the grandparents or great grandparents being common ancestors will be referred to close consanguinity. Very close consanguinity insinuates a conscious decision from the spouses to marry kin whereas distant consanguinity reflects the demographic history of a region, as well as the marital market (Gagnon, 2000 ; Vézina et collab., 2004 : 72). In turn, even though the Church disapproved closely related marriages, the colony was semi-closed to migration; as a result distant consanguinity was inevitable (Bouchard et De Braekeleer, 1991a ; Mayer, 1993 ; Trudel, 1968 : 273-274).

Furthermore, close and distant consanguinity do not have the same length of identical haplotypes. In other words, the length of identical combinations of alleles due to a common ancestor on a chromosome is more likely to be greater when consanguinity is close compared to distant. This distinction in length is due to more meiosis, or cell division, when many generations separate the individual to their common ancestor. Therefore, if an individual is consanguineous due to distant common ancestors, then the more frequent cell divisions reduce the length of identical haplotypes, even if the inbreeding coefficient may be equal to that of an individual who is closely consanguineous, and who thus presents longer stretches of IBD in their genome. Consequences of consanguinity may differ in terms of those lengths, rather than a consanguinity binary, such that consanguinity effects may be predominantly observed in individuals with longer stretches of IBD (closely consanguineous) in their genome.

1.3.1. Consanguinity in Colonial Quebec

Catholic canon law prohibits marriage between related spouses and religious dispensations were necessary if partners were related up to the fourth degree (Trudel, 1968 :

273-274). Emond (1992 : 25) presents in his thesis the importance of studying consanguinity in Colonial Quebec by means of genealogical reconstitution because the marriage dispensations of the Church underestimate consanguineous unions in the colony. Gagnon (2000) describes that the marriage dispensations recorded allow researchers to identify general trends of consanguineous unions in a particular region. However, one may not confidently compare trends across regions since it is unknown to what extent the recordings of dispensations differed from one region to the next, and therefore from one priest to the next (Gagnon, 2000 : 70). This is why, in this study, consanguinity will be measured according to the genealogical data of the RPQA-IMPQ database by specifying the coefficient of inbreeding (F) attributable to each person recorded. This allows for distant consanguinity to also be measured. Since consanguinity was prohibited by the respected institution that is the Catholic Church, it is natural to expect low proportions of consanguineous individuals in Colonial Quebec. However, since both distant and close consanguinity are studied, some higher inbreeding coefficients may prevail in this research.

In fact, Vézina et collab. (2004) studied close inbreeding calculated with five ascending generations in Colonial Quebec. Their results show that the highest mean levels of the inbreeding coefficient per region are between 0.005 and 0.0075 (Îles-de-la-Madeleine, Gaspésie, Charlevoix), and the lowest mean levels of the indicator are below 0.0025 (Bas-Saint-Laurent, Saguenay-Lac-Saint-Jean, Côte-du-Sud, Québec City, Richelieu, Rive-Sud-de-Montréal, Île-de-Montréal, Laurentides, Témiscamingue). However, the study of consanguinity appreciably changes once distant consanguinity is included; Îles-de-la-Madeleine remains the region with the highest level of mean inbreeding coefficient, but now attains almost 0.0218. All regions have mean distant inbreeding coefficients of over 0.0025, though 16 out of the 26 regions studied have coefficients of over 0.005.

1.3.2. Causes of Consanguinity

Several studies identify factors that might contribute to a greater occurrence of consanguinity (Bener et collab., 2007 ; Bener et Mohammad, 2017 ; Bittles et Black, 2010a, 2010b ; Gagnon, 2000). Worldwide, countries have varying degrees of consanguinity and these rates may also differ according to historical periods. Research suggests several factors

associated with consanguinity, beginning with geography (rural/urban, isolation of the population) and marital markets.

Gagnon (2000 : 71) concludes in a chapter of his thesis that regional differences in consanguinity levels come from factors related to the availability of spouses in the marital market of Colonial Quebec. In the twentieth century, consanguineous marriages were mostly present in underpopulated rural areas of Western countries (Bittles, 2003 : 572). By linking both observations, we can assume that the regions with the greatest migratory flow would be those with the least consanguinity, because immigration would feed the marital market and thus, the genetic pool. Indeed, Emond suggests that immigration lowers the rates of consanguineous marriages and increases the genetic diversity of the region (1992 : 26). His thesis notes that parishes in rural Saguenay-Lac-Saint-Jean (SLSJ), a region of Quebec known for its high rates of consanguinity, have higher consanguinity coefficients than the urban areas of this city (Emond, 1992 : 27).

Next, social acceptance of behaviour and religion influence the occurrence of consanguinity. It is possible that there is a social transmission of consanguineous behaviour. Thus, consanguineous families normalize the behaviour and potentially encourage their children to also practice this type of union. For instance, in a study of urban and semi-urban areas of Qatar, Bener et collab. (2007) found high fertility among consanguineous couples. At the time of the study, about 50% of marriages were consanguineous, while the previous generation had only about 40% consanguineous unions (Bener et collab., 2007 : 264-265). However, according to the research of Robert et collab. (2009 : 674, 676) in Colonial Quebec, specifically among the descendants of women born in 1879 in the Saguenay-Lac-Saint-Jean region, the distant consanguinity coefficients are very high, but the coefficients calculated with only the last three generations are lower than expected if there was random pairings. This suggests non-random consanguinity. In other words, distant consanguinity is high but close consanguinity is low in this region of Quebec as these couples avoided closely consanguineous marriages potentially due to the strong social influence of the Catholic Church in the colony (Bouchard et De Braekeleer, 1991a ; Mayer, 1993).

In the past and, in part, in the present, marriages are consecrated by religious institutions, which may authorize or prohibit consanguineous marriages according to the proscribed criteria

(Bener et collab., 2007 : 262 ; Bittles, 2003 : 571-572 ; Bittles et Black, 2010b : 196 ; Bittles et Neel, 1994 : 561-562). This criteria is very different from one religion to the other and one region to another (Bittles et Black, 2010b : 196). This pattern also suggests the possibility of variation within regions. Bittles confirms that perceptions and social practices of consanguinity depend on religious and cultural prohibitions (2003 : 571). However, in this study, all individuals share the same religion. In addition, it is possible that officials in some parishes may be more tolerant of consanguineous marriages and may issue more dispensations if the marriage market of the region is restricted. Social realities form the norms of consanguineous marriages such that certain types of consanguineous unions are more acceptable than others under certain conditions (Bittles, 2003 : 571).

1.3.3. Expected Consequences of Consanguinity

In general and in the case of Colonial Quebec, the effects of consanguinity depend on living conditions and the relationships between genetic and socio-demographic factors (Robert et collab., 2009 : 673, 676). Most studies focus on the negative consequences of consanguinity, often describing the higher risk of death by disease (autosomal recessive disorders) in consanguineous individuals (Bittles et Black, 2010b ; Bittles et Neel, 1994 ; Boisvert et Mayer, 1994 ; Crimmins et Finch, 2006 ; Lyons et collab., 2009). Increased frequency of deleterious recessive genes being identical by descent (IBD) on the same loci has the potential to increase the risk of rare inheritable diseases that, in turn, increases the mortality risk (Rédei, 2008c : 420). The lower "fitness" caused by the deleterious recessive homozygous genes of consanguineous people is referred to as inbreeding depression (Rédei, 2008c : 420). Few researchers study the potential positive consequences to consanguinity (Bittles et Black, 2010a ; Denic et collab., 2008 ; Neel, 1962 ; Sanghvi, 1966). Both positive and negative consequences will be reviewed.

Negative Consequences

Research suggests there is a higher rate of infant and child mortality amongst consanguineous children compared to non consanguineous children (Bittles, 2003 : 573 ; Bittles et Black, 2010b : 201 ; Bittles et Neel, 1994 : 118-119 ; Stoltenberg et collab., 1999 : 522). In their study of 2.14 million individuals living in 15 countries of four different continents, Bittles

and Black (2010b : 200-201) determined that children from first cousins at ages six months of gestation to 12 years had a risk of death that was just 3.5% higher than that of children of non consanguineous couples for the entire period of observation. Emond (1992 : 77) concluded that infant mortality of consanguineous children in Saguenay-Lac-Saint-Jean, a specific region of Colonial Quebec, was significantly higher from that of non consanguineous individuals. However, there was no distinction in child mortality rates for 1- to 15-year-old consanguineous and non consanguineous individuals (Emond, 1992 : 77). This result may explain why Bittles and Black observed an excess mortality of smaller degree (3.5%).

The consequences of inbreeding often include diseases that seem to be more common among people whose parents are related. First, Bener et collab. (2007 : 265) list several diseases that are associated with consanguinity in a statistically significant way: cancer, blood disorders, intellectual disabilities, cardiovascular diseases, bronchial asthma, hypertension, hearing loss and diabetes. The diseases discussed so far apply mostly to adults. Inbred children are more vulnerable to childhood illnesses such as childhood cancers, learning disabilities and rare diseases (Bittles, 2003 : 573). Bittles and Black (2010b : 198, 203) also list other diseases common to consanguineous children such as congenital malformations, single gene-related diseases, intellectual or physical disabilities, higher cholesterol levels, and cardiovascular disadvantages. Indeed, Colonial Quebec's founder effect explains the higher prevalence of Leber's hereditary optic neuropathy (LHON) in Quebec today (Milot et collab., 2017), introduced by a *Fille du Roy* who carried a mutation that causes LHON in her mitochondrial DNA and which resulted in her descendants, especially males, losing their vision in early adulthood. The study by Milot et collab. (2017) also hinted at higher infant mortality levels for her male descendants during colonial times.

Research on consanguinity focuses mainly on the association between consanguinity and health or the association between consanguinity and infant mortality (Bener et collab., 2007 ; Bener et Mohammad, 2017 ; Bittles, 2003 ; Bittles et Black, 2010b ; Stoltenberg et collab., 1999). Inbreeding depression is also heavily studied in animal population. In fact, Keller and Waller (2002 : 236) summarize many studies on various animal species and most of them conclude negative survival, especially in the first month or year of life, as well as negative reproductive success. Further, the authors show examples of how the negative survival

consequences of inbreeding may appear during environmental crises, such as severe storms (Keller et Waller, 2002 : 235).

Studies on consanguinity in Quebec's historical populations focus mainly on the Saguenay-Lac-Saint-Jean region (SLSJ). For example, Robert et collab. (2009) compared the fertility of couples in consanguineous marriages and non consanguineous marriages in SLSJ; these authors found that the second half of the reproductive period of consanguineous couples were significantly less productive than the first half of their reproductive period compared to non consanguineous couples. Among all regions, Colonial Quebec is particularly relevant to the study of consanguineous persons due to the almost complete genealogical reconstitution available , the semi-closed environment of the territory prior to the nineteenth century and the increased potential to form unions among kin when the marital market did not permit otherwise (Dillon et collab., 2017 : 22 ; Gagnon, 2000 : 71).

Positive Consequences

Despite all these negative consequences, the consanguinity measure may also be associated with social and cultural advantages that may overturn the genetic disadvantages so commonly evoked (Bittles et Black, 2010a : 1783-1784). A social advantage possible in the context of this study is that consanguineous individuals may live closer to their extended family and therefore the blood-related parents may receive more caregiving help (Engelhardt et collab., 2019 : 653). This reduced geographic distance between consanguineous children and their grandmothers increases the grandmother effects (Engelhardt et collab., 2019 : 653). Advantageous grandmother effects include increased survival of the child, especially during infancy, and a younger age at first birth for the mother (Engelhardt et collab., 2019 : 652-653 ; Voland et Beise, 2002 : 435). In the context of consanguineous practices today, the social and cultural advantages include “enhanced female autonomy, more stable marital relationships, greater compatibility with in-laws, lower domestic violence, lower divorce rates, and the economic benefits of reduced dowry and the maintenance of any landholdings” (Bittles et Black, 2010a : 1784). These authors claim that the genetic disadvantages exceed social benefits in urban areas and “developed countries with better living and public health conditions”. Yet not every researcher of consanguinity agrees with this negative perspective on consanguinity.

Besides the biological disadvantages, there may be situational genetic benefits to inbreeding. First, Sanghvi explored the purging of harmful genes with generationally close inbreeding in India (1966 : 301). Denic et collab. (2008) explain this briefly through simulation. They discuss inbreeding (carriers of α+-thalassemia alleles) as a prevention to endemic malaria, and this reduced malaria mortality overshadowed the excess death due to autosomal recessive disorders, also referred to inbreeding diseases (Denic et collab., 2008 : 157). They claim inbreeding to be a "facilitator of adaptation" as the process of natural selection occurs faster in these populations, thus eliminating recessive lethal alleles (Denic et collab., 2008 : 156-157). Bittles and Black (2010b : 203) choose to approach the positive consequences of consanguinity historically; they discuss how consanguinity was inevitable when humans lived in small communities and how they must have resorted to inbreeding to survive, thus purging disadvantaged genes and increasing beneficial gene complexes (Jacquard et Reynès, 1968 : 644). However, in Colonial Quebec's case, consanguinity was not practiced enough to purge the negative genes from the population.

CHAPTER 2: OBJECTIVE, HYPOTHESES AND CONCEPTUAL FRAMEWORK

Chapter 2 offers the aim and hypotheses of this study and introduces a framework for how the research will be tested and presented.

2.1. Objective

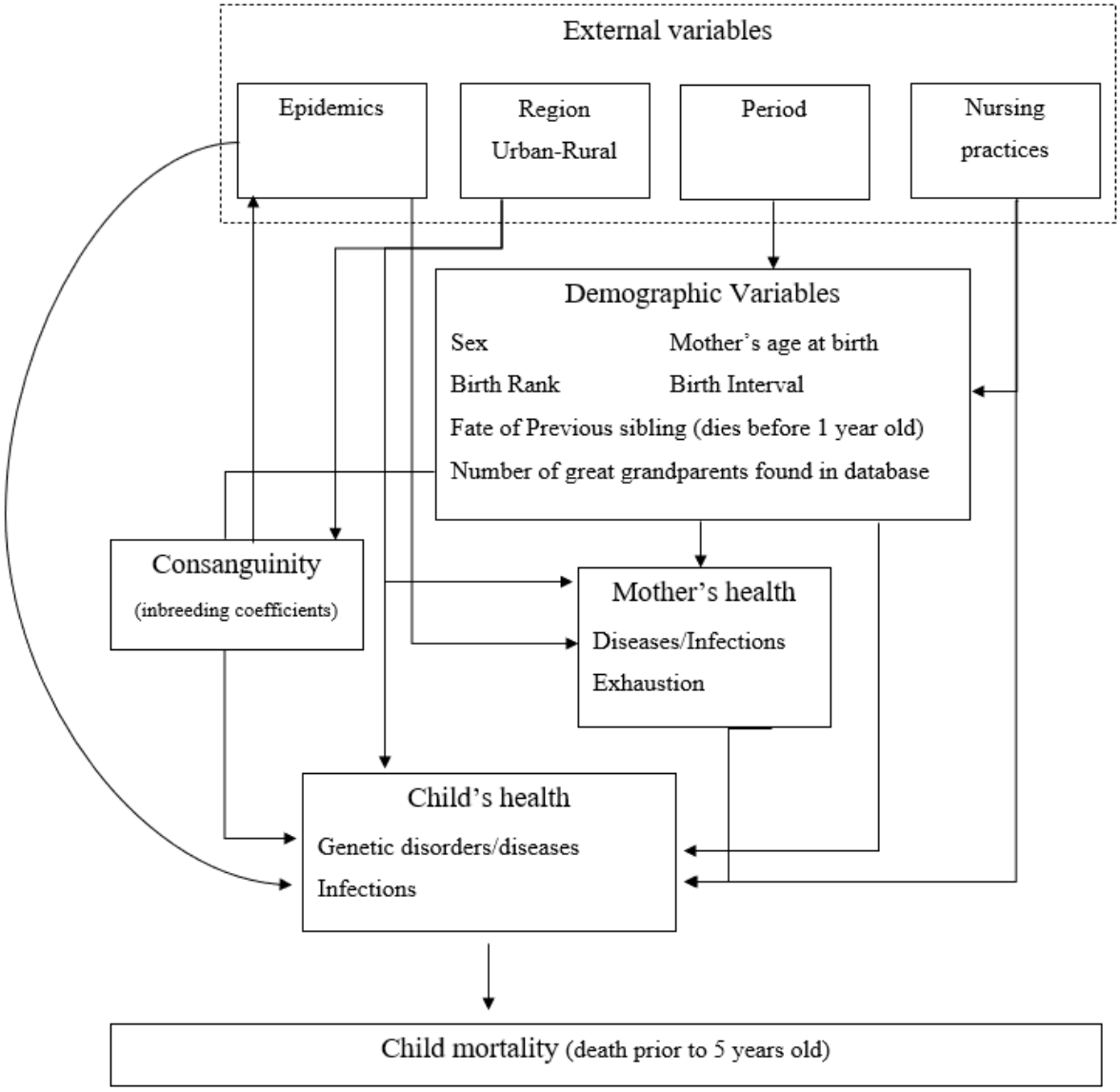
Considering the previous research on child mortality, epidemics and consanguinity, an element remains to be studied in Quebec and elsewhere: the comparative mortality of consanguineous children during epidemic periods. As previously mentioned, Gauvin's doctoral thesis suggests that consanguineous individuals have lower frequencies of white blood cells, which may reduce the effectiveness of their immune system (2016 : 114). As a result, consanguineous individuals may have been more susceptible to infections. In fact, consanguineous individuals' susceptibility to environmentally inflicted mortality was also mentioned in the literature review (Keller et Waller, 2002 : 235). Gauvin's hypothesis concerning the association of consanguinity with a compromised immune system motivates this research on the mortality risks during periods of epidemics in Quebec. Following this advance, consanguineous individuals of Quebec would also be more susceptible to mortality crises in the form of epidemics compared their non consanguineous counterparts. Studying this research question in a historical context allows us to conduct a sort of natural experiment, comparing the mortality risks of consanguineous and non consanguineous persons during epidemics in the absence of modern medical techniques, which would otherwise contribute to survival chances. The objective of this project is to evaluate the association between consanguinity and child survival up to five years of age during epidemic periods of Colonial Quebec between 1720 and 1830.

2.2. Conceptual Framework

The following conceptual framework was inspired by Lalou (1997 : 206). It shows the factors that influence child mortality as well as some known relationships between the factors. Foremost, on the left side of the figure, appears the variables of interest, epidemics and

consanguinity. The literature suggests that they negatively affect child's health, which in turn may cause child mortality. This project aims to explore how both variables interact with one another and how that interaction is expected to disfavour child survival. Epidemics may also affect the mother's health, which in turn affects the child's health and care which may lead to child mortality.

Figure 1. Conceptual Framework for Analysing Child Mortality, Consanguinity and Epidemics in Colonial Quebec



The geographic external variables, region and urban or rural residence, affect both consanguinity and child's health. As mentioned in the literature review, consanguinity in

Colonial Quebec was very much related to the marital market and partner availability. Therefore, levels of consanguinity are higher in certain rural regions compared to other regions, implying a geographic relation with consanguinity. Furthermore, the region also affects both the child's and the mother's health as urban areas usually had poorer hygienic and sanitary conditions compared to rural regions, with, for example, more polluted water from human wastes. As reviewed above, these unfavourable sanitary conditions caused diseases and deaths.

In addition, depending on the period of observation, the number of great grandparents found in the database changes. This difference is possible as the quality of the data collected may oscillate with time and the more time has passed, the more likely it may be to observe the number of generations necessary to have individuals with great grandparents in the database, which serve as a basis to define close consanguinity. Moreover, the mother's age at birth may also differ according to the period of observation as women married younger at the beginning of the Colony due to a sex ratio imbalance; by 1720, the sex ratio imbalance had become minimal (Charbonneau et collab., 2000 : 105, 109-114).

Another aspect of mother's behaviour, her nursing practices, affected both the mother's birth interval, and ergo, her overall health and level of exhaustion, as well as the child's health. The child's health is impacted negatively through the mother's health when birth intervals are short, but also positively through passive immunity and nursing practices. Both effects may play concurrently in the survival or death of the child. Nursing practices cannot be properly controlled as the data is not available. The best clue within the variables included in this study is the birth interval between the studied child and their previous sibling, as it may show some indication as to if the mother breastfeeds. However, this assumption is a stretch as breastfeeding practices may vary from child to child.

Next, early life mortality is dependent on the relationship between demographic variables and mother and child health. As mentioned in the beginning of the literature review, boys are more fragile than girls, especially as infants. Then, the mechanisms between mother's age at birth, birth rank and birth interval described in Chapter 1 affect the mother's overall health and exhaustion level, which in turn determines the child's health and care, as well as their odds at survival. In fact, an older mother has a higher risk of producing irregular gametes (sex cells) and having a high-risk pregnancy, therefore, genetically affecting her child's health and survival

(Maheu, 2001 : 112). The fate of the sibling preceding the studied child may suggest intrinsic or extrinsic conditions unfavourable to child survival if the sibling died prior to one year. Lastly, the number of great grandparents found in the database is related to close consanguinity, as it hints to whether there are enough generations in a family to potentially observe it. Further, this variable serves as a proxy for deep rooting in the colony. For instance, if a family has been in the colony for several generations, the child may benefit from lower mortality because of the better socioeconomic conditions of these families (ex. established land and community).

The factors of child mortality in Colonial Quebec described in the conceptual framework are indicative of the variables to be used in the study. These are further explained in the following chapter.

2.3. Hypotheses

The aim of this study is to evaluate the relationship between child mortality and consanguinity in epidemic periods of Colonial Quebec. A negative association between consanguinity and child survival is hypothesized: children of consanguineous couples will manifest a significantly higher mortality compared to children of non consanguineous couples, once controlling for demographic, geographical and temporal variables. This assumption is based on studies that conclude that infant mortality is higher among consanguineous individuals (Bittles, 2003 : 573 ; Bittles et Black, 2010b : 201 ; Stoltenberg et collab., 1999 : 522), as well as those that demonstrate the poorer health of consanguineous individuals (Bener et collab., 2007 : 265 ; Bittles et Black, 2010b : 198, 203). As mentioned in Chapter 1, there is evidence of a higher risk for consanguineous individuals versus non consanguineous individuals to have autosomal diseases, further weakening consanguineous persons (Bener et collab., 2007 : 264 ; Bener et Mohammad, 2017 : 318). Paired with Gauvin's advance on reduced immunity for consanguineous French Canadians (2016 : 114), the negative association between consanguinity and survival seems further probable as it would suggest that consanguineous children have a deadly disadvantage during epidemics due to their weak immune system (2016: 114). As a reminder to the reader, her thesis suggests lower frequencies of white blood cells in consanguineous individuals which reduces the efficiency of their immune systems, in turn, making these individuals more susceptible to infections (Gauvin, 2016: 114). Further, two

studies on two distinct animal species, Soay sheep *Ovis aries* and song sparrows *Melospiza melodia*, suggest "that inbreeding makes individuals more susceptible to environmentally inflicted mortality", specifically parasites for the sheep and severe storms for the birds (Keller et Waller, 2002 : 235). Thus, even if the susceptibility of being infected during an epidemic is equal for consanguineous and non consanguineous children, the ability to survive this infection would tip the balance in favour of non consanguineous children.

However, there is a bias to the measure of consanguinity with genealogy in this study; the variable also reflects the level of settlement of the family. An individual with a considerable inbreeding coefficient has ancestors that have lived in the colony for a (usually) substantial amount of time, especially when all ascending generations are considered. Thus, a positive association between consanguinity and child survival may reduce the negative association expected as these settled consanguineous individuals may manifest lower mortality risks because they benefit from the availability and presence of a more closely bound network. Genetically related couples are potentially more involved in their family network. This involvement and family closeness could benefit the children of related couples by increasing the levels of supervision and care provided to them. In other words, it is possible for these families to have more available guardians in proximity, even if they may have a reduced number of great grandparents because of common ancestry. Engelhardt et collab. (2019 : 653) described how the grandmother's protective effect on grandchildren survival increased when the distance between the grandmother and the grandchildren decreased. Voland and Beise (2002 : 435) added that this protective effect especially targeted infancy survival. Consanguineous individuals may have extended maternal and paternal families living in the same region. As a result, a positive, yet spurious association between consanguinity and childhood survival is also possible, which could even be compounded during periods of crisis mortality due to an epidemic. In fact, Mazan (2011a) discussed in his thesis that children of people who were settled in the colony since birth (non immigrants) had lower mortality during the measles epidemic of 1714-1715 compared to the children of immigrants. Thus, this extended family settlement measure included in the measure of consanguinity may imply reduced risk of mortality during epidemics. Moreover, a genetic advantage may also exist; the potentially identical alleles of a locus in consanguineous

individuals may be favourable genes that aid in survival. However, there is no way of knowing this.

Therefore, there are circumstances that can be negatively associated with the survival of consanguineous children (such as deleterious recessive genes) or positively associated with the survival of these individuals (due to the measure of settlement). If settlement is positively associated to survival, then the auspicious survival will be further seen in distantly consanguineous children. This situation underlines the need to begin the study with a descriptive analysis of the levels of consanguinity in Colonial Quebec using inbreeding coefficients (F) described as close and distant consanguinity, and calculated with the rich genealogical data available in the RPQA-IMPQ database.

CHAPTER 3: DATA AND METHODOLOGY

The literature review and hypotheses discussed in chapters 1 and 2 help to frame the context of the database used in this research. In this chapter, the longitudinal data is first presented. A description of the essential manipulation of the data with the use of a computerized program to determine the inbreeding coefficients in the genealogical database ensues. This is followed by a thorough review of variables to be studied and the methods of survival analysis employed in the project.

3.1. RPQA-IMPQ Database

From 1630 to 1830, the Catholic population naturally increased to become the majority population of Quebec; indeed, historic French Canada is first and foremost nourished symbolically and institutionally structured by the Catholic Church (Warren, 2007 : 22). As a result, births, marriages and deaths of this population have been systematically recorded in Catholic parish registers acting as civil records (Dillon et collab., 2017 ; Trudel, 1968 : 275). The reconstitution of those records is available in the *Registre de population du Québec ancien* (RPQA) and *Infrastructure intégrée des microdonnées historiques de la Population du Québec* (IMPQ) databases. The RPQA-IMPQ consists of all acts of baptisms, marriages and burials from Catholic parish registers in the province of Quebec from 1621 to 1830 (Dillon et collab., 2017 : 20-21). The RPQA-IMPQ database specifies the date and place of births, marriages and deaths, intergenerational and intragenerational links, as well as additional information, such as the density of families in a specific geographic location (Dillon et collab., 2017 : 28). The high fertility of Quebecers and limited immigration to and emigration from Quebec facilitates the matching of family members, especially when kinship networks are dense. The longitudinal data resulting from the reconstitution are extremely rich and dense; they allow clear identification of marriages between blood-related individuals and make up family files spread across several generations (Dillon et collab., 2017 : 27). As a result, the inbreeding coefficients, as well as the age at death of the individuals under study can be clearly identified.

The RPQA-IMPQ database has some limitations. First, it focuses on the Catholic population of Colonial Quebec. In other words, English Protestants and Aboriginals,

marginalized but common in the Quebec region, are typically not included in the register unless they married and had children with a Catholic inhabitant. Moreover, despite the scarcity of the following situations, it is important to note that in a case of an out-of-province migration or a religious conversion, the subjects under study are not followed (Dillon et collab., 2017 : 23). These unobserved departures are forms of attrition in this study. However, migrations between parishes are well documented in this database since it includes all parishes in Quebec between 1621 and 1830. Furthermore, some parish registers may have been lost or destroyed, though Catholic priests of Colonial Quebec transcribed in double the acts they performed (Desjardins, 1999 : 215 ; Dillon et collab., 2017 : 3, 1982 : 376). This double transcription limited the data lost, and therefore limited the number of individuals with incomplete life histories. Moreover, some individuals with partial histories are observed in the RPQA-IMPQ database; for example, some were followed from birth to marriage, while others are observed for the first time at marriage and, thereafter, at death. The prospective aspect of the data permits inference that may compensate for lost data from missing acts. For instance, if a baptismal act is missing, information from a marriage act or death act for this individual will include information pertinent to their date of birth as their age appears on this marriage or burial act. The systematic work of the Catholic priests, transformed into machine-readable data, permit the creation of the variables necessary for this research.

3.2. Variables

An essential variable for this research which is not directly included in the RPQA-IMPQ database is the inbreeding coefficient. A R (2021) package named GENLIB (Gauvin et collab., 2015) is readily available to analyse intricate genealogical data and calculate distant inbreeding coefficients for all the individuals in the imported database using four standard variables: the identification number of the individual in the database, their sex, the identification number of their mother and the one of their father (Gauvin et collab., 2015 : 162). The program identifies the multiple loops of kin relationships of each individual and outputs their inbreeding coefficient for the number of ascending generations chosen. The GENLIB package consists of many functions and allows other indicators to be calculated such as the genealogical depth of an imported database.

The RPQA-IMPQ database required some manipulation to contain only the four explicit numerical variables mentioned above: identification of the subject and their parents, as well as the subject's sex. All observations of the RPQA-IMPQ that did not have known sexes or parents required additional manipulation to be considered in the calculation of the inbreeding coefficient; a default value for the variable sex was temporarily given to these subjects and identification of their parent was put to 0. Nonetheless, the subjects who lack an identifiable mother and father will be excluded from the study, as their inbreeding coefficients calculated by genealogy are incomplete (see study population below). Even those with only one parent identified, the inbreeding coefficient is 0 (no consanguinity) simply due to lack of information. Consequently, subjects studied must at least have an identification number for each parent.

To analyse child mortality in Colonial Quebec, the risk of death at various age intervals from an individual's birth and their fifth anniversary is observed. The event variable of the survival analyses is death, or censorship, at an exact age, thus the difference in days between the date of death (or censorship, described below) and the date of birth. Therefore, only individuals with a date of birth of adequate quality are studied. This condition leads to the removal of individuals whose date of birth is informed by another researcher, deduced by the age of the individual in another act, or missing. Furthermore, certain inferences are made about the age at death or the age at censorship. Individuals that are present in an act after their fifth birthday or eventually married are assumed to have survived the childhood period of observation. The observation period from birth to five years old reflects the usual early childhood cut-off for mortality (Government of Canada, 2019, 2020 ; UNICEF, 2021). Often, death rates are described for infants first, and then for children one to four years, until the day prior to their fifth anniversary. Those without a date at death and no proof of childhood survival are censored at the date of the last act they are present in. Otherwise, the children who die during childhood have their exact age at death calculated with their date of birth and their date of death, which is in early childhood, so its quality is reasonable.

The independent variables of interest are consanguinity and epidemics. Consanguinity is separated into three categories: distant, close, and no consanguinity, based on the child's inbreeding coefficient. Close consanguinity is calculated with three ascending generations in this thesis. It describes the individuals whose parents are more closely related than double

second cousins (or first cousins once removed) (Bittles, 1994: 566). Further, both categories, distant and close consanguinity, are mutually exclusive. Thus, distantly consanguineous individuals are those with an inbreeding coefficient greater than 0.0039 when considering all possible ascending generations. The remaining people in the database are considered non consanguineous. Since close inbreeding is calculated with three ascending generations within this study, there are five possible inbreeding coefficients in the population with close consanguinity: F= 0, 0.03125, F=0.0625, F=0.125 and F=0.25. In the descriptive analyses (section 4.1.2), the inbreeding coefficients are further separated into categories of consanguinity: having a higher probability of IBD genes will be referenced as "strongly" consanguineous (F>0.03125), and having a lower probability of IBD genes, but still remaining consanguineous, will be referenced as "weakly" consanguineous (F=0.0039-0.03125). This classification is not used in the multivariate analyses as the sample sizes do not permit adequate results. The close inbreeding coefficients are usually higher and describe stronger consanguinity. Weakly consanguineous individuals are those with parents who are equivalent to third cousins up to double second cousins (or first cousins once removed) (Bittles and Black, 2010: 195; Bittles, 1994: 566). The "strong" versus "weak" categorisation used for distant consanguinity and close consanguinity in the descriptive analysis may suggest differential child mortality.

The time-varying epidemic variable is created by using the complete database, which includes individuals of all ages in the colony. The number of deaths per month and age group (0-1 years old, 1-4 years old, 4-7 years old, 7-15 years old, 15-49 years old and greater than 50 years old) are graphically represented for the east and west of the colony to identify peaks of mortality by age group in time (See Annex, Figures A1 and A2 which show the graphs for the first 4 age groups). These graphs permit the identification of mortality crises for the regions east and west of Trois-Rivières. The east and the west of the colony often experience epidemics within a few months of each other, reflecting the time that the epidemic takes to disperse. The age group distinction aids to detect epidemics like smallpox which, as mentioned in section 1.2.1, cycled about every seven years in the late eighteenth century and therefore affected mostly 0- to 7-year-olds (Bruckner et collab., 2018). The epidemics are identified by looking at the peaks and the date, in months, for the epidemics. The exact dates of the epidemics are required

in the survival analysis as the age at death is calculated in days. The dates of the epidemics are estimated from the first day of the starting month to the last day of the ending month in the peak. In Colonial Quebec, epidemics usually lasted a few months. To paint a complete picture of epidemics in Colonial Quebec, Table 3 identifies the dates of the peaks of the epidemics for the east and west, respectively, from 1670 to 1830, even if this study limits itself to 1720-1830. The epidemic peaks that are believed to be smallpox are in bold.

There is not a lot of information on epidemics from 1800 to 1830, so the smallpox epidemics are assumed using Bruckner et coll.'s seven-year cycle (2018). After the known 1798 smallpox epidemic, another smallpox epidemic is presumed in 1804 (6 years later), and in 1810 (6 years later). The small peaks between 1810 and 1819 are ignored because of the War of 1812 and the consequences of the Tambora eruption (1815-1816), which both incurred mortality in the colony for non epidemic reasons (McGuigan, 2016 ; Stommel et Stommel, 1983). Next, Barbeau (2007) claims that variola decimated the Native population in Colonial Québec from 1819 to 1821. Thus, the 1820-1821 epidemic is also considered smallpox. Further, in the West, the two epidemics that span a full year, 1803 to 1804 and 1820 to 1821, have peaks of mortality at the dates listed below.

Table 3. Epidemic Dates for the East and West of Colonial Quebec, 1670-1830

EAST				WEST			
Start date	End date	Start date	End date	Start date	End date	Start date	End date
01 Jan 1670	**31 Jan 1670**	**01 Feb 1770**	**31 May 1770**			**01 Mar 1770**	**30 June 1770**
01 Aug 1687	30 Sept 1687	01 Dec 1772	31 Dec 1772	01 Oct 1687	30 Nov 1687	01 Nov 1772	30 Nov 1772
01 Jan 1699	31 Jan 1699	**01 Jan 1777**	**30 Apr 1777**	01 Aug 1699	31 Aug 1699	**01 Dec 1776**	**31 Mar 1777**
01 Jan 1703	**31 Mar 1703**	**01 Jan 1784**	**31 May 1784**	**01 Mar 1703**	**31 May 1703**	**01 Feb 1784**	**31 May 1784**
01 Sept 1714	31 Oct 1714	01 Mar 1786	30 Apr 1786	01 Oct 1714	30 Nov 1714	01 May 1786	31 May 1786
01 Feb 1717	28 Feb 1717	**01 Mar 1791**	**31 May 1791**	01 Dec 1716	31 Dec 1716	**01 May 1791**	**31 May 1791**
		01 Mar 1795	30 Apr 1795	01 Oct 1727	31 Oct 1727		
01 Dec 1729	28 Feb 1730	**01 May 1798**	**31 May 1798**	01 Dec 1729	31 Mar 1730	**01 Mar 1797**	**30 Sept 1797**
01 May 1733	**30 June 1733**	**01 Jan 1802**	**31 Mar 1802**	**01 Feb 1733**	**30 Apr 1733**	**01 Mar 1803**	**31 Mar 1803**
01 Nov 1747	30 Nov 1747			01 Aug 1748	31 Aug 1748	**01 Mar 1804**	**30 Apr 1804**
01 Nov 1748	31 Dec 1748	**01 Dec 1808**	**28 Feb 1809**	01 Dec 1748	31 Dec 1748	**01 Apr 1810**	**31 Aug 1810**
01 Oct 1749	31 Dec 1749	**01 Dec 1820**	**28 Feb 1821**			**01 Jan 1820**	**31 Apr 1820**
01 Aug 1755	**30 Nov 1755**			**01 Oct 1755**	**31 Dec 1755**	**01 Aug 1820**	**30 Sept 1820**
01 Dec 1757	31 Jan 1758			**01 Sept 1758**	**30 Sept 1758**	**01 Mar 1821**	**31 May 1821**
01 Sept 1758	**30 Sept 1758**	01 July 1826	31 Aug 1826	01 Aug 1759	31 Aug 1759	01 June 1825	31 Aug 1825
01 Sept 1759	31 Dec 1759			01 Nov 1760	30 Nov 1760	01 July 1826	31 Aug 1826
01 Apr 1765	**31 Aug 1765**	**01 Jan 1830**	**30 Apr 1830**	**01 Apr 1765**	**30 June 1765**	**01 Apr 1830**	**31 May 1830**

Source: RPQA-IMPQ.

Note: The peaks of smallpox are bolded.

Past research and availability of the information in the database guide the choice of control variables. Following the variables studied in child mortality research in Colonial Quebec, a few broad categories of control variables will be included in this master's thesis: external variables, parental variables, and child's socio-demographic variables. The genetic variable of interest is related to parental variables in ways that affect all the children of two parents equally; the inbreeding coefficient is the same for each child of the same two parents. Other parental variables affect each child differently such as the mother's age at birth. No variable controls for their similar familial environment and each sibling will be studied independently. Variables related to the child include their sex, birth rank, birth interval and the survival until 1 year old of the sibling that precedes the index child in rank. Note that only live births are documented, so if the child's mother had a miscarriage prior the child observed, it will not be considered in the study. However, it may be partly considered with the birth interval variable. Further, both the birth interval and birth rank are calculated with reference to the mother of the individual of interest.

Next, the number of grandparents is a proxy for extended family settlement or "rooting" within the colony. As mentioned in Chapter 2, consanguinity is closely related to settlement in this study and this "rooting" may play in a different direction than consanguinity on mortality for socioeconomic reasons. Consequently, a categorical variable describing the number of great grandparents identified in the database (0-2, 3-5, 6-8), will act as a proxy to settlement to partially control for this effect. Lastly are the external variables which describe geographic and environmental conditions of each individual in time. As epidemics, our time-varying variable of interest, arise in specific regions and different periods, models are controlled by region and a 15- to 20-year period of observation. The region variable includes thirteen delimited regions of Colonial Quebec which are specified as either rural or urban, and one other or unknown region value. The regions follow the ten delimitated regions set by Gagnon (2000) in his thesis with the region surrounding Trois-Rivières separated into the north and the south, and further distinction of urban Montreal and urban Quebec from their rural regions. The variable assumes that the region of birth is the region the individual lived in their early childhood if they did not die. If they died during their childhood, region of death is used as the region of the individual. The region used is only different from the region of birth in 8,455 cases, so, less than 1.5% of

the study population. Ergo, the hypothesis that a child who survived early childhood still lives in their region of birth is fair, and the precision of the region of death permits for the proper designation of the epidemic dates in the case where the child died in an epidemic crisis.

The variables of interest and the control variables may have confounding interactions in the study of child mortality and may mute the intensity of a result if their respective effects work in opposite directions. For example, as mentioned in the literature review, the proportion of consanguinity in the population should be lower in urban areas versus rural areas as the marital market should be more limited in rural areas due to geographic isolation. On the other hand, urban areas of Colonial Quebec have higher infant mortality rates than rural areas due to their less favourable sanitary conditions (Amorevieta-Gentil, 2010 : 145), but rural areas tend to have higher consanguinity coefficients (Emond, 1992 : 27). Therefore, the relative risk of consanguinity may be muted or enhanced, depending on the effect of consanguinity on mortality, when comparing urban and rural regions in a poorly controlled model. Hence, it is essential to control for region and period in every model. This control should be sufficient as most external factors increasing the risk of child mortality, for instance, urban sanitary conditions, affect the consanguineous and non consanguineous populations similarly.

3.3. Methodology

This study uses survival analysis since the longitudinal data contains censoring and the epidemic time-varying variable. Furthermore, the transition between life and death is a process in time; this acts as a third motive for using survival analysis, which is also referred to as event history analysis in demography, and social sciences in general (Allison, 2010). Indeed, this type of analysis allows us to study the risk of an individual experiencing an event of interest: child mortality, in this case. Each individual is identified as having died in distinct episodes of their early childhood or being censored; if a child dies prior to their fifth anniversary, they will be identified as experiencing the event at their precise age at death (measured in days); however, if a child has survived the period of observation, they will be right censored at five years old. This right censoring is also known as Type I censoring (Allison, 2010 : 11). The case of random censoring occurs when attrition occurs (Allison, 2010 : 12) which is whenever an individual is lost in the database in our period of observation and there is therefore no proof that they survived

early childhood. The consideration for censoring allowed in survival analysis is an essential reason for its use in this study of childhood mortality of consanguineous individuals in Colonial Quebec.

Foremost, descriptive statistics will be studied using the Kaplan-Meier method. Explicitly, the survival and the hazard estimates of the study population plotted in function of age allows for preliminary conclusions. Individuals are removed from observation at their age at death, or their age at censorship. In fact, the survival function presents the proportion of children at risk of dying, given that they have not experienced the event of interest (death) in function of time (in days), where the origin (time 0) is the individual's birth. Furthermore, multiple survival curves may be superimposed in terms of consanguinity levels or sex to identify potential relationships between age-specific survival and status. Similarly, hazard ratios are plotted in function of time (in days) demonstrating the risk of death at specific ages. The Kaplan-Meier graphs allow us to test the proportionality hypothesis essential to the Cox survival analysis which will also be presented in Chapter 4.

Cox models allow the researcher to concentrate on the effects of the independent variables without specifying the baseline hazard function. This function is often unknown which explains the prevalence of Cox models in research. Cox models are proportional risk models. In other words, the proportional hazards hypothesis is necessary to use this model. Proportionality may be verified graphically by comparing log-negative-log functions of survival (see Kaplan-Meier survival curves in section 4.1.3). For the hypothesis to be respected, the curves must be subjectively parallel. Proportionality may also be verified with Schoenfeld residuals, a hypothesis test examining if the log hazard-ratio function is constant over time. STATA easily tests if the log hazard-ratio function is constant over time (H0) using Schoenfeld residuals. If the null hypothesis is rejected, the proportional-hazards assumption is not respected in the model. The detailed results are not presented in this thesis, but if a variable or a model resulted in a p-value lower than 5%, ergo rejecting the null hypothesis of a zero slope with 95% confidence, it is mentioned in the results and further discussed. For a large sample size, the test using Schoenfeld residuals easily rejects the null hypothesis. The log-negative-log curves are further observed in those cases to graphically identify the extent of the violation of the proportional hazards assumption. If the violation is minor, "the [hazard ratio (HR)] varies over

time and changes in magnitude but not direction", then "the overall HR can be interpreted as an average HR over time" (Barraclough et collab., 2011 : 981). If the proportional hazards hypothesis is in major violation for a variable, stratification permits the model to be controlled for this variable. In this case, all individuals included in each stratum are said to be placed on their own specific baseline hazard, against which the role of other independent variables included in the model are assessed. However, in such cases, one cannot identify the effect that the stratified variable has on the dependant variable.

The Kaplan-Meier curves in the next chapter suggest how some variables affect child mortality differently according to the age observed. Further, the proportional hazards hypothesis cannot be respected from birth up until five years old. Consequently, separate models for distinct age groups are analysed. The age intervals chosen for the Cox models are partially based on infant and child mortality categories. Before age 1, they are the early neonatal (0-6 days) deaths, excluding birth day deaths to avoid stillbirths, the late neonatal (7-27 days) deaths, and the post neonatal (28 days-1 year) deaths, which are themselves separated into three intervals: 28 days to 3 months, 3 to 6 months and 6 to 12 months). The remainder of under-five mortality intervals are annual (1-2 years; 2-3 years; 3-4 years; 4-5 years).

The effects of consanguinity and epidemics on infant and child mortality are analysed for each age interval described above, revealing the differential longitudinal effects of these variables. All the models control for time periods of at most two decades, regions and urban status. Some models are further controlled for parental and child variables previously discussed. Thus, the Cox models are stratified by geographic and temporal variables and models are separated by sex and age intervals. Each stratum, which separate the subjects into disjoint groups, has its own uncalculated baseline hazard, but the variables added to the model, which must have proportional risks for individuals of the same strata, have effects independent of the strata. In other words, the coefficients of the hazard function will be the same for each stratum.

The dataset is split into episodes with epidemics and without epidemics. This way, each observation for an individual can indicate the time-varying variable's dichotomous value of epidemic. Further, the dataset is split into the time intervals of maximum one year chosen to ensure proportionality of the variables of interest: early neonatal, late neonatal, post-neonatal, remaining infancy, toddlerhood and remaining early childhood. An individual therefore appears

in every episode split until their death. In other words, if the individual dies or is censored prior to 5 years, no additional observations are created for them after their censorship or death. Lastly, models are evaluated with the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), which are criterions that promote parsimony as compared to likelihood ratios.

3.4. Study Population

The Catholic population of colonial Quebec from the seventeenth to the early nineteenth century is our study population. This population includes single-birth individuals who were born in Quebec (non-immigrants) between January 1st, 1720 and December 31st, 1830 and who have been successfully linked to their mothers and fathers, allowing the calculation of consanguinity. The study population cannot include the individuals born after 1830 as their linkage to their ancestry is not yet complete. Consequently, December 31st, 1830 is the censorship date for everyone still in observation at that date. If the start date and the end date of the survival analysis could not be identified, ergo the birth date and the end of observation date, the individuals were removed from the analysis. This selection permits 783,145 boys and girls to be studied among the 1,818,295 individuals mentioned in the complete RPQA-IMPQ database. However, survival analysis in STATA excludes all subjects that die or are censored on their date of birth (StataCorp., 2021 : 441), therefore reducing the number of individuals observed in our survival analysis models to 610,412. This exclusion allows the omission of very early deaths which may be stillbirths.

CHAPTER 4: RESULTS

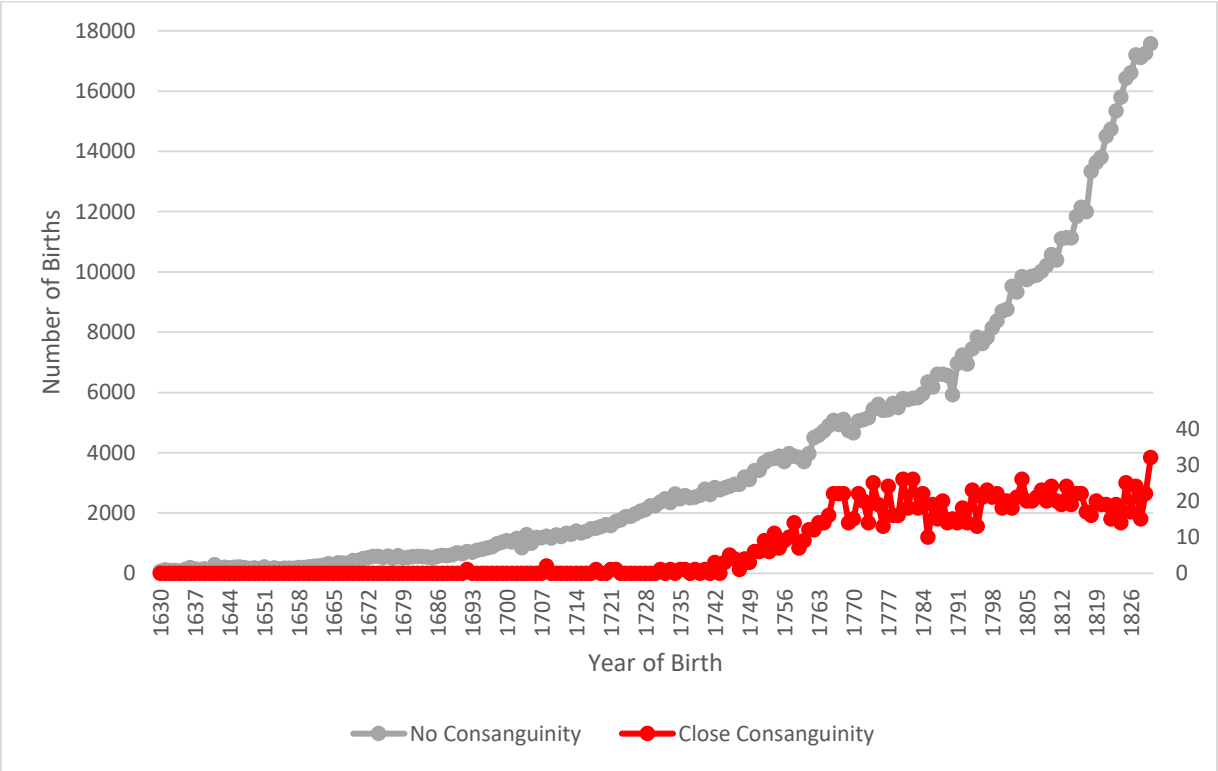
In this chapter, graphs presenting levels of consanguinity over time precede the descriptive statistics which present the frequency and percent distributions of the population at risk across all variables included in the analysis. Then, Kaplan-Meier survival graphs show potential relationships between specific characteristics and childhood mortality. These also test whether the proportionality hypothesis necessary in Cox survival analyses is violated for each key variable. These relationships will be further observed using Cox survival analysis later in the chapter. The stratified Cox models for under-five mortality in Colonial Québec are separated by sex and age groups: from birth to six days (early neonatal period), seven days to 27 days (late neonatal period), 28 days to three months (post-neonatal period), from three months to six months, six months to one year, and every year until five years, the end of observation.

4.1. Descriptive Analysis

4.1.1. Consanguinity and Time

Quebec's history is characterised by its semi-isolated nature. Having been colonized by a distinct group of founders and having had limited immigration, distant consanguinity is related to time within the colony. The next graphs compare the growth of the number of consanguineous ($F \geq 0.0039$) and non consanguineous individuals in the colony by their year of birth. Figure 2 shows the more stable and constant number of closely consanguineous individuals over time, while Figure 3 shows a growing number of distantly consanguineous individuals over time.

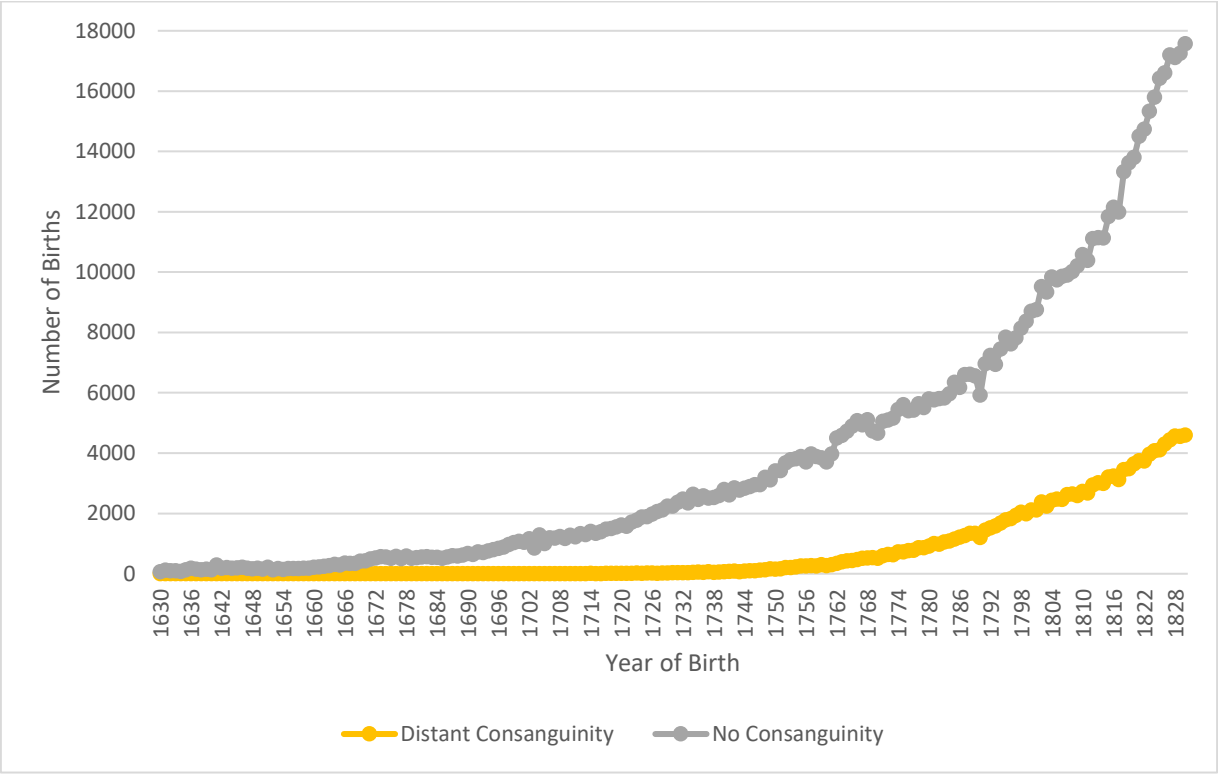
Figure 2. Number of Closely Consanguineous and Non Consanguineous Individuals by Year of Birth, Colonial Quebec, 1630-1830



Source: RPQA-IMPQ data
Denominator: The entire colony
Note: Close consanguinity - considering three ascending generations - follows the secondary axis on the right.

In the 1600's, when the colony experienced very few births, close consanguinity cannot be found in the study population. The first closely consanguineous baby is born in 1692, and the next two closely consanguineous babies are born in 1708. Closely consanguineous babies make up about 0.05% of the births per year until 1740. Beginning in the 1740's until 1830, consanguineous babies make up between 0.1% and 0.4% of the births per year (Figure 2). The proportion of closely consanguineous babies remains very low and relatively stable throughout the observation period, confirming the aversion of the Christian population to closely consanguineous marriages.

Figure 3. Number of Distantly Consanguineous[1] and Non Consanguineous Individuals by Year of Birth, Colonial Quebec, 1630-1830



Source: RPQA-IMPQ data
Denominator: The entire colony
[1] excludes close consanguinity

By the mid 1790's, consanguineous babies, with distant consanguinity, make up about 20% of the births per year (Figure 3). This is an important rise considering that they only made up 6% of the births in the 1750's and less than 1% prior to 1725. This trend illustrates the relationship between time and distant consanguinity in the semi-isolated colony. As decades passed, it became more likely that parents share similar ancestors in their genealogies. Nevertheless, these ancestors are often so many generations distant into the genealogies that the accumulated inbreeding coefficients for their children is below the cut-off (F=0.39%) which deems them non consanguineous.

Table 4 presents the number of distant and close consanguineous individuals in the study population by period and excluding unknown sexes. The increase in consanguinity with time is indubitable, especially for exclusively distant consanguinity which continues to increase in frequency even when the interval of time is reduced.

Table 4. Period and Consanguinity, Number of Individuals, Colonial Quebec, 1630-1830

Period	Distant consanguinity [a]	Close consanguinity
1630-1700	24	1
1700-1720	78	3
1720-1740	561	7
1740-1760	2,950	101
1760-1780	10,393	319
1780-1800	26,721	341
1800-1815	39,161	307
1816-1830	55,490	268
Total	135,378	1,347

Source: RPQA-IMPQ. [a] excludes close consanguinity.
Denominator: Study population including children who died on their date of birth.

Table 4 shows that it is statistically unadvisable to include the seventeenth century as well as the first two decades of the eighteenth century in this analysis because consanguinity levels are too low during that period, especially for close consanguinity, which only has one case in the seventeenth century.

4.1.2. Frequencies and Percentages

Table 5 presents the distribution of individual characteristics in frequencies and percentages for every category of the variables in this study, including period, region, sex, distant consanguinity, close consanguinity, living an epidemic (during the time at risk), number of known great grandparents in the database, birth interval of the mother at individual's birth, fate of previous child (i.e., whether the previous sibling died prior to age 1), birth rank, and mother's age at childbirth. The table also presents the descriptive statistics for the interaction terms between epidemics and consanguinity included later in multivariate analyses.

Table 5. Descriptive statistics of the Study Population – Temporal, Geographic and Demographic Variables, Colonial Quebec, 1720-1830

Variables	Frequency	Percentage
Region		
Nord de Montréal	108,336	17.75%
Ile de Montréal (urban)	31,949	5.23%
Ile de Montréal (rural)	47,371	7.76%
Sud de Montréal	148,148	24.27%
Trois-Rivières (north)	32,754	5.37%
Trois-Rivières (south)	21,701	3.56%
Portneuf	19,799	3.24%
Lotbinière, Beauce, Lévis et Bellechasse	58,124	9.52%
Ville de Québec (urban)	31,517	5.16%
Région de Québec (rural)	13,798	2.26%
Île d'Orléans	9,988	1.64%
Beaupré & Charlevoix	22,837	3.74%
Bas St-Laurent & Beauce	63,038	10.33%
Unknown or Other	1,052	0.17%
Consanguinity		
Distant consanguinity		
Strong Consang.	3,604	0.60%
Weak Consang.	95,998	15.73%
Close consanguinity		
Strong Consang.	1,005	0.16%
Weak Consang.	76	0.01%
No consanguinity		
No Consang.	509,729	83.51%
Epidemic during time at risk [1]		
Epidemic	347,777	56.97%
No Epidemic	262,635	43.03%
Interaction: Epidemics and Consanguinity		
Epidemic*Distant Consang.	55,448	9.08%
Epidemic*Close Consang.	615	0.10%
Epidemic*No Consang.	291,714	47.79%
No Epidemic*Distant Consang.	44,154	7.23%
No Epidemic*Close Consang.	466	0.08%
No Epidemic*No Consang.	218,015	35.72%
Fate of previous sibling [2]		
Dies prior to 1 year	192,395	31.52%
Survives 1st year	308,342	50.51%
Unknown	109,675	17.97%
Interaction: Death of previous sibling prior to 1 year old and Consanguinity level		
Distant Consanguinity	31,311	31.44%
Close Consanguinity	363	33.58%
No Consanguinity	160,721	31.53%
Total	610,412	

Variables	Frequency	Percentage
Period		
1720-1740	33,794	5.54%
1740-1760	56,155	9.20%
1760-1780	91,067	14.92%
1780-1800	136,855	22.42%
1800-1815	144,647	23.70%
1816-1830	147,894	24.23%
East/West		
West	390,259	63.93%
East	219,101	35.89%
Unknown	1,052	0.17%
Sex		
Male	310,961	50.94%
Female	299,451	49.06%
Unknown	Removed	Removed
Number of great grandparents identified		
0	4,043	0.66%
1	18	0.00%
2	1,607	0.26%
3	146	0.02%
4	36,562	5.99%
5	123	0.02%
6	17,561	2.88%
7	2,455	0.40%
8	547,897	89.76%
Birth interval		
<15 months	99,877	16.36%
15-20 months	165,556	27.12%
21-29 months	179,481	29.40%
30-35 months	37,328	6.12%
>35 months	41,648	6.82%
N/A or Unknown	86,522	14.17%
Mother's age at birth		
<20	33,570	5.50%
20-24	131,628	21.56%
25-29	151,816	24.87%
30-34	129,113	21.15%
35-39	93,007	15.24%
40-44	41,710	6.83%
45+	4,022	0.66%
Unknown	25,546	4.19%
Total	610,412	

Table 5 *Continued*

Rank of birth (according to mother)					
Variables	Frequency	Percentage	Variables	Frequency	Percentage
1	86,522	14.17%	12	15,484	2.54%
2	77,694	12.73%	13	10,877	1.78%
3	70,109	11.49%	14	7,250	1.19%
4	63,074	10.33%	15	4,698	0.77%
5	56,618	9.28%	16	2,831	0.46%
6	50,299	8.24%	17	1,611	0.26%
7	44,136	7.23%	18	880	0.14%
8	38,160	6.25%	19	426	0.07%
9	32,211	5.28%	20	175	0.03%
10	26,414	4.33%	21	88	0.01%
11	20,792	3.41%	22-25	63	0.01%
			Total	610,412	

Source: RPQA-IMPQ.
Denominator: Study population
[1]In the Kaplan-Meier and Cox analyses, a time-varying variable is used and therefore it is not equivalent to the variable presented in this table.
[2]The previous sibling can have a birthdate of any quality.

The most populated regions of the study are in and around Montreal, in the West (64%). The eastern and western regions were not subject to the same environmental factors and did not experience the same mortality rates (Mazan, 2011a ; Mazan et collab., 2009). In fact, the time-varying epidemic variable is calculated according to this geographic delineation. The differential mortality was not just observable between eastern and western regions, but also north and south of the St-Lawrence River (Gagnon, 2012). Further, the urban areas had differential mortality too, which is why the are considered in the delimitation of regions.

Next, the population grows significantly as the colony ages. About half of the study population are children present in the last 30 years of observation (1800-1830) and 57% of the children experienced at least one epidemic during the 5-year observation.

As mentioned in Chapter 1, males' and females' mortality patterns differ. There is a slight male majority (51%) in the analysed population comprised of males and females, corresponding to the typical sex ratio that is 105 males per 100 females (51.2%) (Cavalli-Sforza et Bodmer, 1971 ; Henry et Blum, 1988 ; Pressat, 1983 ; Sieff et collab., 1990). There is also a male majority in the general population comprised of males, females, and unknown sexes (not shown). However, Table 5 describes statistics excluding children of unknown sex as they follow

a different mortality pattern (see section 3.3). If such children were included, about 1% of the 1630-1830 population comprised of 830,860 individuals would have an unknown sex.

In Table 5, very few individuals (<1%) with close consanguinity are observed within the two centuries of study. This is expected as the Catholic Church proscribed marriage between kin. However, almost 16% of the population studied have distantly consanguineous coefficients considered weakly consanguineous, which implies a cumulated consanguinity equivalent to, at least, the inbreeding coefficient of third cousins. When observing all ascending generations (close and distant consanguinity), 16.5% of the population is consanguineous, whereas when observing only three ascending generations (close consanguinity), less than 0.2% of the population is consanguineous. This distinction illustrates the importance of observing a larger number of ascending generations to determine consanguinity, especially in a colony that experienced limited immigration.

When observing close consanguinity with three ascending generations Table 5 also shows the expected greater proportion of strongly consanguineous individuals (F>3.125%), individuals with a higher probability of having genes identical be descent, compared to the weakly consanguineous individuals (F≤3.0125%), individuals with a lower percentage of having genes identical by descent, but are still considered consanguineous. Naturally, the probability of sharing identical genes is greater when individuals are close to each other in a genealogical pedigree. Thus, distant consanguinity is expected to have mostly weakly consanguineous individuals, and it does by holding 15.7% of the study population. Almost all of the consanguineous individuals have distant consanguinity of low levels.

Knowing that the prevalence of close consanguinity in Colonial Quebec is low, let's observe the number of individuals in each level of consanguinity per sex as it may later aid in the understanding of the Cox models which are separated by sex.

Table 6. Number of Males and Females Analysed in the First Period of Observation per Consanguinity Level

	Male	**Female**
Distant Consanguinity	50,801	48,801
Close Consanguinity	545	536
No Consanguinity	259,615	250,114

Source: RPQA-IMPQ.
Denominator: Study population

Considering the lower sample size of closely consanguineous children, some hazard ratios will have to be very high to be minimally significant ($p<0.05$). For example, the 95% confidence interval for the stratified Cox model of the first age group (see section 4.2.1), when the sample size of close consanguinity is at its highest, is between 1.12 and 2.55 for boys and between 1.38 and 3.14 for girls when considering only consanguinity as a variable. As a comparison, the 95% confidence interval of the same stratified Cox model for distantly consanguineous newborns is 1.00 and 1.14 for boys and 0.97 and 1.12 for girls. There is obviously a considerable variability in the 95% confidence interval due to the small sample size of closely consanguineous children and this variability will make it "harder" to have significant effects.

Further, considering that consanguinity and epidemics are key subjects of this thesis, we verify if both variables seem independent. In fact, for close consanguinity and distant consanguinity, the levels of consanguinity are similarly distributed whether individuals experienced an epidemic during the time at risk or not. For instance, the expected number of individuals to be closely consanguineous and live during an epidemic is 616 (1,081*57%). The actual frequency of this group is 615, which is equivalent. The complete example below shows that consanguineous children did not particularly experience epidemics more or less than non consanguineous children. The odds of being at risk of dying in an epidemic should therefore be similar for all groups.

Table 7. Expected versus Actual Number of Individuals per Level of Consanguinity and Having Lived during an Epidemic

Interaction: Epidemics and Consanguinity			
	Actual frequency	*Expected Frequency*	*Difference*
Epidemic*Distantly Consanguineous	55,448	56,773	-1,325
Epidemic*Close Consanguineous	615	616	1
Epidemic* non consanguineous	291,714	290,545	1,169

Source: RPQA-IMPQ.
Denominator: Study population

Next, most of the study population has complete genealogies for three ascending generations since 90% of individuals have eight identified great grandparents, which is the maximum amount possible. The way the variable is calculated, the closely consanguineous

individuals with complete genealogies at three ascending generations still show a full number of great grandparents identified, even if, with their common ancestor, they have less than eight distinct great grandparents. Basically, the number of great grandparents were calculated without forcing them to be distinct individuals. 36,562 individuals (6%) have just four great grandparents. Some of these may be children with one settled parent, usually the mother, who would have a complete genealogy, and one immigrant parent, who has no genealogy. However, this category is not limited to this scenario. It is possible that two settled parents simply have partial genealogies.

Further, Amorevieta-Gentil (2010) identified distinct categories for birth intervals in her doctoral thesis pertaining to Colonial Quebec and these are the categories observed in Table 5 with one additional category: not applicable or unknown. The majority (56.5%) of birth intervals observed are between 15 and 29 months, which is the birth interval that incurs the least infant mortality (Amorevieta-Gentil, 2010). Birth intervals of some individuals of the study population (14%) cannot be identified as their older sibling's date of birth is unknown or they are their mother's first-born child. The 16% of children with a birth interval less than 15 months may be at risk of death because the mother's body may not have had enough time to recover from the previous birth (Amorevieta-Gentil, 2010). Further, the death of the previous sibling prior to his first birthday may hint at harsh living conditions for this family and increased risk of death for the individual at study. The previous sibling of children with close consanguinity has a higher percentage of death prior to one (33.5%) compared to distant consanguinity and no consanguinity (31.4% and 31.5%, respectively). Thus, harmful genetic conditions are also accounted for in this variable.

Quebec mothers from our study period had many children. In fact, each rank from 1 to 6 contains 8%-15% of the study population, amounting to 66% of the study population in these common ranks. Only 15% of the study population are from ranks greater than or equal to 10. Rank is related to the mother's age as the older the mother becomes, the more likely she births children of higher rank (Amorevieta-Gentil, 2010). Almost 68% (412,557) of the mothers of the study population gave birth to their child between 20 and 35 years old, the most beneficial age group for labour to lessen risks of infant mortality. Another 15% of them gave birth between 35 and 40 years old. Age at birth below 20 years old or above 40 years old are less common as they

each hold only 5-7% of the population, but they are also riskier pregnancies. There are 25,546 individuals in our study population (4%) whose mother's age at birth cannot be determined.

The next table observes the distribution of the study population by consanguinity level and sex. Though the study ultimately excludes them, unknown sexes and children censored or deceased on their date of birth are included in the following table to show the higher percentage of strongly consanguineous individuals in the unknown sex category.

Table 8. Level of Consanguinity by Sex and Number of Ascending Generations Considered, Frequencies and Column Percentages, Colonial Quebec, 1720-1830

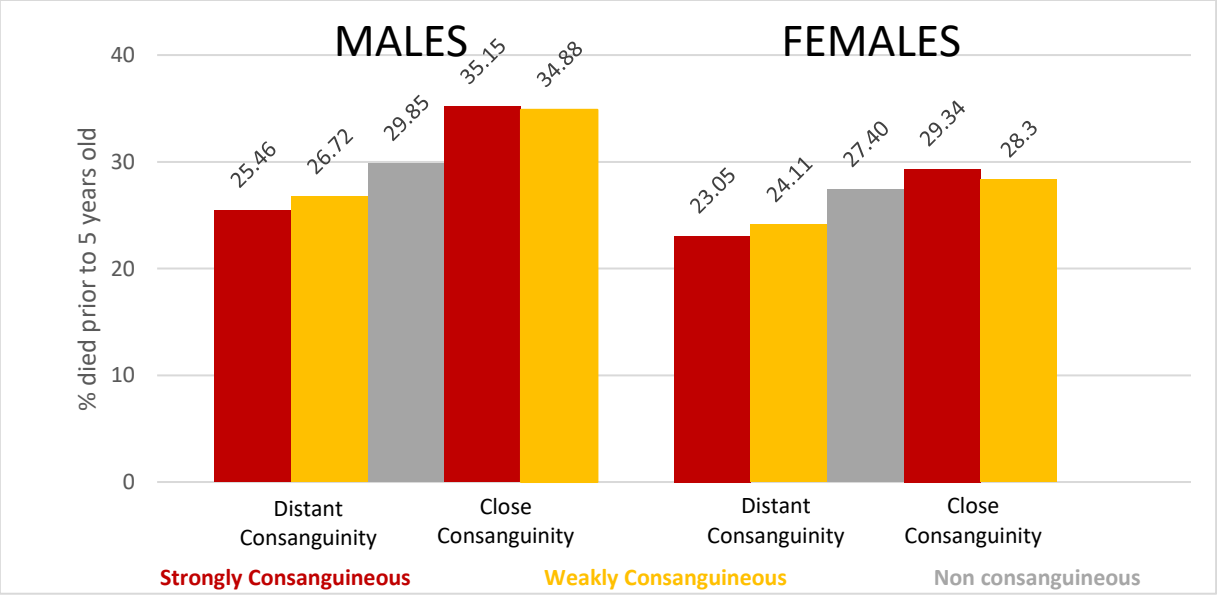
	Considering >3 Ascending Generations (Distant Consanguinity)				Considering 3 Ascending Generations (Close Consanguinity)			
Level of Consanguinity (inbreeding coefficients)	**Male** Frequency %	**Female** Frequency %	**Unknown** Frequency %	**Total** Frequency %	**Male** Frequency %	**Female** Frequency %	**Unknown** Frequency %	**Total** Frequency %
Strongly Consanguineous (F=0.03125-0.25)	2,464 0.61%	2,270 0.60%	68 0.65%	**4,802** 0.61%	623 0.15%	593 0.16%	25 0.24%	**1,241**[A] 0.16%
Weakly Consanguineous (F=0.0039-0.03125)	66,169 16.46%	62,607 16.49%	1,698 16.15%	**130,474** 16.47%	47 0.01%	53 0.01%	2 0.02%	**102**[B] 0.01%
Non consanguineous (F=0.0000-0.0039)	333,480 82.93%	314,839 82.91%	8,751 83.21%	**657,070** 82.93%	402,113 99.83%	379,716 99.83%	10,517 99.74%	**792,346** 99.83%
Total (N=)	**402,113**	**379,716**	**10,517**	**792,346**	**402,783**	**380,362**	**10,544**	**793,689**

Source: RPQA-IMPQ. Percentages are column percentages. For example, 83% of the population is non consanguineous when observing distant consanguinity (excluding close consanguinity), more than 3 ascending generations.
Denominator: Study population including children who died on their date of birth.
[A]. Most (1189) strongly consanguineous individuals with close consanguinity (considering 3 ascending generations) have inbreeding coefficients of 0.0625 (ex. parents are first cousins), some (51) individuals have F=0.125 (ex. parents are half-siblings or double first cousins) and 1 individual has F=0.25 (ex. parents are siblings).
[B]. All (102) weakly consanguineous individuals in close consanguinity have inbreeding coefficients of 0.03125 (ex. parents are first cousins once removed).

When focusing on the data of Table 8 by sex, there is the same proportion of strongly consanguineous boys and girls for distant consanguinity (0.60%-0.61%) and for close consanguinity (0.15%-0.16%). When looking at only three ascending generations, the

proportions between both sexes are equivalent for weakly consanguineous (0.01%) and non consanguineous (99.83%) boys and girls. They remain similar when observing distant consanguinity (16.5% and 83.9% respectively). The unknown sexes have slightly higher proportions of strongly consanguineous individuals when observing distant and close consanguinity. There is a slightly smaller proportion of weakly consanguineous individuals of unknown sex (16.2%) compared to known sex (16.5%) when observing distant consanguinity. The proportion of weakly consanguineous individuals of unknown sex is quite similar to known sexes for close consanguinity (0.02%). Children of unknown sexes are often children who die within a week of their birth. Knowing individuals of unknown sex often die rapidly and that there is a higher proportion of strongly consanguineous children of unknown sex, the proportion of children who die prior to their fifth anniversary by consanguinity level and sex is verified in Figure 4. Note that children who die the day of their birth are accounted for in the next figure, so it is normal if the proportion shown is not identical to what is seen in the Cox models.

Figure 4. Death Prior to Age 5 per Level of Consanguinity by Sex and Type of Consanguinity, Frequencies and Percentages[1], Colonial Quebec, 1720-1830



Source: RPQA-IMPQ.
Denominator: Study population including children who died on their date of birth.
[1]Percentages are based on the number of males and females in each type and level of consanguinity (see Table 8). For example, when observing close consanguinity, 35.15% of (625) strongly consanguineous boys die prior to 5 years old.

Almost all the individuals with an unknown sex died prior to age five. In fact, most of them (98%) died within a week of their birth (see Annex, Figure A4). As expected and discussed in the literature review, males have a higher child mortality than females and closely consanguineous children have a higher mortality than distantly consanguineous children of the same sex and consanguinity "strength" (strongly consanguineous – $F>3.125\%$, weakly consanguineous – $F\leq3.125\%$, non consanguineous – $F<0.39\%$). As discussed in the literature review, the number of meiosis occurring between common ancestors and individuals with close consanguinity is reduced compared to individuals with distant consanguinity. Thus, the length of IBD genes is much greater for closely consanguineous children. The percentages hint at a higher risk of childhood mortality when identical genes are of longer length due to close family ties, but the differences in survival may not be statistically significant.

When analysing close consanguinity with three ascending generations, there are very few incidences of death for consanguineous children as there are few of them in our database (≈625 childhood deaths, all sexes considered). For this type of consanguinity, the percentage of weakly consanguineous children who died prior to 5 years old is 35% for boys and 28% for girls, which is equivalent to the proportions observed for strongly consanguineous (35% for boys and 29% for girls), but higher than the proportions for non consanguineous children (30% for boys and 27% for girls). Weakly consanguineous boys and girls also seem slightly disadvantaged compared to strongly consanguineous children when observing distant consanguinity. However, this effect may not be statistically significant. Further analyses are necessary to comprehend the mechanism behind these descriptive statistics, but this may hint at an interaction between both hypotheses: (1) children of consanguineous couples will manifest a higher mortality compared to children of non consanguineous couples, (2) consanguineous individuals, as measured in this study, manifest lower mortality risks because they benefit from their ancestors' settlement in the colony. Kaplan-Meier curves can observe the phenomenon on a timeline. This can help understand when certain disadvantages come into effect.

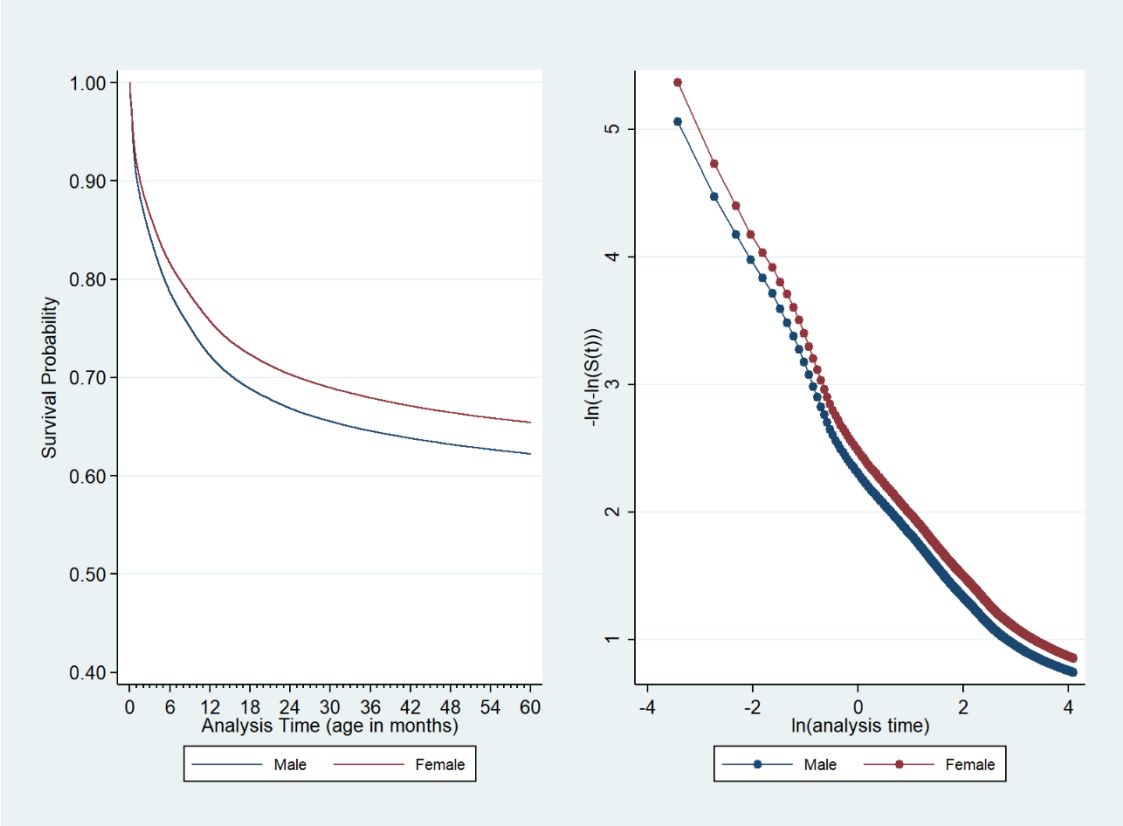
4.1.3. Kaplan-Meier Survival Curves

As discussed in Chapter 2, the Kaplan-Meier survival curves show the proportion of children at risk of dying, as a function of time, where the origin (time 0) is the individual's birth.

The age is calculated in days although tick markers on the X-axes are in months. The associated log-negative-log functions are also shown. These Kaplan-Meier estimates allow for preliminary interpretations of the role of each variable and hint at proportional hazards violations, which would affect the Cox models presented later in the chapter.

We begin by observing sex, period, and region variables.

Figure 5. Kaplan-Meier Survival Curves by Sex, Colonial Quebec, 1720-1830

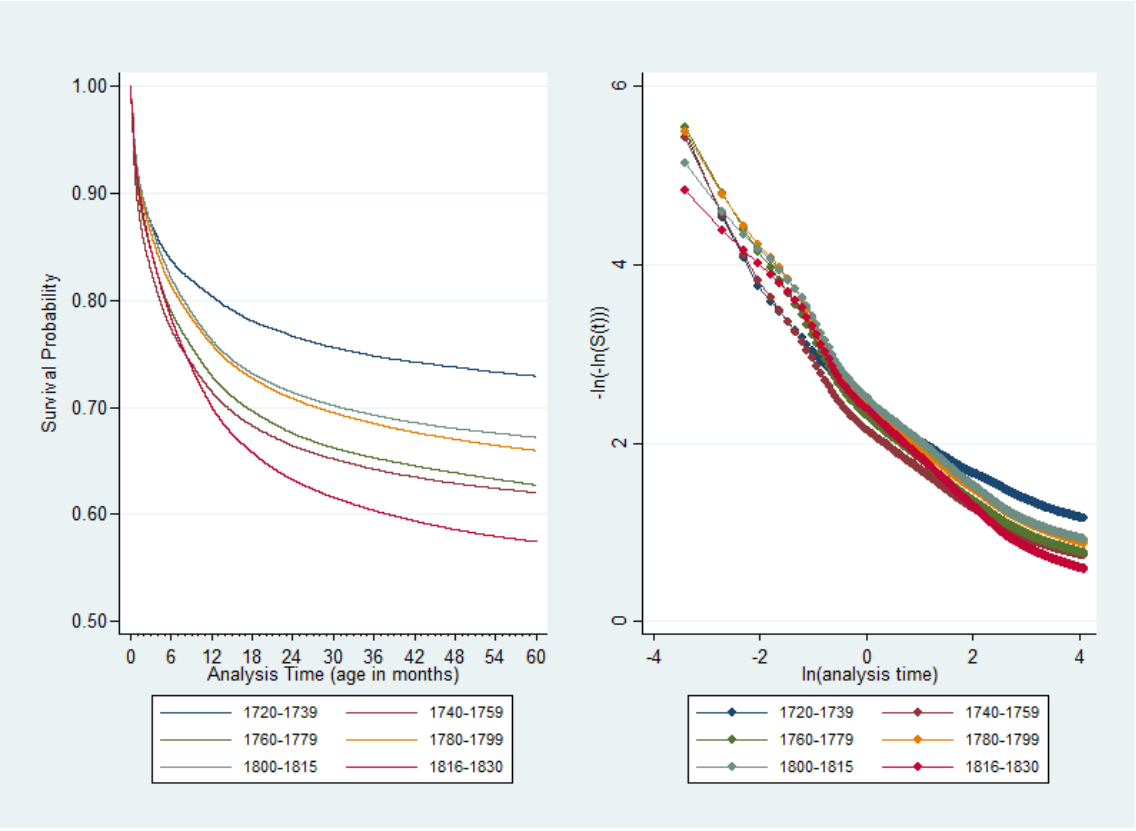


Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

The survival probability curves show that girls have a higher survival than boys, as expected and seen in Table 5. Otherwise, both sexes seem the follow a similar trend, demonstrating proportional hazards of death. If curious about the unknown sexes removed from the study, the curve does not follow the proportionality hypothesis (see Annex, Figure A4).

Now, the survival curves for six different periods of about two decades are observed.

Figure 6. Kaplan-Meier Survival Curves by Period, Colonial Quebec, 1720-1830

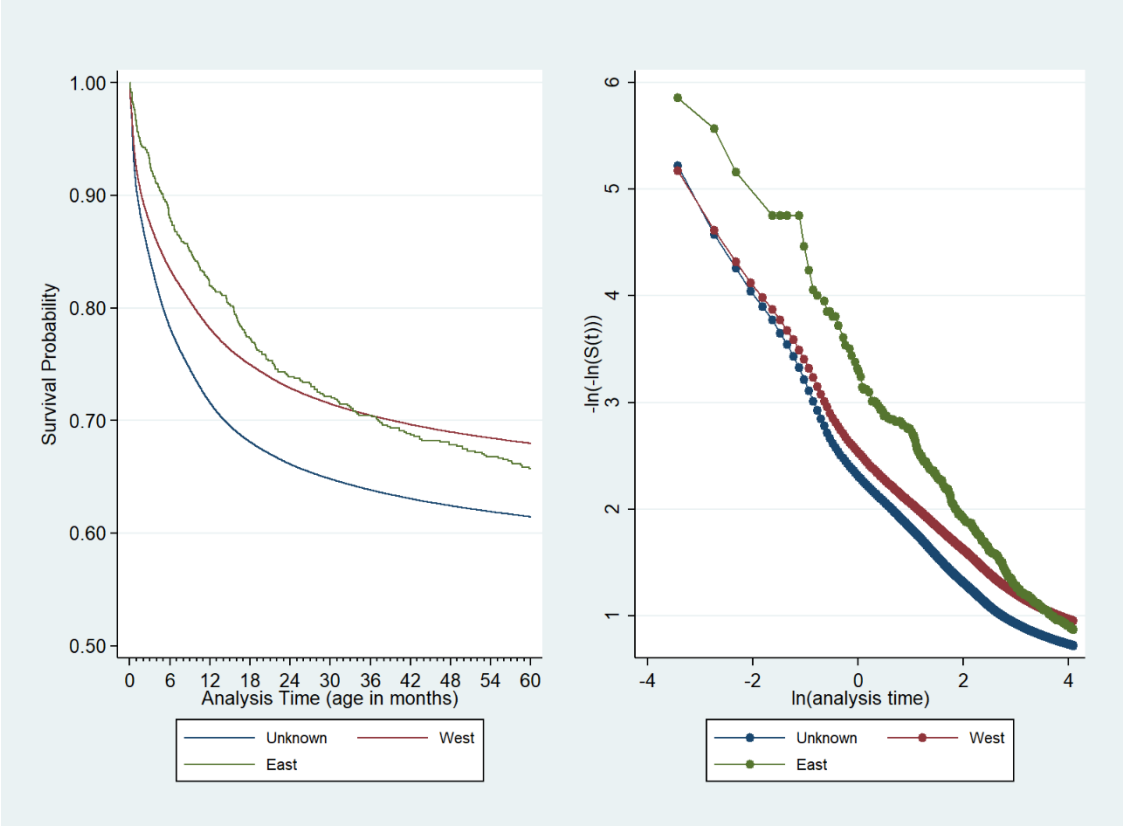


Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

The survival curves show that the early nineteenth century was overall deadlier for children over one and under five years old than the eighteenth century, but infant mortality seems less important for 1816-1830. This may also be an underestimation of infant mortality due to incomplete data (see Chapter 5). Otherwise, children born from 1720 to 1739 showcase the best survival curve. All log-negative-log curves overlap signifying violations of the proportionality hazards assumption. The absence of parallel lines justifies stratifying the Cox analyses by period. The expected group effect of period on childhood mortality must be considered in the Cox model to control this effect in the hazard function. The disadvantage of the stratification is that we cannot numerically evaluate how significant the effect of period is on childhood mortality, which may be interesting research, especially when noticing the slight plateauing of deaths in the late neonatal period for 1760 to 1830.

Next, the region will be observed, first by east and west, and then by smaller regions.

Figure 7. Kaplan-Meier Survival Curves by Cardinal Direction (East/West), Colonial Quebec, 1720-1830

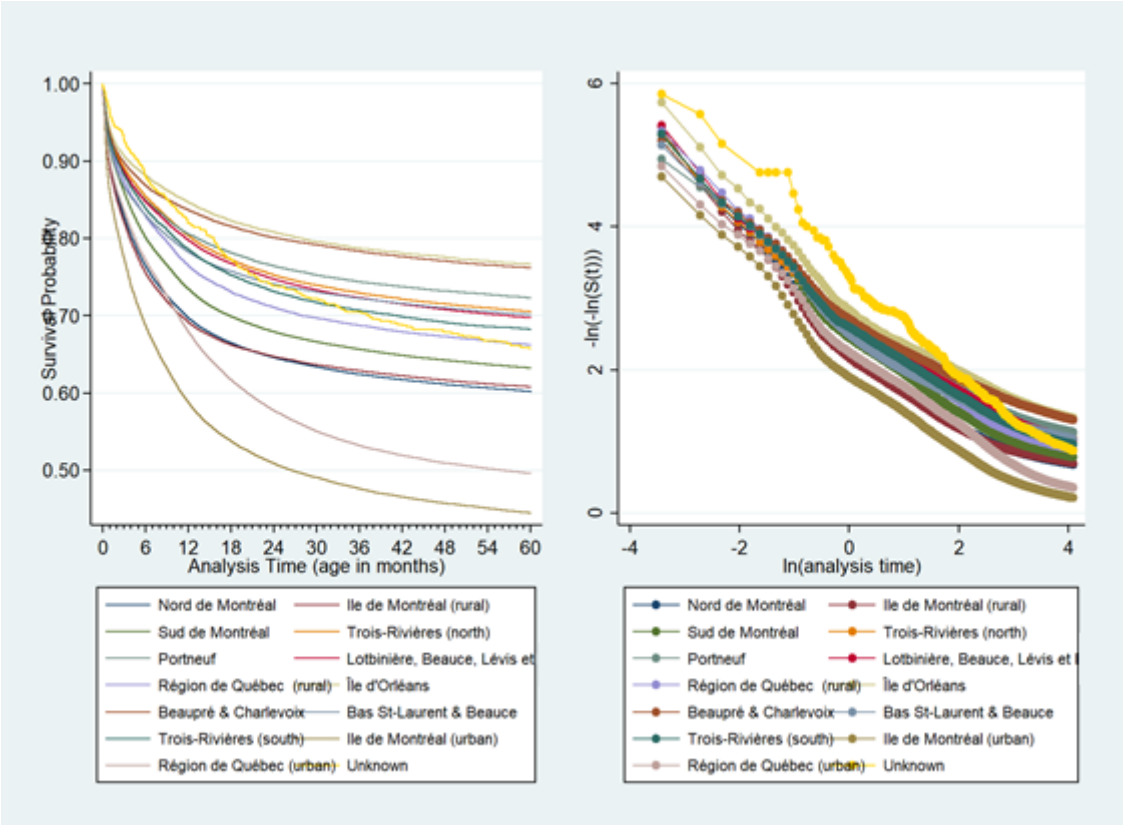


Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

Children in the west of the colony have a greater mortality compared to the children in the east until about three years old. At that age, the survival curves for the eastern and western children cross, showing the higher survival probability of western children from three years old until the end of the observation period. The children of an unknown region experience the lower survival probabilities at all ages observed. These include individuals without a registered birth region, if they survived until early childhood or were censored, or individuals without a death region, if they perished prior to five years old. The log-negative-log curves are not parallel, confirming the violation of the proportionality hypothesis. Consequently, the Cox analyses will be stratified by a more detailed geographical variable than east-west. Using the detailed region variable will control for some external factors not included in the models such as the weather, geographic resources and other environmental factors. Further, epidemics were identified

according to east-west, so stratifying by region remains consistent in the research. Here are the region survival curves.

Figure 8. Kaplan-Meier Survival Curves by Region and Urban Status, Colonial Quebec, 1720-1830



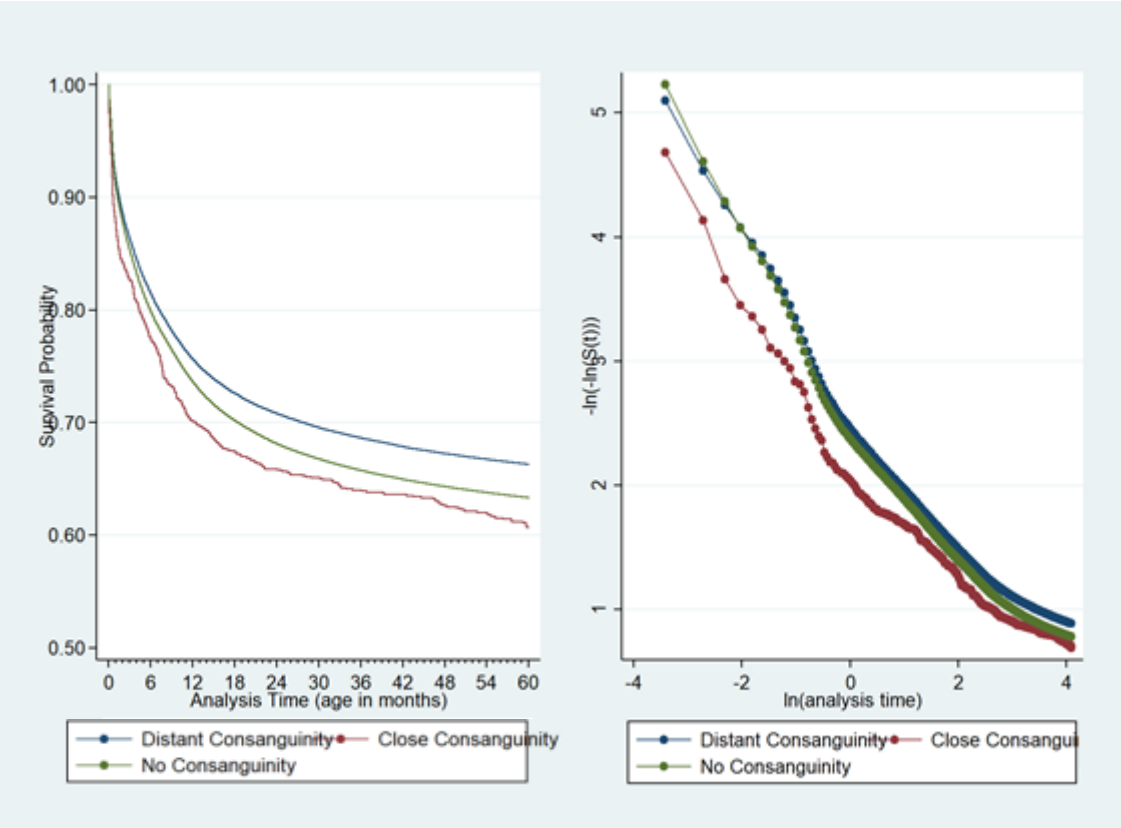
Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

Due to the lack of proportionality, the Cox models will be stratified by period and region so there are no coefficients for the effect of these variables on childhood mortality in the Cox models. The models will also be separated by sex as the objective of the study is to observe the overall effects of consanguinity and epidemics on child mortality and child mortality is different per sex.

Next, the Kaplan-Meier curves for our variables of interest, consanguinity and epidemics, will be observed. Throughout the chapter, consanguinity has been described in mutually exclusive categories of consanguinity: exclusive distant consanguinity, close consanguinity, and no consanguinity (thus, not even distantly consanguineous). The strongly and weakly consanguineous persons are grouped together for each degree of consanguinity,

close and distant, as observing them separately in multivariate analyses would involve many categories and would often involve small samples, which is not convenient for attaining statistical significance.

Figure 9. Kaplan-Meier Survival Curves by Consanguinity, Colonial Quebec, 1720-1830

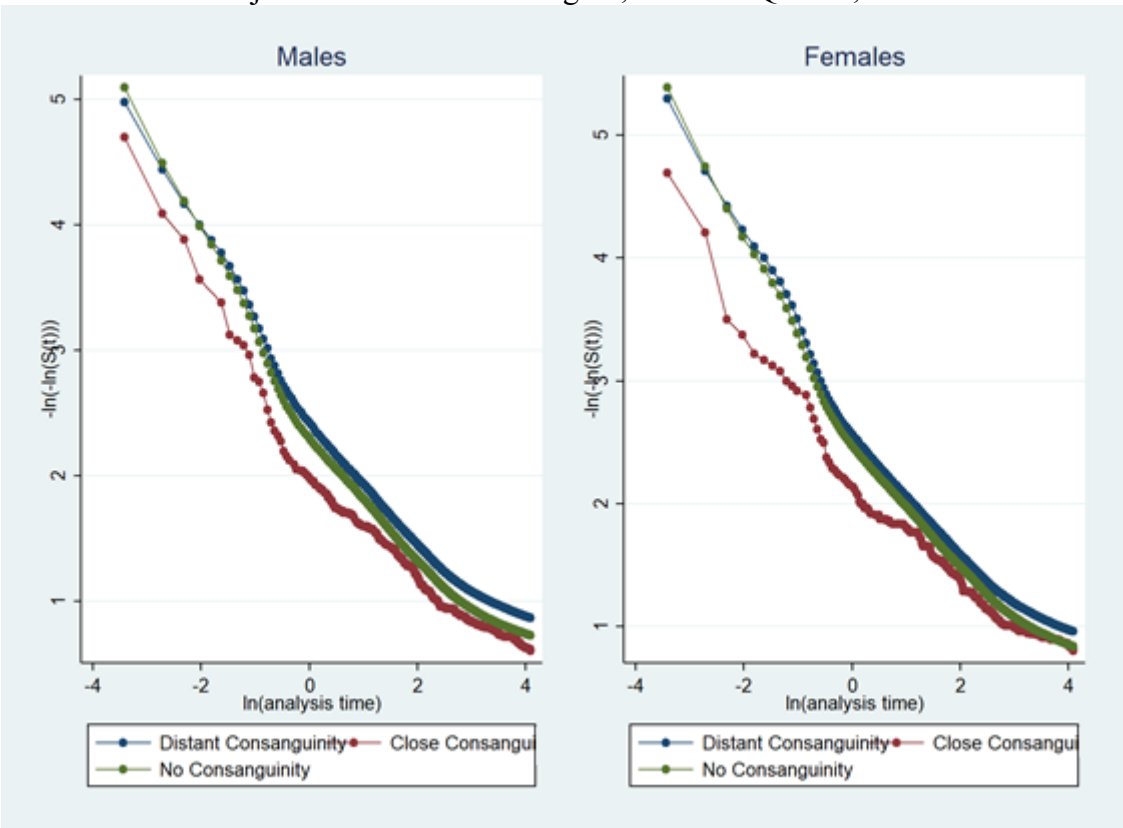


Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

The Kaplan-Meier survival curves of this consanguinity variable show distinct curves for each category; the closely consanguineous children have the lowest survival, followed by the non consanguineous. The distantly consanguineous children have the most advantageous survival curve, though this may be due to their family's extended settlement in the colony. The log-negative-log curves confirm that this variable does not follow the non proportional assumption necessary for the Cox model if the period of observation is not separated into shorter episodes. We cannot stratify our Cox model by the variable of interest, but we can check if proportionality can be claimed for consanguinity when controlling for other variables that will be in the models. For instance, the log-negative-log curves if we control for sex, period and region (the

stratification variables), show certain time intervals where consanguinity seem proportional: the late neonatal phase (7-27 days), the second half of infancy, etc.

Figure 10. Log-Negative-Log Survival Curves by Consanguinity and Sex, Adjusted for Period and Region, Colonial Quebec, 1720-1830



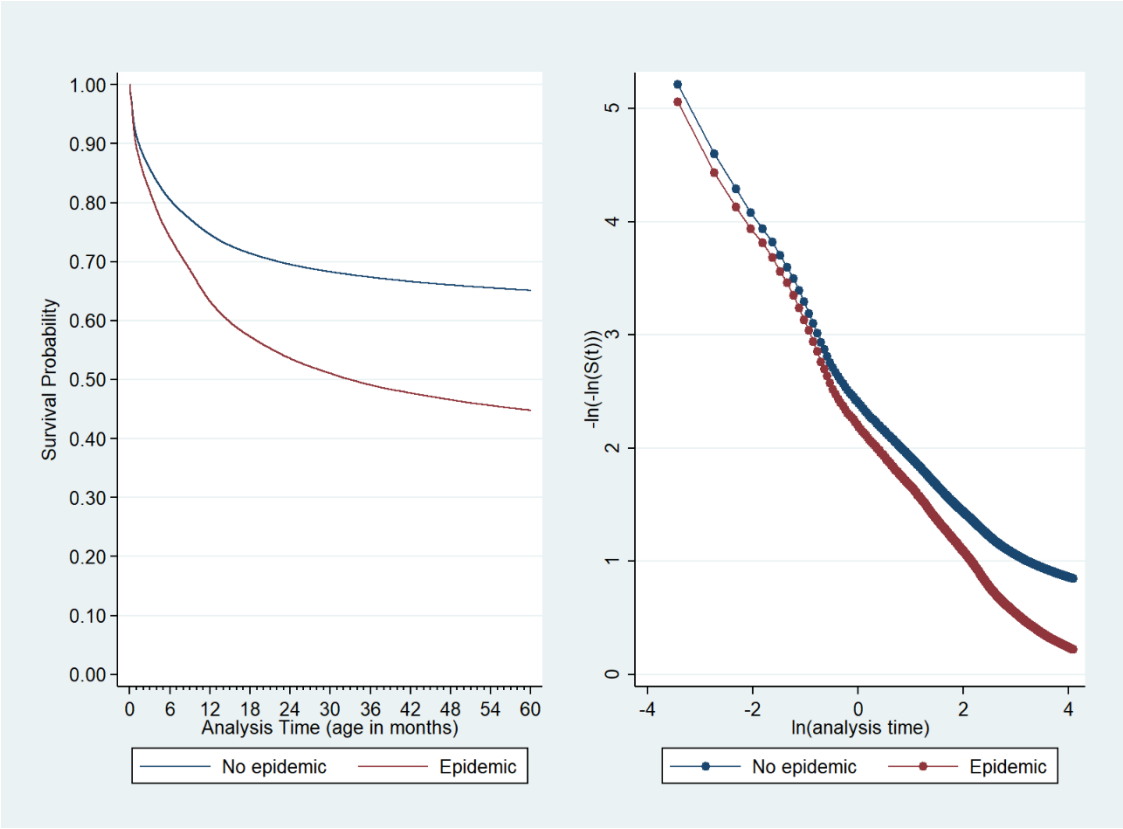
Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

The curves for distant and non consanguineous children are not quite parallel the first week after birth, but then seem to remain distinct enough for the rest of the observation period, suggesting proportionality or minor violation of proportionality. Recall, a minor violation of proportionality allows one to interpret the hazard ratio from the Cox model as an averaged effect on the period of study (Barraclough et collab., 2011 : 981). Close consanguinity is almost parallel to the other curves from three months old to a little over one year old and then, another section of the curve seems parallel for the rest of the observation period. Another time interval where the close consanguinity function is somewhat parallel to the other consanguinity curves is from 12 days to two months. Basically, if we make different models for shortened time intervals, consanguinity can follow the proportionality hypothesis required for Cox models.

The curves per sex (not shown) follow the same tendencies as the survival functions for consanguinity, but girls have a better survival than boys. The survival curve for closely consanguineous girls overlaps the curve for non consanguineous boys until one and a half years old, then it overlaps the male and distantly consanguineous curve (which is very close to the female and non consanguineous survival curve).

The other variable of interest is epidemics. In Table 5, the epidemics variable was described as "epidemic" if a child lived through at least one epidemic and "no epidemic" if a child did not live to go through an epidemic. However, in the multivariate analysis, the epidemic variable is operationalised as a time-varying variable. That is, for the same individual, the variable alternates between "epidemic" and "no epidemic" for the exact age the child lived through an epidemic. This information allows for precise analysis of survival during an epidemic. To show how widely different these two variables are, the Kaplan-Meier curves for the non varying measure of epidemics is available in the Annex (Figure A5).

Figure 11. Kaplan-Meier Survival Curves by Epidemics, Colonial Quebec, 1720-1830



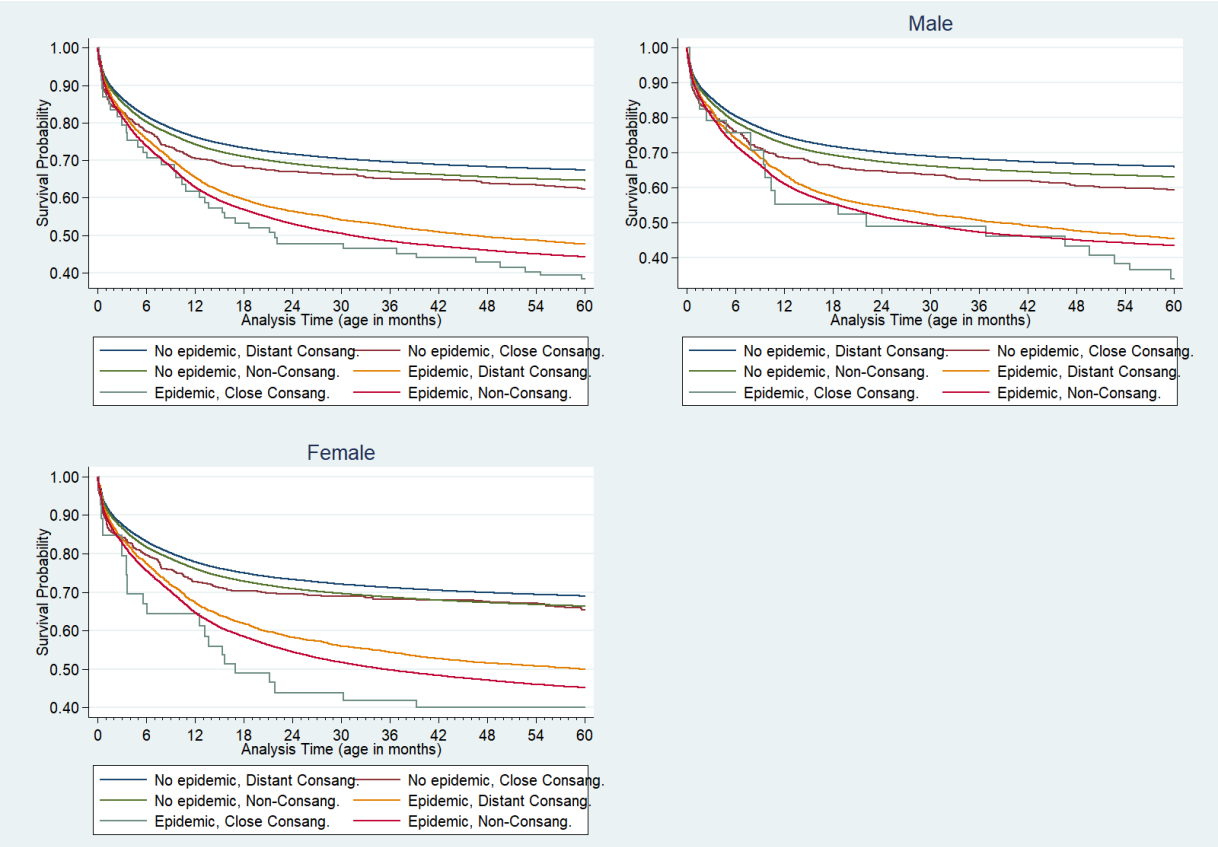
Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

With this time-varying epidemic variable, children living through an epidemic are on the "epidemic" curve during the days, weeks or months they lived the epidemic. The instant the peak of the epidemic is over, the children who survived are then described on the "no epidemic" curve for the ages that they were not in an epidemic. The Kaplan-Meier curves show that children are more at risk of dying during an epidemic than out of one. The survival curve for the children in an epidemic is much lower compared to children not in an epidemic. This is not surprising as the occurrence of epidemics are often determined by the death rate at the population level, especially the death rate of infants. The variable follows the proportional hazards necessary to be used in the Cox analysis for certain time intervals of the period of observation. For instance, in the first month of life, the log-negative-log curves are parallel. Furthermore, it appears proportionality is observed from one year old to five years old (date at censorship). In between the post neonatal period and one year old, there are periods where the log-negative-log curves are fairly parallel. This motivates analysing separate Cox models for certain time intervals, for instance, the first six days after birth (early neonatal period), then from seven days to 27 days (late neonatal period), then from 28 days to three months, etc.

The epidemic survival curves per sex suggest that sex and epidemics act independently on child survival as the functions seem identical for males and females. The only difference is that girls have a better survival compared to boys (see Annex, Figure A6).

Now, the interaction effect of the independent variables of interest, consanguinity and epidemics, on child survival is observed. There are for the six categories of survival curves that describe each interaction possible: no epidemic and distantly consanguineous, no epidemic and closely consanguineous, no epidemic and not consanguineous, epidemic and distantly consanguineous, epidemic and closely consanguineous, and epidemic and not consanguineous. The first graph includes all sexes and subsequent graphs are separated by sex.

Figure 12. Kaplan-Meier Survival Curves by Epidemic and Consanguinity for All Sex United and for each Sex, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data. Denominator: Study population.

The Kaplan-Meier curves suggest that living an epidemic and being closely consanguineous follow different trends in childhood survival dependent on sex. When not living through an epidemic, males have distinct survival curves dependent on their consanguinity level, whereas closely consanguineous females seem to only have excess death prior to two years old, potentially suggesting a selection process where the weakest closely consanguineous girls die at young age and the remaining ones are robust enough to follow the survival of non consanguineous girls. On the other hand, when experiencing an epidemic, closely consanguineous girls show a more distinct and disadvantaged survival trend compared to distantly and non consanguineous females, whereas closely consanguineous boys have a survival function that overlaps their counterparts living an epidemic until four years old. In other words, the closely consanguineous boys living an epidemic only seem to have a distinct and disadvantaged survival curve at the end of the observation period. However, these closely

consanguineous children are present at much the lower frequencies in the database, so this tendency may not be significant, especially in the later ages when the sample is reduced, and the risk of death is overall lower.

In summary, these Kaplan-Meier survival functions suggest a disadvantage in survival for children with close consanguinity. Further, epidemics imply excess mortality, which is in line with what was discussed in the literature review. Then, for the interaction of both variables on survival, there seems to be different effects and selection processes dependent on the sex of the child. However, distantly consanguineous children still seem mostly disadvantaged. Still, to know if these effects are statistically significant and if they still hold when controlling for geographic and temporal factors, more robust models are necessary. In the next section, Cox models will be analysed per sex. The Kaplan-Meier curves do imply that proportionality is possible if the Cox models are separated by age intervals.

4.2. Multivariate Analysis

In this section, the effects of consanguinity and epidemics on child mortality are analysed for the age intervals: from birth to six days (early neonatal period), seven days to 27 days (late neonatal period), 28 days to three months (post-neonatal period), from three months to six months, six months to one year, and every year until five years. The age divisions allow for the analysis of the differential longitudinal effects of these variables. The models also control for various variables such as time periods of at most two decades, regions and their urban status, and sex.

4.2.1. Consanguinity and Epidemics: Stratified Cox Models per Sex and Age

First, the stratified models including the consanguinity and epidemic variables, as well as their interaction, are presented for each age group. The variables of interest are analysed with distinct models per sex. Four stratified models are analysed; the first two are single variable models, one with consanguinity, and the other with epidemics. The third model includes both variables' effect on mortality. Lastly, the fourth model includes the main effects of consanguinity and epidemics, as well as the interaction variable of consanguinity and epidemics. The reference categories are no consanguinity and no epidemic. Further, each Cox model was evaluated for proportional hazards using Schoenfeld residuals. If the proportional-hazards assumption is not respected in the model by this method, it is indicated in the model (see notes under the tables) and verified graphically (see Annex). The neonatal period is analysed first.

4.2.1.1. Neonatal Period: Birth to 27 days

The stratified Cox models including only consanguinity and epidemic variables show the significantly greater neonatal mortality in consanguineous infants, especially in the early neonatal period (0 to 6 days; Table 9). Closely consanguineous infants are the most disadvantaged when observing the early neonatal period, and this effect appears stronger for girls in all models ($HR_{Male} \approx 1.7$, $p<0.05$; $HR_{Female} \approx 2.1$, $p<0.001$). We have seen in the literature review that males generally have higher baseline mortality than females and that they are more susceptible to intrauterine death, which may explain in part why the hazard ratio appears stronger for females. For boys, the excess mortality of consanguineous children is statistically

significant for both types of consanguinity, distant ($HR_{Male} \approx 1.1$, $p<0.05$) and close ($HR_{Male} \approx 1.7$, $p<0.05$). However, for girls, the early neonatal mortality of distantly consanguineous newborns is not significantly different from that of non consanguineous girls. For distantly consanguineous girls, the excess mortality is rather observed in the late neonatal period (7 to 27 days). Interestingly, the excess mortality of closely consanguineous girls is no longer significant for that age group (7 to 27 days), as if most of the weakest females with close consanguinity died in the first week of life and only the more robust ones were remaining in the late neonatal period. In fact, a slightly greater percentage of closely consanguineous girls were lost in the early neonatal period (7.8%) compared to the boys (7.7%), which may hint at an underlying selection process closely consanguineous since a greater male infant mortality was expected (see % lost in Table 9).

The effect of consanguinity remains at the same order when we control for epidemics (in Model 3), but the main effect slightly increases when we control for the interaction in the early neonatal period. The deadly effect of epidemics also remains the same whether we look at Model 2 or Model 3 for the entirety of the neonatal period. This stability hints that the main effect of epidemics is independent of consanguinity since the HR doesn't change whether we control for consanguinity or not. The hazard ratio of epidemics is also higher and more significant for girls compared to boys in the first week of life ($HR_{Male} \approx 1.1$, $p<0.05$; $HR_{Female} \approx 1.3$, $p<0.001$), probably, once again, owing to the overall higher mortality of boys in the early neonatal period. In the late neonatal period, the intensity of the effect of epidemics increases for males ($HR_{Male} \approx 1.2$, $p<0.001$) and remains of similar order for females ($HR_{Female} \approx 1.3$, $p<0.001$). Therefore, girls living an epidemic between 7 and 27 days have a similar relative risk of dying (ref=not in an epidemic) compared to the ones living an epidemic in their first week of life.

Now, looking at the interaction between epidemics and consanguinity the effects tend in an opposite direction than expected, except for males in the late neonatal period. Nonetheless, all hazard ratios are not significant if considering a 95% confidence interval, probably due to the low sample size of consanguineous children who die in their first week of life while living an epidemic (1 230 distantly consanguineous and 27 closely consanguineous). However, the estimate of the ratio of hazard rates still shows an averaged tendency for distantly

Table 9. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models Per Sex, Neonatal Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs.	Log Likelihood	AIC	BIC
[0 days, 7 days[	**Male**	Mod. 1	1.069*	1.691*				310,961	313,184	-70,876	141,755	141,776
		Mod. 2			1.109*					-70,878	141,758	141,769
		Mod. 3	1.069*	1.694*	1.109*					-70,873	141,752	141,784
		Mod. 4	1.084*	1.773**	1.151**	*0.779†*	-			-70,870	141,751	141,804
		% lost	-	-								
[0 days, 7 days[	**Female**	Mod. 1	1.043	2.085***				299,451	301,688	-54,972	109,948	109,970
		Mod. 2			1.265***					-54,968	109,937	109,948
		Mod. 3	1.043	2.085***	1.265***					-54,962	109,930	109,962
		Mod. 4	1.059	2.146***	1.310***	0.786	0.598			-54,961	109,931	109,984
		% lost	-	-								
[7 days, 28 days[	**Male**	Mod. 1	1.014	1.357*				292,655	299,820	-165,370	330,743	330,764
		Mod. 2			1.175***					-165,358	330,717	330,728
		Mod. 3	1.014	1.360*	1.175***					-165,356	330,717	330,749
		Mod. 4	1.01	1.364*	1.163***	1.071	0.942			-165,355	330,720	330,773
		% lost (6.1%)	4.7%	7.7%								
[7 days, 28 days[	**Female**	Mod. 1	1.059*	1.263				284,162	291,163	-136,035	272,073	272,094
		Mod. 2			1.287***					-136,009	272,020	272,030
		Mod. 3	1.058*	1.267	1.287***					-136,005	272,016	272,048
		Mod. 4	1.060*	1.226	1.289***	0.982	1.585			-136,005	272,019	272,072
		% lost (5.3%)	4.3%	7.8%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See Figure A7 in Annex.

consanguineous boys of having a lower relative risk of dying compared to non consanguineous boys when living an epidemic in the first week of life, though the limited number of cases only makes the interaction significant with 90% confidence ($HR_{Male} \approx 0.8$, $p<0.1$). In the late neonatal period, the distantly consanguineous boys living an epidemic have a hazard ratio greater than one. The HR is not significant, but it is now in the hypothesized direction, perhaps due to the distantly consanguineous boys having had experienced less mortality in the previous age interval.

In all, the hazard ratios for close consanguinity are much greater than those for epidemics in both neonatal periods, hinting that intrinsic mortality, that is mortality because of genetic factors or endogenous factors in general, is more important in the first month of life compared to extrinsic mortality, that is mortality due to environmental hazards. The statistically significant excess mortality of closely consanguineous children is however not perceivable in epidemic periods in the neonatal period, so Model 3 tends to be the best overall model for this period, based on loglikelihood ratios, AICs or BICs.

4.2.1.2. Post-neonatal Period: 28 days to 3 months

Now, let's turn to the period of life from 28 days to three months old. First, notice the amount of consanguineous children lost in the neonatal period. Considering that selection may be an underlying process in the models and that it is not controlled, the proportion of children lost in each category of consanguinity can help hint at selection. In the late neonatal age group, 8.7% of boys, and 6.9% of girls, with close consanguinity inbreeding coefficients were lost from the sample, which is higher than the 6.8% and 5.8% of non consanguineous boys and girls. A lower proportion of children with distant consanguinity is lost, 6.2% and 5.4%, respectively.

Consanguinity does not affect child mortality for the post-neonatal period. Distantly consanguineous girls present a marginally significant 4% excess mortality (i.e, at the 90% level of confidence), but otherwise, child mortality for consanguineous persons is statistically the same as non consanguineous individuals. In fact, AIC and BIC statistics confirm that the best model to explain child mortality from 28 days to three months is Model 2 where epidemic is the sole variable in the stratified models per sex. Evidently, if mortality due to consanguinity is not

Table 10. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models Per Sex, Post-Neonatal Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs.	Log Likelihood	AIC	BIC
[28 days, 3 months[	**Male**	Mod. 1	1.016	1.048				273,052	293,157	-165,935	331,875	331,896
		Mod. 2			1.344***					-165,884	331,771	331,781
		Mod. 3	1.016	1.045	1.344***					-165,884	331,774	331,806
		Mod. 4	1.023	1.033	1.362***	0.91	1.146			-165,883	331,777	331,830
	% lost (6.8%)		6.2%	8.7%								
[28 days, 3 months[	**Female**	Mod. 1	1.045†	0.960				267,915	287,831	-142,507	285,019	285,040
		Mod. 2			1.315***					-142,472	284,946	284,957
		Mod. 3	1.044†	0.962	1.315***					-142,470	284,946	284,978
		Mod. 4	1.041†	0.985	*1.307****	1.044	*0.619*			-142,470	284,950	285,003
	% lost (5.8%)		5.4%	6.9%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See Figure A8 in Annex.

well reflected for infants 28 days to 3 months old, it is not well reflected in epidemic periods either. As the estimates for consanguinity approach 1 compared to the neonatal period, the HR for epidemic rise above previous values for this variable ($HR_{Male} \approx 1.34$, $p<0.001$; $HR_{Female} \approx 1.32$, $p<0.001$). The post-neonatal period seems to be when extrinsic mortality becomes more important than intrinsic mortality, at least in the intensity of the relative measure. The insignificant effect of close consanguinity on child mortality may be because the closely consanguineous children are very selected in the neonatal periods, and then, in the post-neonatal period, only the more robust ones remain. Particularly, not many external factors differ from one to three months old compared to the environment lived in in the neonatal period. This stability may aid in keeping the hazard ratios insignificant. The next age groups will indicate if this reasoning is probable.

4.2.1.3. Weaning Period: 3 months to 12 months

In Colonial Quebec some infants were weaned off breastmilk as early as 3 months old (Amorevieta-Gentil, 2010). However, it is believed that most children were weaned off breastmilk starting at 6 months old. Similarly to children aged 28 days to 3 months, infants 3 months old to 6 months old only have one significant factor affecting mortality amongst those considered and it is living an epidemic. This extrinsic mortality variable continues to increase compared to the previous age group (increase in HR from 33% to 48%), showing the greater gap in mortality hazards between epidemic and non epidemic periods as subjects age within the first year of life.

After two age intervals spanning 5 months where consanguinity was not significant (from 28 days to 6 months), the effect of close consanguinity is again significant for boys observed from 6 months to 1 year and in the same direction for girls, yet not significant. One can hypothesize that the strong excess mortality observed in the neonatal period "purged" the weakest consanguineous children which resulted in a short time interval where mortality between consanguineous and non consanguineous infants were similar. From 6 month to 1 year old, closely consanguineous boy are at a 47% greater risk of dying compared to non consanguineous boys ($p=0.012$, Model 1). The epidemic time-varying variable is not

Table 11. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models Per Sex, Weaning Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs.	Log Likelihood	AIC	BIC
[3 months, 6 months[	**Male**	Mod. 1	1.029	1.027				253,076	278,859	-147,153	294,309	294,331
		Mod. 2			1.478***					-147,070	294,142	294,152
		Mod. 3	1.028	1.026	1.478***					-147,069	294,144	294,176
		Mod. 4	1.025	1.073	1.471***	1.037	0.435			-147,068	294,147	294,200
		% lost (7.4%)	6.9%	7.4%								
[3 months, 6 months[	**Female**	Mod. 1	1.007	1.026				250,606	276,761	-122,357	244,718	244,739
		Mod. 2			1.487***					-122,285	244,572	244,583
		Mod. 3	1.006	1.028	1.487***					-122,285	244,576	244,608
		Mod. 4	1.0109	0.9733	1.500***	0.933	1.804			-122,284	244,579	244,631
		% lost (6.5%)	6.2%	6.1%								
[6 months, 1 year[	**Male**	Mod. 1	1.009	1.466*				235,199	283,486	-163,411	326,826	326,847
		Mod. 2 (NP)			*1.852***					-163,142	326,285	326,296
		Mod. 3 (NP)	1.009	1.472*	*1.852***					-163,139	326,283	326,315
		Mod. 4 (NP)	1.012	1.422*	*1.859***	0.97	1.416			-163,138	326,286	326,339
		% lost (7.2%)	6.5%	6.8%								
[6 months, 1 year[	**Female**	Mod. 1	1.013	*1.173*				235,446	284,664	-143,990	287,984	288,006
		Mod. 2 (NP)			*2.031***					-143,659	287,321	287,331
		Mod. 3 (NP)	1.012	1.163	*2.031***					-143,659	287,324	287,355
		Mod. 4 (NP)	*1.017*	1.279	*2.049***	0.955	0.227			-143,657	287,324	287,376
		% lost (6.1%)	5.6%	6.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, ŧp<0.1

[1] Stratified by period, region-urban. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_t}{[\text{N subjects}]_{t-1}}$.

NP=Non proportional model by Schoenfeld residuals. HR *in italic* = Non proportional variable by Schoenfeld residuals. See Figure A11 & Figure A12 in Annex.

proportional according to Schoenfeld residuals, making models 2, 3 and 4 non proportional for both sexes. However, the graphs in the Annex show that proportionality can easily be assumed for epidemics (Figure A11 & Figure A12). The log-negative-log survival graph confirms the violation of the proportion hazards assumption for consanguinity in the Cox models for girls six months to a year old. Absence of proportionality was foreseeable as risks of death decrease exponentially over time and there are selection processes affecting the models. Therefore, models 1 and 4 are grayed out and cannot be interpreted. This is not so much of an issue because the retained model for girls observed from six to twelve months is Model 2 which only contains the epidemic variable that is proportional according to graphical methods. The hazard ratio for epidemics jumps to twice the relative risk of death for girls in an epidemic episode compared to girls out of one. This is the greatest change in HR as the effect doubles compared to the previous age group. For boys, the relative effect of epidemics on child mortality also highly increases, going from 1.48, for the observation period three to six months, to 1.85, or the observation period six to twelve months ($p<0.001$).

In summary, for the weaning period, excess mortality due to consanguinity is not observed in epidemic or non epidemic periods. However, living an epidemic in this age group implies double the risk of death compared to non epidemic periods. This increase in hazard ratio follows the trend observed for all infant age groups, such that the mortality hazard for epidemic periods increases with time. At this point, extrinsic mortality is more important than intrinsic mortality.

4.2.1.4. Toddlerhood: 1 year to 3 years

The observed trend of epidemic hazard ratios with time seems to continue for 1- to 2-year-old children, increasing from 1.85 to 2.15 for boys, and from 2.03 to 2.32 for girls ($p<0.001$). However, the effect of consanguinity on child mortality significantly changes directions for individuals with distant consanguinity. They experience 6% to 7% less mortality compared to non consanguineous 1- to 2-year-olds. The effect of close consanguinity on mortality is insignificant, but also in this opposite direction. As briefly discussed, selection processes may partly contribute to a relative risk in this unexpected direction. In fact, in the weaning period, 11.6% of closely consanguineous boys were lost and 9.1% of close

consanguineous girls. These proportions are higher than the respective 8.7% and 7.9% observed for non consanguineous infants. Nonetheless, Model 4 shows that closely consanguineous 1- to 2-year-old girls experience significant excess mortality in epidemic periods. The HR for the interaction variable shows that their risk of dying is thrice that of non consanguineous girls living an epidemic. Interestingly, the boys of the same observation period have a HR for the interaction variable, epidemic and close consanguinity, in the opposite direction, however, it is insignificant.

When observing 2- to 3-year-olds, the effect of distant consanguinity loses its significance compared to the previous age group, but the HR is still in the same advantageous direction. Males with close consanguinity have an insignificant HR, but in the direction suggesting a higher mortality for them. To continue with the selection hypothesis described in the models for other age groups, the proportion of closely consanguineous boys lost in the last observation period, 1- to 2-years old, is much lower than the proportion of non consanguineous boys lost, 6.3% and 8.3%, respectively. This change in direction for close consanguinity hazard ratios, despite the insignificant p-value, suggests that, though closely consanguineous boys are typically robust in this age group, the excess mortality effect may appear at a later age group as their robustness may be temporary.

Two- to three-year-olds living an epidemic have a much greater risk of dying compared to the ones not living an epidemic. The effect is greater than two and a half times the risk for both sexes ($HR_{Males} \approx 2.6$, $p<0.001$; $HR_{Females} \approx 2.7$, $p<0.001$). Actually, Model 2 with only the epidemic variable is the preferred model as its AIC and BIC statistics are the lowest for girls and boys observed from 2 years to 3 years. Ergo, consanguineous toddlers do not show any significant signs of differential mortality compared to non consanguineous children. However, interacted with epidemics, girls with distant consanguinity have an almost significant ($p<0.1$) 18% lower risk of death during an epidemic compared to children with no consanguinity. Hypotheses as to why the advantage for distantly consanguineous is only in epidemic periods will be discussed in the next chapter. In this interaction model (Model 4), the effect of an epidemic is higher compared to Model 3, which does not include the interaction variable. This suggests that epidemics affect the mortality hazard at greater proportions than consanguinity

Table 12. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models Per Sex, Toddlerhood, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs.	Log Likelihood	AIC	BIC
[1 year, 2 years[	**Male**	Mod. 1	0.942*	0.804				214,918	304,209	-132,019	264,043	264,064
		Mod. 2			2.147***					-131,640	263,283	263,293
		Mod. 3	0.943*	0.802	2.147***					-131,637	263,279	263,311
		Mod. 4	0.939*	0.815	2.136***	1.037	0.855			-131,636	263,283	263,336
		% lost (8.7%)	8.0%	11.6%								
[1 year, 2 years[	**Female**	Mod. 1	0.931**	0.835				217,072	308,468	-128,247	256,498	256,519
		Mod. 2 (NP)			*2.322***					-127,781	255,564	255,574
		Mod. 3 (NP)	0.931**	0.823	*2.322***					-127,776	255,558	255,590
		Mod. 4	0.937*	0.627†	*2.334***	*0.943*	3.183**			-127,773	255,555	255,608
		% lost (7.9%)	7.4%	9.1%								
[2 years, 3 years[	**Male**	Mod. 1	0.95	1.118				197,411	282,973	-55,224	110,451	110,472
		Mod. 2 (NP)			*2.633***					-54,925	109,852	109,863
		Mod. 3	0.949	1.115	*2.633***					-54,924	109,854	109,885
		Mod. 4	0.967	1.304	2.689***	0.871	-			-54,921	109,852	109,905
		% lost (8.3%)	7.3%	6.3%								
[2 years, 3 years[	**Female**	Mod. 1	0.952	0.592				198,948	287,404	-54,717	109,438	109,459
		Mod. 2			2.741***					-54,391	108,784	108,795
		Mod. 3	0.953	0.594	2.741***					-54,389	108,784	108,816
		Mod. 4	0.978	0.594	2.814***	0.821†	0.998			-54,387	108,784	108,837
		% lost (8.5%)	7.8%	7.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

NP=Non proportional model by Schoenfeld residuals. HR *in italic* = Non proportional variable by Schoenfeld residuals. See Figures A13 - A16 in Annex.

does; once again, reflecting the superior role of extrinsic mortality compared to intrinsic mortality after the neonatal period.

4.2.1.5. Childhood: 3 years to 5 years

For 3- to 4-year-old boys, Model 2 with only the epidemic variable in the stratified model is retained. The hazard ratio remains of similar order as the past age group (2-3 years old) and therefore shows the expected strong relative effect of epidemics (HR=2.53, $p<0.001$). For girls of this age interval, the distant consanguinity advantage to survival is only significant with 90% confidence when the model is not controlled by the interaction variable. Therefore, the retained model is also Model 2 as it has the lowest BIC statistic out of the four models. The hazard ratio is like the past age group at 2.6 ($p<0.001$).

Close consanguinity has a significant negative effect on survival for 4- to 5-year-old boys and girls. Further, the interaction variables between consanguinity and epidemic are also significant for boys and show that the excess mortality of closely consanguineous boys is present during epidemics. As for girls, the excess mortality when closely consanguineous is present out of epidemic episodes. Epidemic hazard ratios remain extremely high, but the reduced gap between the effect of intrinsic and extrinsic mortality from 4- to 5- years is particular and hard to explain as the effect of the excess mortality when closely consanguineous differs according sex and epidemics.

This result may be due to the small sample sizes remaining for closely consanguineous children in this age group. In fact, by then, only about 300 boys and girls remain and only a dozen of them are lost for each sex from four years old to five years old. Other variables known to affect child mortality may help us better interpret the intensity of the effects of consanguinity and epidemics on mortality for children up to five years old in Colonial Quebec.

Adding control variables to the above models do not change the order of the effect of consanguinity or epidemics (see Annex II). The only exception to this stability is the interaction variable between epidemic and close consanguinity which had a HR=0.94 for boys in the late neonatal period and, now controlled, an HR=1.09. The estimates in the initial models and in these controlled models are insignificant but are tending in opposite directions. As soon as mother's birth interval or the fate of the previous sibling are added to the models, the HR

Table 13. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models Per Sex, Childhood, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs.	Log Likelihood	AIC	BIC
[3 years, 4 years[	**Male**	Mod. 1	1.045	*1.58*				189,001	271,753	-32,490	64,983	65,004
		Mod. 2			2.526***					-32,330	64,662	64,672
		Mod. 3	1.046	*1.593*	2.527***					-32,329	64,663	64,695
		Mod. 4	1.023	*1.449*	2.460***	1.167	1.808			-32,328	64,665	64,718
		% lost (4.3%)	4.1%	4.0%								
[3 years, 4 years[	**Female**	Mod. 1	0.926†	0.517				189,589	274,229	-33,769	67,541	67,562
		Mod. 2			2.635***					-33,582	67,165	67,176
		Mod. 3	0.927†	0.513	2.635***					-33,579	67,164	67,196
		Mod. 4	0.927	0.455	2.632***	1.001	1.83			-33,579	67,168	67,221
		% lost (4.7%)	4.8%	2.9%								
[4 years, 5 years[	**Male**	Mod. 1	0.994	2.080*				183,028	265,454	-22,504	45,011	45,032
		Mod. 2			2.604***					-22,381	44,764	44,774
		Mod. 3	0.993	2.092*	2.604***					-22,379	44,764	44,795
		Mod. 4	0.948	1.347	2.460***	1.359*	5.005*			-22,374	44,759	44,811
		% lost (3.1%)	3.4%	4.1%								
[4 years, 5 years[	**Female**	Mod. 1	0.942	1.965*				182,625	265,932	-22,717	45,439	45,460
		Mod. 2			2.921***					-22,552	45,106	45,116
		Mod. 3	0.941	1.961*	2.922***					-22,549	45,105	45,136
		Mod. 4	0.971	2.370**	3.018***	0.807	-			-22,546	45,103	45,155
		% lost (3.6%)	4.0%	2.7%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See Figure A17 in Annex. When the violation of the proportional hazards assumption is confirmed by the graphical method, models are grayed out and will not be interpreted (3- to 4- year old boys).

increases to almost one. The estimate increases above one when mother's birth interval, birth rank, fate of the previous sibling, and mother's age at birth are added to the model (HR=1.08). Otherwise, the only differences in hazard ratios are of small order for close consanguinity; for instance, a difference of maximum 0.1 in HR is visible for females 0 to 7 days old and males 4 to 5 years old. The largest difference of HR for the variables of interest when comparing models with control variables to the ones without is 0.18 for males 2 to 3 years old (Model 6A, HR=1.12 and HR=1.30, respectively). Considering that close consanguinity is the variable of interest the most affected, its small sample size could be why such disparities are seen.

As for the role of the control variables, the almost exclusively reproductive variables lose importance as the children at study get older. In contrast, the epidemic variable remains significantly explanatory for childhood mortality. The birth interval between the child at study and their previous sibling loses its significant effect on mortality with time, but when it is significant, in infancy, the shorter birth intervals are especially detrimental to survival. Further, if this previous sibling died prior to one year, the risk of death prior to 3 months for the child at study is highly significant. The control variable remains significant, but at lower order until about 2 to 3 years old, so it also loses importance with time. The mother's age at birth is hardly ever significant past the neonatal period and never significant after 2 years old. Following the trend of the other reproductive variables, rank is only significant in infancy and being the first born seems to incur the highest risk of mortality. Having the greatest number of grandparents, the indicator of settlement, was expected to be beneficial for survival. It often is, but there are also hazard ratios in the other direction, suggesting that the indicator may not be a robust enough proxy to settlement. This is further discussed in the discussion (Chapter 5).

4.2.2. Consanguinity and Smallpox: Stratified Cox Models per Sex and Age

Now that the effects of consanguinity on child mortality have been observed for the cumulation of all epidemics, let's briefly observe a case study of a particular epidemic to see if the effects signaled prior hold. The following models are identical to the previous models in section 4.2.1, except for the epidemic variable. In this case, the epidemic variable does not include all the epidemics the Colony experienced. Instead, it analyses only the smallpox epidemics, or what was assumed to be smallpox epidemics (see Table 3 in section 3.2). As

mentioned, smallpox epidemics cyclically targeted the young population of Colonial Quebec every seven years starting in the late eighteenth century (Bruckner et collab., 2018). There are, therefore, great chances that many individuals of the study population experienced a smallpox epidemic as the period of observation encompasses almost the entire cycle (five years). Furthermore, smallpox is known to be a child killer. However, as discussed in the literature review, maternal milk from a mother immune to this virus has protective effects for the child. The following models are briefly observed to see how consanguinity fits with these known smallpox mortality effects, and most importantly, if the effects of consanguinity and epidemics differ from the effects observed when we considered all epidemics regardless of their infection or disease.

First, the frequency table for smallpox epidemics per consanguinity level is shown. The frequencies for the other variables are the same as in Table 5.

Table 14. Number of Children who Experienced a Smallpox Epidemic Prior to Five Years Old per Consanguinity Level, Colonial Quebec, 1720-1830

Smallpox during time at risk [1]		
Epidemic	294,608	48.26%
No Epidemic	315,804	51.74%
Interaction: Epidemics and Consanguinity		
Epidemic*Distant Consang.	50,354	8.25%
Epidemic*Close Consang.	544	0.09%
Epidemic*No Consang.	243,710	39.93%
No Epidemic*Distant Consang.	49,248	8.07%
No Epidemic*Close Consang.	537	0.09%
No Epidemic*No Consang.	266,019	43.58%

Source: RPQA-IMPQ. [1]In the Kaplan-Meier and Cox analyses, a time-varying variable is used and therefore it is not equivalent to the variable presented in this table.

When specifically looking at smallpox epidemics, almost half of the study population lived through one during their time at risk. Knowing that individuals with close consanguinity constitute of a small sample size, it is relieving to see that a little more than half of them experienced a smallpox epidemic since the following Cox models should not be much more affected by sample size as the ones in the remainder of the chapter. Main findings are described briefly, and the models are available in the Annex (Table A1 – Table A5).

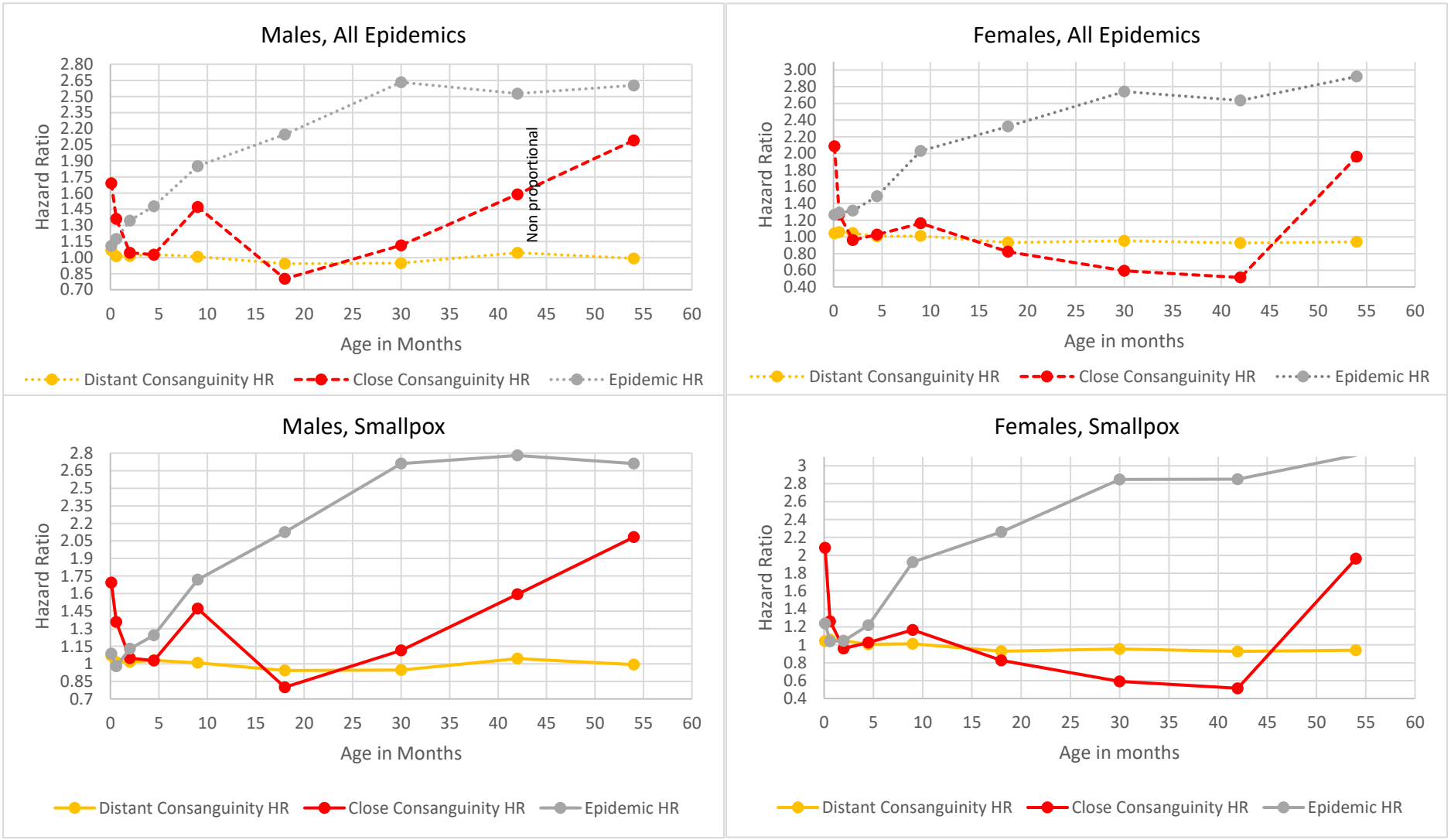
Now, to compare the intensity of these effects on childhood mortality, Model 3's hazard ratios are plotted at the midpoint of each age interval by sex. The plots show the increasing

importance of epidemics with age, the relatively stable risk of mortality for distant consanguinity and the varying close consanguinity effect. When we compare Model 3 to Model 3S (Model 3 but for smallpox epidemics only), consanguinity relatively follows the same trend and intensity per sex whether all the epidemics are observed or only smallpox. The major differences in HRs occur before and partly during the weaning period; the results show much smaller hazard ratios for smallpox compared to all epidemics, hinting that the passive immunity of breastfeeding has an important protective effect with smallpox, stronger than the protective effect of all cumulated epidemics. In fact, the hazard ratios for smallpox are between 1.00 and 1.25 until 6 months old, when it achieves hazard ratios of at least 1.71. Overall, smallpox seems less deadly during infancy compared to all the cumulated epidemics. But, after the weaning period, the hazard ratios approach the levels seen for all epidemics, and then become more important to childhood mortality at later childhood years (3 to 5 years old) compared to the cumulated epidemics model.

The models with the interaction term also show differences between the hazard ratios. With the main effect of the smallpox epidemic less intense during infancy, the interaction term between smallpox and close consanguinity tends to be higher. Distant consanguinity coefficients are basically identical whether one looks at smallpox only or all the epidemics. The main effects of close consanguinity are similar too. Principally, through the interaction model, smallpox signals a weaker effect of mortality during infancy compared to all epidemics, and stronger effects after 3 years old. However, the interaction between smallpox and close consanguinity hints at a higher mortality of closely consanguineous children during smallpox epidemics as opposed to all epidemics.

This chapter presented results which signaled at effects of consanguinity and epidemics. The next chapter will attempt to further reason for these effects, as well as bring forth limits of this study that do not permit a complete comprehension of the effects. The results obtained hint at selection effects, especially with close consanguineous children, along with different results for infant mortality and childhood mortality.

Figure 13. Plot of Hazard Ratios, Consanguinity and Epidemic, Model 3 and 3S of Stratified[1] Cox Models Per Sex, 0 to 5 years old, Colonial Quebec, 1720-1830



Generated by Excel with RPQA-IMPQ data. [1] Stratified by period, region-urban.

No consanguinity is the reference for distant and close consanguinity; No epidemic is the reference for the upper graphs (All Epidemics); No smallpox epidemic is the reference for the lower graphs (Smallpox). These same graphs on a logarithmic scale are available in the Annex (see Figure A18)

CHAPTER 5. DISCUSSION AND CONCLUSION

This final chapter of the thesis discusses the results obtained in the previous chapter and further interprets them in light of other studies. The methodological choices of the project are questioned, and limits are brought forth. This leads to ideas for further explorations of early childhood mortality in relation to consanguinity and epidemics. Finally, the contribution of this research to literature concludes the thesis.

The overall effect of consanguinity tends towards excess apparent mortality during infancy, although the effect is not always significant. The mortality is described as apparent because the data is not corrected for underreporting and may show lower mortality than if correction factors were used (Amorevieta-Gentil, 2010 ; Mazan, 2011a ; Nault et collab., 1990). However, if the correction factors affect all individuals more or less the same way, then the hazard ratios are still adequate as the correction would be present at both the numerator and denominator, essentially canceling each other out.

The change of statistical significance in infancy hazard ratios suggest a selection process since consanguineous children are especially targeted at certain age groups, and thus, in the next age group, the basin of consanguineous children becomes smaller and likely more robust. Precisely, closely consanguineous boys have a smaller HR in the early neonatal period (HR=1.7, $p<0.05$) compared to girls (HR=2.1, $p<0.001$). Therefore, consanguineous boys seem to be less selected at this early age group, perhaps partly due to their higher susceptibility to intrauterine deaths (Catalano et collab., 2005). Nonetheless, their pool after the early neonatal period is sufficiently large and frail to continue having a significant effect on mortality in the late neonatal period (HR=1.4, $p<0.05$). However, close consanguineous girls are heavily selected at birth and no longer have a significant effect in the late neonatal period (HR=1.3).

This selection process is also observed in the variation in direction of the effect of consanguinity on child mortality after infancy. But especially noticeable is the interaction variable that, even if often insignificant, switches direction at almost every age group. For instance, the interaction of epidemic and close consanguinity has a HR greater than one from 7 days to 28 days, a HR of less than one from 28 days to 3 months, and, once again, a HR of more than one from 3 months to 6 months.

The next table presents a summary of the direction of the effect of consanguinity on mortality based on Models 1 to 4 presented in section 4.2.1. If the hazard ratios are greater than 1, ergo, signaling excess mortality for consanguineous children versus non consanguineous children, then a "+" is indicated. Otherwise, a "-" is indicated in the table.

Table 15. Direction of the Effect of Consanguinity per Age Group and Sex in the Stratified[1] Cox Models, Colonial Quebec, 1720-1830

		Males				Females			
	Age Group	Distant Consang	Close Consang	Epidemic x Distant	Epidemic x Close	Distant Consang	Close Consang	Epidemic x Distant	Epidemic x Close
Infancy	0d-7d	+ *	+ *	- *		+	+ *	-	-
	7d-28d	+	+ *	-	-	+ *	+	-	+
	28d-3m	+	+	-	+	+ *	-	+	-
	3m-6m	+	+	+	-	+	+	-	+
	6m-12m	+	+ *	-	+	+	+	NP	NP
Childhood	1y-2y	- *	-	+	-	- *	-	-	+ *
	2y-3y	-	+	-		-	-	- *	-
	3y-4y	+	+	+	+	- *	-	+	+
	4y-5y	-	+ *	+ *	+ *	-	+ *	-	

Source: RPQA-IMPQ.
[1] Stratified by period, region-urban. *: the effect is significant at least at α=0.10. The interaction variables are based on Model 4 only.

The results of the study signal that all types of consanguinity are hazardous to infant survival. This is seen by the "+" signs for the main effects in the table. However, from one to five years old, distantly consanguineous children, especially girls, tend to have an overall better survival than their non consanguineous counterparts. Even closely consanguineous girls from one to four years old signal positive survival main effects, though the interaction HR are sometimes greater than one, suggesting a potential disadvantage for survival specifically in epidemic episodes. This disadvantage during epidemics is also signaled for boys of both types of consanguinity. After infancy, closely consanguineous boys still indicate a trend towards excess mortality that becomes significant when observing four- to five-year-olds. This is the only age group where closely consanguineous girls show disadvantaged survival in childhood. This seesawing in outcomes may be a mixed effect of settlement and selection. Both selection and the deep rooting in the colony are thoroughly described in this chapter.

As mentioned above, an underlying selection effect seems to frequently alternate the outcome of consanguinity on child mortality during epidemics. When hazard ratios for the interaction variable are below 1, the estimate signals a reduced risk of death compared to the

expected independence of effects of epidemics and consanguinity on child mortality. To highlight the selection process with an example, let's observe the smallpox models; the closely consanguineous girls are strongly selected in the neonatal period ($HR_{close\ consang.0\text{-}7days}=2.1$, $p<0.001$, $HR_{close\ consang.*smallpox7\text{-}28days}=1.5$), which may explain why their HR for the main effect of close consanguinity, and also its interaction with smallpox, are below 1 in the post-neonatal period ($HR_{close\ consang.1\text{-}3months}=0.96$, $HR_{close\ consang.*smallpox1\text{-}3months}=0.91$). The hazard ratios are not significant but do tend in a different direction than initially observed. This, once again, follows the selection processes explained in the literature review (Catalano et Bruckner, 2006 : 1641 ; Palloni, 1990 : 201 ; Willführ et Gagnon, 2013 : 200).

Selection processes influence statistical significance, and this can be seen by the outcome of consanguinity on child mortality at four years old, where the effect is significant especially for boys. However, the mortality risk is lower for everyone during later childhood. For example, less than 3,000 deaths occur for all (183,028) boys from four to five years old, and only a fraction of these deaths occur during an epidemic. Thus, in absolute numbers, not as many deaths are available to lead a significant effect in the later childhood years, especially for the interaction variables since they are based on small sample sizes when overall mortality risk is low; therefore, the selection process is substantial. These small samples also contribute to the seesawing of HR observed. To control for selection, more robust analyses would be required to better interpret the results of this study. An example of such a method of analyses is the Heckman correction.

The role of infections on consanguineous deaths is also amongst this background noise that is not controlled for in this master's thesis. Gauvin's (2016) advance concerning the association of consanguinity with a compromised immune system, essentially the basis of this thesis, would also affect consanguineous mortality during infections which are not epidemics. For instance, the warm summers exacerbated death by infection in the study period, especially in the urban cities of Colonial Quebec (Bruckner et collab., 2018). Since these periods of high mortality are included in the reference mortality of consanguineous children for the interaction variables, it is that much harder to obtain a clear significant effect of consanguinity during an epidemic episode. Controlling for urban status as a variable in the Cox models would have potentially hinted at this effect. In fact, simply having stratified regions and periods reduced the

precision in the estimation of hazard ratios and the power of the analysis. Thus, this conservative method for managing violations of the proportional hazards' assumption makes it harder to achieve significant estimates, especially in this case where the stratified Cox models included 84 strata (i.e., 6 periods x 14 regions). The excessive computational time of the Cox models for 610,412 subjects (further split into epidemic episodes) disallowed a variable describing interaction with time for the non proportional variables.

This said, the more categories added to a variable, the more reduced the sample sizes are, and the more difficult it is to achieve significant results. Thus, it could have been beneficial to group the parishes into broader regions to increase the number of events in each stratum and increase the statistical power of the models while still controlling for many environmental factors. That is exactly what Mazan (2011a : 34) did in his thesis: he separated the colony into six regions when studying the early eighteenth century. However, without enough distinct categories, the level of unobserved heterogeneity will affect the interpretation of the estimates. Recall in the literature review, that epidemics and consanguinity occurrence depended on region. However, by reducing the number of strata, it may have been possible to stratify some other control variables used in the study, something that could not have been done in this thesis since the models already had many strata.

The issue of unobserved heterogeneity discussed above is present in this study, but with the reference category of no consanguinity. Future research on this subject should further classify the no consanguinity category to increase homogeneity in the measure. There should be no issue in sample sizes as "no consanguinity" is the largest category of consanguinity in this study. The unobserved heterogeneity in the reference category may underestimate the hazard ratios for consanguinity. Consanguineous persons are in smaller groups which are heavily selected at the beginning of the observation period. Therefore, these groups approach homogeneity and robustness faster than the reference group of non consanguineous persons. In fact, considering that the strongly consanguineous children studied have 3% to 25% probability of identical genes by descent, it is possible that most of the children remaining after infancy do not even have identical genes due to a common ancestor, and resemble mostly the heterogeneous non consanguineous group. In fact, distantly consanguineous children, whose inbreeding coefficients may be as small as 0.39%, tend to have HR very close to 1 implying a similar risk

of death compared to non consanguineous children. If the reference group were children with no consanguinity and no immigrant parents, then the comparison would be more intuitive and the consanguinity interaction with epidemics would probably be stronger for the reasons evoked in the next paragraphs.

In this research, consanguinity is a confounding measure as it is assessed by genealogical ascendence. It therefore also measures the level of settlement in the colony. When settled in the colony for several generations, parents and grandparents not only have a greater social network as they are rooted in the community, they also (usually) passed through several common epidemics. Thus, for this period of study, they are perhaps immune to these diseases and can properly care for the child if the child gets infected as oppose to unsettled parents or immigrants who have a high risk of death if they never contracted the disease prior to when the child was at risk of contracting it. This may be one of the reasons why girls with distant consanguinity have an almost significant ($p<0.1$) 18% lower risk of death during an epidemic compared to children with no consanguinity at two to three years old. Further, if the infant is being breastfed by a mother who is immune, she transmits passive immunity to her child (Niewiesk, 2014 ; Palloni, 1990), whereas immigrant mothers, though more rare than immigrant fathers in Colonial Quebec (Mazan, 2011a), cannot transmit an immunity they do not possess, and their risk of mortality is high. Ergo, if a mother dies of an epidemic, then the child is not only at risk of death because of the epidemic, but also because of the death of the mother and that lack of care, which could instantly double the risk of death of the child (Pavard et collab., 2005 ; Willführ et Gagnon, 2013 : 199-200).

To further continue on the biases in this study due to the variable of interest also measuring settlement, Mazan (2011a) discussed the potential added support from extended family when settled in the colony which aided in the survival of children after a measles epidemic. In fact, he brought forth that “the majority of flu deaths were among exposed children who had at least one immigrant parent” (Mazan, 2011a : 121). Ergo, among the non consanguineous children in this study, the ones with immigrant parents have higher mortality than the settled non consanguineous children. The immigration status of the child’s parents was initially observed in this study but including the measure in models with consanguinity implied too much collinearity as children with two immigrant parents, parents with no ascending

genealogies in the colony, were non consanguineous. Immigration was not the interest of this study, so the variable was omitted. However, after the exploration done with this research, including the variable in the measure of consanguinity may be a good idea to avoid bias in the measure due to heterogeneity. Instead of one "no consanguinity" category, future research could separate this group into three to increase homogeneity in the measure: no consanguinity and no immigrant parents, no consanguinity and one immigrant parent, and no consanguinity and two immigrant parents. When immigration status of the parents was included in the preliminary analyses of this thesis, the results tended towards a protective effect when the child had no immigrant parents or two immigrant parents, compared to having just one immigrant parent. One can hypothesized that this effect would hold when categorized into the consanguinity measure, therefore probably increasing the estimate of the interaction between consanguinity and epidemics if the reference category is no consanguinity and no immigrant parents.

The effect of the unobserved heterogeneity of parents' immigration status for the non consanguineous group is strongly related to the inadequate control of settlement. The attempted proxy to settlement in this study, the number of great grandparents, was not sufficiently robust. It often tended in the expected direction (the controlled stratified models of Annex II show hazard ratios in the order of 0.8 or 0.9 for the early neonatal period and from six months to four years), but the variable still went against the expected direction a few times. Since the proxy is only partially controlling for settlement, the effect of consanguinity is probably further underestimated as the measure encompasses opposite effects, as discussed in the hypothesis.

Going back to the results, factors other than selection may further explain why boys with distant and close consanguinity would signal excess mortality during an epidemic at four years old. First, males with close consanguinity may have a greater susceptibility of death than girls. Milot et collab. (2017 : S16) studied boys and girls who are T14484C carriers and compared their survival to the overall French-Canadian population of Quebec from 1670 to 1775. They found that "male carriers, but not female carriers, suffered greater infant mortality" compared to the overall population. So, in consanguinity's case, it is possible that mortality due to consanguinity differs by sex, such that boys are more susceptible of dying than girls, as they are in general conditions. This may be the consequence of longer stretches of IBD in boys' genome due to less recombination in the meiosis processes for boys, but this study cannot verify that

(Gagnon et collab., 2005). Milot et collab. (2017) also noticed that, in epidemics, all children had a higher relative risk of dying; being a carrier did not further increase the already high risk of mortality in epidemics significantly (Milot et collab., 2017 : S4, S16). This insignificant interaction with epidemics is in line with the results of this study for many age groups. However, the results of this thesis do show age groups where epidemics and specific types of consanguinity interact with one another to increase risk of death, (ex. girls at one revolved year old) or, otherwise, lessen the fatal effect usually observed (ex. boys in the first week of life), but it was discussed in the beginning of this chapter that selection may be a major underlying effect of this.

The last section of the results also suggests that the effect of the interaction between epidemics and consanguinity differs based on the type of epidemic. Smallpox had more detrimental and significant interaction effects with consanguinity compared to overall epidemics. However, as explained, the main effect of smallpox was of lower order than the hazard ratios of overall epidemics. Though this lower order of HR may be because they are slightly underestimated; having studied only smallpox, the other epidemics were grouped with the reference category, which biased the estimates. Nonetheless, the weaning period accounts for a differential intensity of the effect of all epidemics. This is confirmed by the fact that proportionality cannot be followed by Schoenfeld residuals. As explained in the results section, absence of proportionality was foreseeable as risks of death decrease exponentially over time and there are selection processes affecting the models. From the Kaplan-Meier curves (Figure A11 and Figure A12), the augmented relative effect of epidemics on childhood mortality is unquestionable even if the hazard ratios cannot be further analysed due to non proportional hazards. Ergo, the maternal antibodies seem to have the expected protective effect on the child during epidemics (Niewiesk, 2014 ; Palloni, 1990 : 200). But smallpox, as opposed to other epidemics, is perhaps more efficiently susceptible to maternal protection, the same way the maternal protection is better against measles as opposed to polio (Wodi et Morelli, 2015 : 2). A factor that may contribute to its efficiency is the higher odds for mothers to experience a smallpox epidemic prior to their motherhood, and thus carry the antibodies for it, compared to other less common and cyclical epidemics. However, this cycle of smallpox depends on the period of study. There were more smallpox epidemics in the later study periods, as opposed to

the first half of the timeframe. As the period effects cannot be seen because of stratification, it can be concluded that the smallpox models have evident loss of statistical power due to stratification.

Speaking of time periods, a limit of this study is that all linkage in the database may not be as complete as possible for the period of 1815-1830. Therefore, infant mortality may be underestimated in this time interval. However, since periods are stratified in all the Cox models, effects were controlled.

Pertaining to the epidemic measure, another limit of this study is that the epidemic dates were subjectively chosen by month according to the utmost peak of apparent death on a graph separated between east and west for the colony (Figure A1 and Figure A2). Then, the epidemics were hypothesized to have lasted the entire month. This lack of precision in the calendar dates for epidemics implies that not all the epidemics are observed at the exact time they occurred for each individual. The study assumed that all children selected were exposed to epidemics occurring in the east or west of the colony at the epidemic dates recorded. However, it is possible that some individuals live in a town that was not affected by a certain epidemic, but they are counted in the analyses as if they were at risk, thereby underestimating the epidemic effect. Further, sometimes an epidemic peak started at the end of a month or ended at the beginning of a month, and the way the peaks were identified, it does not permit an exact date in the middle of the month. In other words, lack of precision in the time-varying variable suggests the underestimation of the estimates of epidemic. Thus, epidemics may have an even stronger effect on mortality than observed in this study. Future research may choose to further identify epidemics per region, to better control the equivalent risk of contracting the epidemic hypothesis. It may also choose to look at periods that directly follow an epidemic as well. Remember, Clements and Hussey (2004) discussed another probable selection process, which is higher mortality up to a year after an epidemic due to weakened immunity following an infection. This may be another way the effect of epidemics is underestimated in this thesis.

This study is the first step in exploring the concurrent effects of consanguinity and epidemics on early childhood survival, as well as their interactive consequence. Even with the numerous limits of this explorational study and the conservative choices in methodology, this research on consanguinity, epidemics and early child survival is important for future research,

being that it is the first demographic study on the relation of consanguinity and epidemics on infant and child mortality. This study of consanguinity would not have been possible without the RPQA-IMPQ data that reconstitute hundreds of thousands of individuals and tens of thousands of "family files spread across four or five generations and up to nine generations in some cases" (Dillon et collab., 2017 : 27). The richness and overall completeness of the historical data allowed for the study of these effects in a context where there is no influence of modern medical treatment or vaccination. Modern medicine is very hard to control in a study and its effect will distort findings (Mazan, 2011a). Therefore, the overall historical context of this study must be recognized as this type of research may be difficult to reproduce with modern data. Further, the historical results are relevant in the study of modern epidemics, especially today with the resurgence of interest in extrinsic mortality in demography, particularly due to SARS- CoV-2 (COVID-19).

Despite the limits of this study, the effects of consanguinity and epidemics on childhood mortality are the following: epidemics show their increasing importance with age, distant consanguinity suggests a fairly stable relative risk of mortality that often resembles the reference group, and close consanguinity has varying effects with time hinting at selection processes. Further, the small sample size of close consanguinity, especially at older ages, potentially maximized fluctuations in the results. Perhaps choosing close consanguinity as a measure with four ascending generations, as opposed to three would have increased the sample size sufficiently to allow for clearer results. In fact, a measure calculated like this may further distinguish distant consanguinity from close consanguinity in child mortality and help explore the idea that longer strings of identical genes due to fewer meiosis may have a differential effect on child mortality during crises compared to many short strings of identical genes. However, this thesis is an exploration of epidemics and consanguinity and does not allow for any advances on the source of the signaled genetic disadvantage.

In summary, this explorative thesis does not refute Gauvin's (2016) advance on reduced immunity for consanguineous individuals of Colonial Quebec, but it is not precise enough to fully support it either. This exploration does hint that closely consanguineous children are more selected in epidemic periods if they did not undergo a selection process in the form of excess mortality prior. Research will have to continue to test the validity of this signal of excess

mortality for consanguineous children. Controls would be necessary, either by limiting the context at birth, for example, by following a cohort that was born in an epidemic, or by adding more robust control variables to the model. The underlying selection process has been thoroughly discussed, especially to hypothesize why excess mortality appears to cease in some childhood age groups. Selection seems to occur during infancy and, as this subsample of the population becomes more robust, the consanguineous become "comparatively less susceptible to any form of adversity as [they grow] older" (Willführ et Gagnon, 2013 : 200). Advancement in the research would therefore have to include controlling selection, for instance, by using the Heckman correction (Heckman, 1976, 1979), as well as controlling for other factors such as the unobserved heterogeneity of the non consanguineous group by further categorizing this subpopulation by number of immigrant parents.

BIBLIOGRAPHY

ALLISON, P. D. 2010. *Survival Analysis Using SAS: A Practical Guide*, 2e éd., Cary, NC. SAS Institute Inc.

AMOREVIETA-GENTIL, M. 2010. « Les niveaux et les facteurs déterminants de la mortalité infantile en Nouvelle-France et au début du Régime Anglais (1621-1779) »,Université de Montréal, Montréal.

BAILEY, P. G., N. DELARUE et H. NJOO. 2019. « Tuberculosis », dans *The Canadian Encyclopedia*, Disponible à la page : https://www.thecanadianencyclopedia.ca/en/article/tuberculosis

BARBEAU, M. 2007. Épidémies en Nouvelle-France. Disponible à la page : http://www.genealogieplanete.com/blog/view/id_796/user_/name_Doyen1938/

BARRACLOUGH, H., L. SIMMS et R. GOVINDAN. 2011. « Biostatistics Primer: What a Clinician Ought to Know: Hazard Ratios », *Journal of Thoracic Oncology*, 6, 6 : 978-982.

BATESON, W. et G. MENDEL. 1902. *Mendel's Principles of Heredity: A Defence, with a Translation of Mendel's Original Papers on Hybridisation*, Cambridge. Cambridge University Press. 212 p.

Batten D., P. Schummer et H. Selden (dir.). 2017. « Typhus », dans D. BATTEN, P. SCHUMMER et H. SELDEN (dir.), *Human Diseases and Conditions*, 3rd ed. Farmington Hills, MI. Charles Scribner's Sons : 2099-2102.

BEISE, J. 2004. *The helping and the helpful grandmother - The role of maternal and paternal grandmothers in child mortality in the 17th and 18th century population of French Settlers in Quebec, Canada*, WP-2004-004, Max Planck Institute for Demographic Research, Rostock, Germany.

BENER, A., R. HUSSAIN et A. S. TEEBI. 2007. « Consanguineous Marriages and Their Effects on Common Adult Diseases: Studies from an Endogamous Population », *Medical Principles and Practice*, 16, 4 : 262-267.

BENER, A. et R. R. MOHAMMAD. 2017. « Global distribution of consanguinity and their impact on complex diseases: Genetic disorders from an endogamous population », *Egyptian Journal of Medical Human Genetics*, 18, 4 : 315-320.

BERNARD, P. 1994. « La mortalité différentielle par variole en Nouvelle-France »,Université de Montréal, Montréal.

BITTLES. 2003. « Consanguineous marriage and childhood health », *Developmental Medicine & Child Neurology*, 45, 8 : 571-576.

BITTLES et M. L. BLACK. 2010b. « Consanguineous Marriage and Human Evolution », *Annual Review of Anthropology*, 39 : 193-207.

BITTLES et M. L. BLACK. 2010a. « Consanguinity, human evolution, and complex diseases », *Proceedings of the National Academy of Sciences*, 107, suppl 1 : 1779-1786.

BITTLES et J. V. NEEL. 1994. « The costs of human inbreeding and their implications for variations at the DNA level », *Nature Genetics*, 8, 2 : 117.

BOISVERT, M. et F. M. MAYER. 1994. « Mortalité infantile et consanguinité dans une population endogame du Québec », *Population (French Edition)*, 49, 3 : 685-724.

BOUCHARD, G., H. CHARBONNEAU, B. DESJARDINS, E. HEYER et M. TREMBLAY. 1995. « Mobilité géographique et stratification du pool génique canadien-français sous le Régime français », dans *Les chemins de la migration en Belgique et au Québec, XVIIe - XXe siècles*, Louvain-la-Neuve etc. Éditions Académia [etc.] : 51-59.

BOUCHARD, G. et M. DE BRAEKELEER. 1991a. « Histoire d'un génome. Population et génétique dans l'est du Québec »,Université du Québec, Québec.

BOUCHARD, G. et M. DE BRAEKELEER. 1991b. « Mouvements migratoires, effets fondateurs et homogénéisation génétique », dans *Histoire d'un génôme: population et génétique dans l'est du Québec*, Sillery. Presses de l'Université du Québec : 283-321.

BRUCKNER, T. A., S. GAILEY, S. HALLMAN, M. AMOREVIETA-GENTIL, L. DILLON et A. GAGNON. 2018. « Epidemic cycles and environmental pressure in colonial Quebec », *American Journal of Human Biology*, 30, 5 : e23155.

CADOTTE, M. 2013. « Epidemics in Canada », dans *The Canadian Encyclopedia*, Disponible à la page : https://www.thecanadianencyclopedia.ca/en/article/epidemic

CANADA. DEPT. OF AGRICULTURE. 1873. *Census of Canada. 1870-71*, Ottawa : Printed by I. B. Taylor [&] Maclean, Rogers & co. 544 p.

CATALANO, R. et T. BRUCKNER. 2006. « Secondary sex ratios and male lifespan: Damaged or culled cohorts », *Proceedings of the National Academy of Sciences*, 103, 5 : 1639-1643.

CATALANO, R., T. BRUCKNER, T. HARTIG et M. ONG. 2005. « Population stress and the Swedish sex ratio », *Paediatric and Perinatal Epidemiology*, 19, 6 : 413-420.

CAVALLI-SFORZA, L. L. et W. F. BODMER. 1971. *The Genetics of Human Populations*, 965 p.

CHARBONNEAU, H., B. DESJARDINS, A. GUILLEMETTE, Y. LANDRY, J. LÉGARÉ, F. NAULT, R. BATES et P. COLOZZO. 1993. *The first French Canadians : pioneers in the St. Lawrence Valley*, Newark. University of Delaware Press. 236 p.

CHARBONNEAU, H., B. DESJARDINS, J. LÉGARÉ et H. DENIS. 2000. « The Population of the St. Lawrence Valley, 1608-1760 », dans *A Population History of North America*, New York. Cambridge University Press : 99-142.

CLEMENTS, C. J. et G. D. HUSSEY. 2004. « Measles », dans *The global epidemiology of infectious diseases*, World Health Organization : 75-110.

CRIMMINS, E. M. et C. E. FINCH. 2006. « Infection, inflammation, height, and longevity », *Proceedings of the National Academy of Sciences*, 103, 2 : 498-503.

DECHAMBRE, A. 1885. *Dictionnaire encyclopédique des sciences médicales*, Paris. V. Masson Et Fils. 818 p.

DELÂGE, D. 2006. « Microbes, animaux et eau en Nouvelle-France », *Globe: revue internationale d'études québécoises*, 9, 1 : 113-139.

DENIC, S., N. NAGELKERKE et M. M. AGARWAL. 2008. « Consanguineous marriages and endemic malaria: can inbreeding increase population fitness? », *Malaria Journal*, 7, 1 : 150.

DÉPATIE, S. 1988. « L'évolution d'une société rurale : l'Île Jésus au XVIIIe siècle », (Doctoral Thesis), McGill University, Montréal.

DESJARDINS, B. 1999. « Le registre de la population du Québec ancien », *Annales de Démographie Historique*, 1998, 2 : 215-226.

DESJARDINS, B. 1996. « Demographic Aspects of the 1702-1703 Smallpox Epidemic in the St. Lawrence Valley », *Canadian Studies in Population*, 23, 1 : 49.

DILLON, L. 2016. « 1.2 Historical Demography of Canada, 1608-1921 », dans *Canadian History: Post-Confederation*, Montréal, Québec. .

DILLON, L., M. AMOREVIETA-GENTIL, M. CARON, C. LEWIS, A. GUAY-GIROUX, B. DESJARDINS et A. GAGNON. 2017. « The Programme de recherche en démographie

historique : past, present and future developments in family reconstitution », *The History of the Family*, 23 : 20-53.

EMOND, M. 1992. « Effets de la consanguinité sur la fécondité et la mortalité infantile au Saguenay-Lac-St-Jean (Québec, Canada) », (Master thesis), Université du Québec à Chicoutimi, Québec.

ENGELHARDT, S. C., P. BERGERON, A. GAGNON, L. DILLON et F. PELLETIER. 2019. « Using Geographic Distance as a Potential Proxy for Help in the Assessment of the Grandmother Hypothesis », *Current Biology*, 29, 4 : 651-656.

GAGNON, A. 2012. « Effect of birth season on longevity: Thrifty and hopeful phenotypes in historical quebec », *American Journal of Human Biology*, 24, 5 : 654-660.

GAGNON, A. 2000. « Aux origines du génome québécois: démographie génétique d'une population nouvelle », (Doctoral Thesis), Université de Montréal, Montréal.

GAGNON, A., J. BEISE et J. W. VAUPEL. 2005. « Genome-wide identity-by-descent sharing among CEPH siblings », *Genetic Epidemiology*, 29, 3 : 215-224.

GAGNON, A. et E. HEYER. 2001. « Fragmentation of the Québec population genetic pool (Canada): Evidence from the genetic contribution of founders per region in the 17th and 18th centuries », *American Journal of Physical Anthropology*, 114, 1 : 30-41.

GAGNON, A. et R. MAZAN. 2009. « Does exposure to infectious diseases in infancy affect old-age mortality? Evidence from a pre-industrial population », *Social Science & Medicine*, 68, 9 : 1609-1616.

GAUVIN, H. 2016. « The French Canadian founder population : lessons and insights for genetic epidemiological research »,Université de Montréal, Montréal.

GAUVIN, H., J.-F. LEFEBVRE, C. MOREAU, E.-M. LAVOIE, D. LABUDA, H. VÉZINA et M.-H. ROY-GAGNON. 2015. « GENLIB: An R package for the analysis of genealogical data », *BMC Bioinformatics*, 16, 160

GAUVREAU, D., P. THORNTON et H. VÉZINA. 2010. « Le jumelage des recensements aux mariages du fichier balsac : présentation de l'approche et étude exploratoire des enfants de couples mixtes à la fin du xixe siècle », *Cahiers québécois de démographie*, 39, 2 : 357-381.

GOVERNMENT OF CANADA, S. C. 2020. Mortality rates, by age group. Disponible à la page : https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310071001

GOVERNMENT OF CANADA, S. C. 2019. Methods for Constructing Life Tables for Canada, Provinces and Territories. Disponible à la page : https://www150.statcan.gc.ca/n1/pub/84-538-x/84-538-x2019001-eng.htm

HAMILTON, W. D. 1964. « The genetical evolution of social behaviour. I », *Journal of Theoretical Biology*, 7, 1 : 1-16.

HARRIS, C. et G. J. MATTHEWS. 1987. *Historical atlas of Canada. Volume 1, From the beginning to 1800*, Toronto. University of Toronto Press.

HECKMAN, J. 1976. « The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models », *Annals of Economic and Social Measurement,* 5, 4 : 475-492.

HECKMAN, J. J. 1979. « Sample Selection Bias as a Specification Error », *Econometrica*, 47, 1 : 153-161.

HENRY, L. et A. BLUM. 1988. *Techniques d'analyse en démographie historique*, L'institut national d'études démographique. Paris. . 180 p.

JACQUARD, A. 1974. *The Genetic Structure of Populations*, Berlin Heidelberg. Springer-Verlag.

JACQUARD, A. et F. REYNÈS. 1968. « Mesure démographique du fardeau génétique », *Population*, 23, 4 : 625-648.

KELLER, L. F. et D. M. WALLER. 2002. « Inbreeding effects in wild populations », *Trends in Ecology & Evolution*, 17, 5 : 230-241.

KOELLNER, C. M., K. A. MENSINK et W. E. HIGHSMITH. 2018. « Chapter 5 - Basic Concepts in Human Molecular Genetics », dans W. B. COLEMAN et G. J. TSONGALIS (dir.), *Molecular Pathology*, 2e éd., Academic Press : 99-120.

LALOU, R. 1997. « Endogenous Mortality in New France: At the Crossroads of Natural and Social Selection », dans A. BIDEAU, B. DESJARDINS et H. PEREZ BRIGNOLI (dir.), *Infant and child mortality in the past*, Clarendon Press. Oxford University Press : 203-215.

LALOU, R. 1990. « Des enfants pour le paradis : la mortalité des nouveau-nés en Nouvelle-France », (Doctoral Thesis), Université de Montréal, Montréal.

LYONS, E. J., A. J. FRODSHAM, L. ZHANG, A. V. S. HILL et W. AMOS. 2009. « Consanguinity and susceptibility to infectious diseases in humans », *Biology Letters*, 5, 4 : 574-576.

MAHEU, S. 2001. « Impacts de la consanguinité maternelle sur la mortalité de la progéniture dans une population endogame du Québec », (Master thesis), Université du Québec à Montréal, Montréal (Québec, Canada).

MALÉCOT, G. et L. BLARINGHEM. 1948. *Les mathématiques de l'hérédité*, Paris. Masson et Cie.

MAYER, F. M. 1993. « BOUCHARD, Gérard et Marc DeBRAEKELEER, Histoire d'un génôme. Population et génétique dans l'est du Québec. Québec, Presses de l'Université du Québec, 1991. xxv-607 p. », *Revue d'histoire de l'Amérique française*, 46, 3 : 503.

MAYR, E. 1904-2005. 1942. *Systematics and the origin of species from the view-point of a zoologist.*, N.Y. Columbia University Press.

MAZAN, R. 2012. « Delayed measles mortality among exposed children who survived the epidemic of 1714-15 in New France », *Canadian Studies in Population*, 39 : 9-22.

MAZAN, R. 2011a. « Analyzing Epidemics in New France: The Measles Epidemic of 1714-15 »,The University of Western Ontario,

MAZAN, R. 2011b. « Risk factors of mortality among French Canadian children during the measles epidemic of 1714-1715 », *Genus*, 67, 2 : 31-56.

MAZAN, R., A. GAGNON et B. DESJARDINS. 2009. « The Measles Epidemic of 1714-1715 in New France », *Canadian Studies in Population*, 36

MCGUIGAN, P. 2016. 1816: The Year Without Summer. Disponible à la page : http://www.canadashistory.ca/explore/environment/1816-the-year-without-summer

MCINNIS, M. 2000. « The Population of Canada in the Nineteenth Century », dans *A Population History of North America*, New York. Cambridge University Press : 371-432.

MILOT, E., C. MOREAU, A. GAGNON, A. A. COHEN, B. BRAIS et D. LABUDA. 2017. « Mother's curse neutralizes natural selection against a human genetic disease over three centuries », *Nature Ecology & Evolution*, 1, 9 : 1400-1406.

MORGAN, K. 2016. « Population Genetics », dans *The Canadian Encyclopedia*, Disponible à la page : https://www.thecanadianencyclopedia.ca/en/article/population-genetics

NAULT, F., B. DESJARDINS et J. LÉGARÉ. 1990. « Effects of reproductive behaviour on infant mortality of French-Canadians during the seventeenth and eighteenth centuries », *Population Studies*, 44, 2 : 273-285.

NEEL, J. V. 1962. « Diabetes Mellitus: A “Thrifty” Genotype Rendered Detrimental by “Progress”? », *American Journal of Human Genetics*, 14, 4 : 353-362.

NIEWIESK, S. 2014. « Maternal Antibodies: Clinical Significance, Mechanism of Interference with Immune Responses, and Possible Vaccination Strategies », *Frontiers in Immunology*, 5 : 446.

PALLONI, A. 1990. « Assessing the levels and impact of mortality in crisis situations », dans *Measurement and analysis of mortality: new approaches*, Oxford. Clarendon Press : 195-228.

PÂQUET, M. 1999. « «Diminuer le danger par de bons règlemens intérieurs». État colonial et contrôle médical des migrations au Bas-Canada et au Canada-Uni, 1795-1854 », *Canadian Bulletin of Medical History*, 16, 2 : 271-291.

PAVARD, S., A. GAGNON, B. DESJARDINS et E. HEYER. 2005. « Mother’s death and child survival: the case of early Quebec », *Journal of Biosocial Science*, 37, 2 : 209-227.

PERRY, R. T. et N. A. HALSEY. 2004. « The Clinical Significance of Measles: A Review », *The Journal of Infectious Diseases*, 189 : S4-S16.

PRESSAT, R. 1983. *L’analyse démographique*, Paris. P.U.F. 295 p.

R CORE TEAM. 2021. *R: A language and environment for statistical computing*, Vienna, Austria. R Foundation for Statistical Computing.

RÉDEI, G. P. 2008a. « Inbreeding Coefficient », dans *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Disponible à la page : https://link.springer.com/referenceworkentry/10.1007/978-1-4020-6754-9_8397

RÉDEI, G. P. 2008b. « Locus », dans *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Disponible à la page : https://link.springer.com/referenceworkentry/10.1007%2F978-1-4020-6754-9_9530

RÉDEI, G. P. 2008c. « Inbreeding Depression », dans *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*, Disponible à la page : https://link.springer.com/referenceworkentry/10.1007/978-1-4020-6754-9_8398

ROBERT, A., B. TOUPANCE, M. TREMBLAY et E. HEYER. 2009. « Impact of inbreeding on fertility in a pre-industrial population », *European Journal of Human Genetics*, 17, 5 : 673-681.

SANGHVI, L. D. 1966. « Inbreeding in India », *Eugenics Quarterly*, 13, 4 : 291-301.

SCITABLE BY NATURE EDUCATION. 2014. Allele | Learn Science at Scitable. Disponible à la page : https://www.nature.com/scitable/definition/allele-48/

SIEFF, D. F., L. BETZIG, L. CRONK, A. G. FIX, M. FLINN, L. SATTENSPIEL, K. GIBSON, D. A. HERRING, N. HOWELL, S. R. JOHANSSON, Z. PAVLÍK, J. W. SHEETS, E. A. SMITH, E. VOLAND et E. SIEGELKOW. 1990. « Explaining Biased Sex Ratios in Human Populations: A Critique of Recent Studies [and Comments and Reply] », *Current Anthropology*, 31, 1 : 25-48.

SPAULDING, W. B. et M. FOSTER-SANCHEZ. 2019. « Smallpox in Canada », dans *The Canadian Encyclopedia*, Disponible à la page : https://thecanadianencyclopedia.ca/en/article/smallpox

STATACORP. 2021. *Survival Analysis Reference Manual Release 17*, College Station, Texas. Stata Press. 537 p.

STOLTENBERG, C., P. MAGNUS, A. SKRONDAL et R. T. LIE. 1999. « Consanguinity and recurrence risk of stillbirth and infant death. », *American Journal of Public Health*, 89, 4 : 517-523.

STOMMEL, H. M. et E. STOMMEL. 1983. *Volcano weather: the story of 1816, the year without a summer*, Newport, R.I. Seven Seas Press.

THORNTON, P. A. et S. OLSON. 2011. *Peopling the North American city : Montreal, 1840-1900*, Montreal. McGill-Queen's University Press. 524 p.

TRUDEL, M. 1968. *Initiation à la Nouvelle-France : histoire et institutions*, Montréal. Holt, Rinehart et Winston.

UNICEF. 2021. Child Mortality. Disponible à la page : https://data.unicef.org/topic/child-survival/under-five-mortality/

VÉZINA, H., M. TREMBLAY et L. HOUDE. 2004. « Mesures de l'apparentement biologique au Saguenay-Lac-St-Jean (Québec, Canada) à partir de reconstitutions généalogiques », *Annales de demographie historique*, no 108, 2 : 67-83.

VOLAND, E. et J. BEISE. 2002. « Opposite Effects of Maternal and Paternal Grandmothers on Infant Survival in Historical Krummhörn », *Behavioral Ecology and Sociobiology*, 52, 6 : 435-443.

WARREN, J.-P. 2007. « L'invention du Canada français: le rôle de l'Église catholique », dans M. PÂQUET et S. SAVARD (dir.), *Balises et références. Acadies, francophonies*, Presses de l'Université Laval : 21-56.

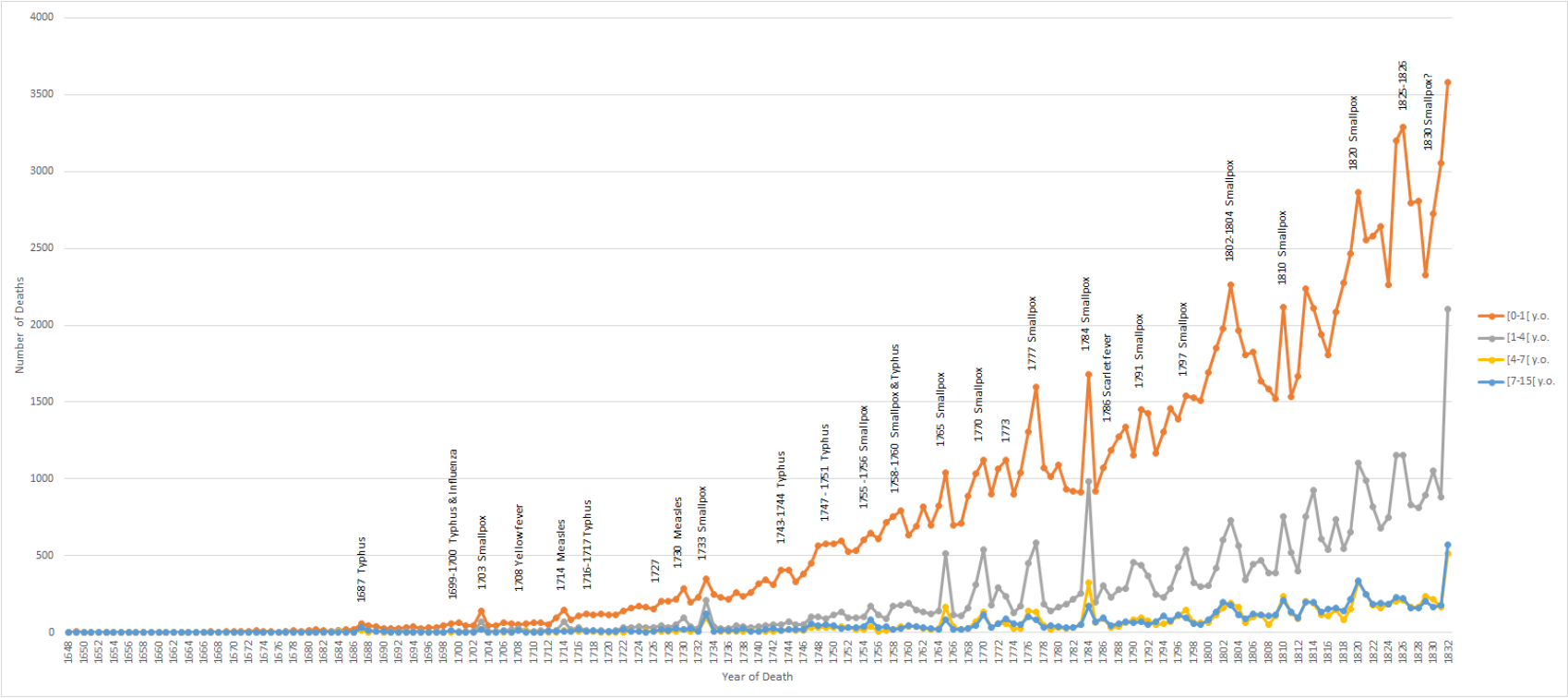
WILLFÜHR, K. P. et A. GAGNON. 2013. « Are Stepparents Always Evil? Parental Death, Remarriage, and Child Survival in Demographically Saturated Krummhörn (1720–1859) and Expanding Québec (1670–1750) », *Biodemography and Social Biology*, 59, 2 : 191-211.

WODI, A. P. et V. MORELLI. 2015. « Principles of Vaccination », dans J. HAMBORSKY, A. KROGER et C. (Skip) WOLFE (dir.), *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 13e éd., Washington D.C. Public Health Foundation : 1-8.

1982. « Le programme de recherche en démographie historique du département de l'Université de Montréal », *Annales de Démographie Historique*, 1982, 1 : 375-382.

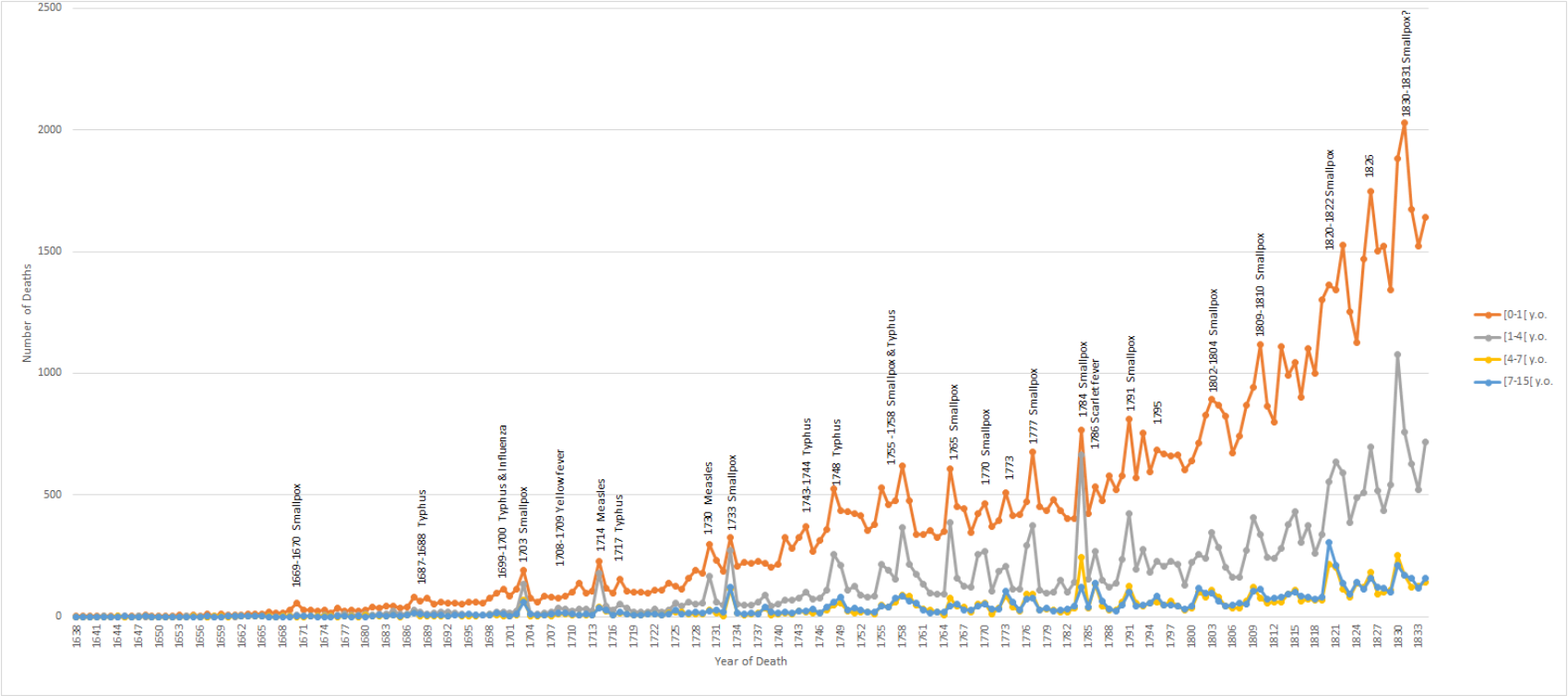
ANNEX I

Figure A1. Identifying Epidemics by Number of Deaths per Year, Children 0 to 15 years old, West Colonial Quebec, 1648-1832



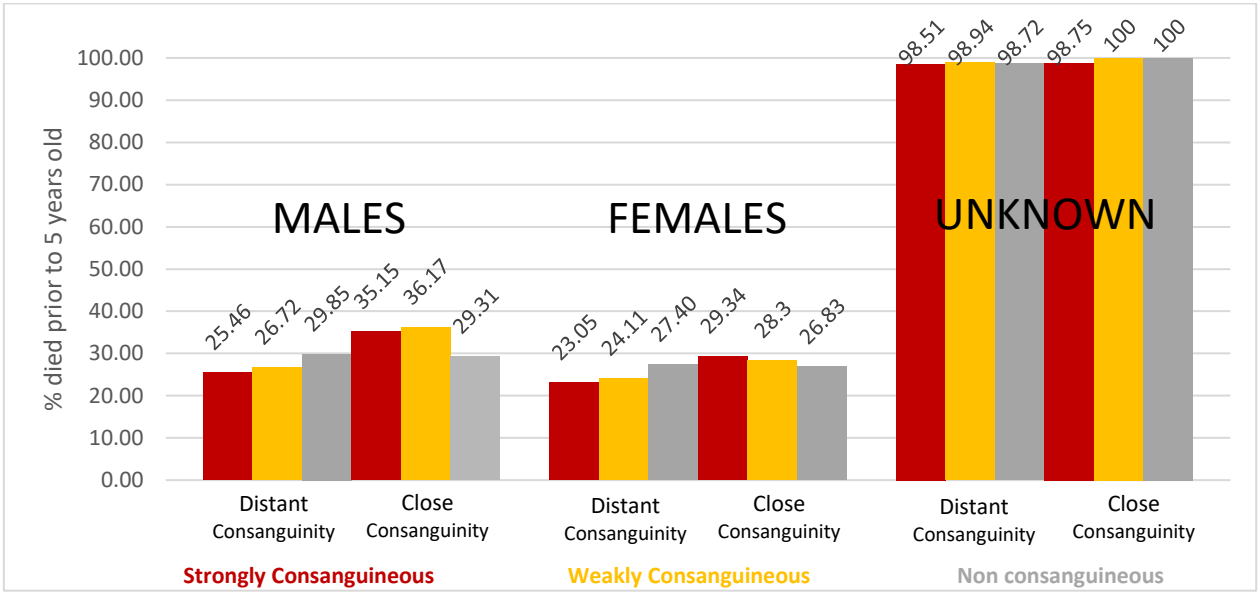
Generated by Excel with RPQA-IMPQ data.

Figure A2. Identifying Epidemics by Number of Deaths per Year, Children 0 to 15 years old, East Colonial Quebec, 1648-1832



Generated by Excel with RPQA-IMPQ data.

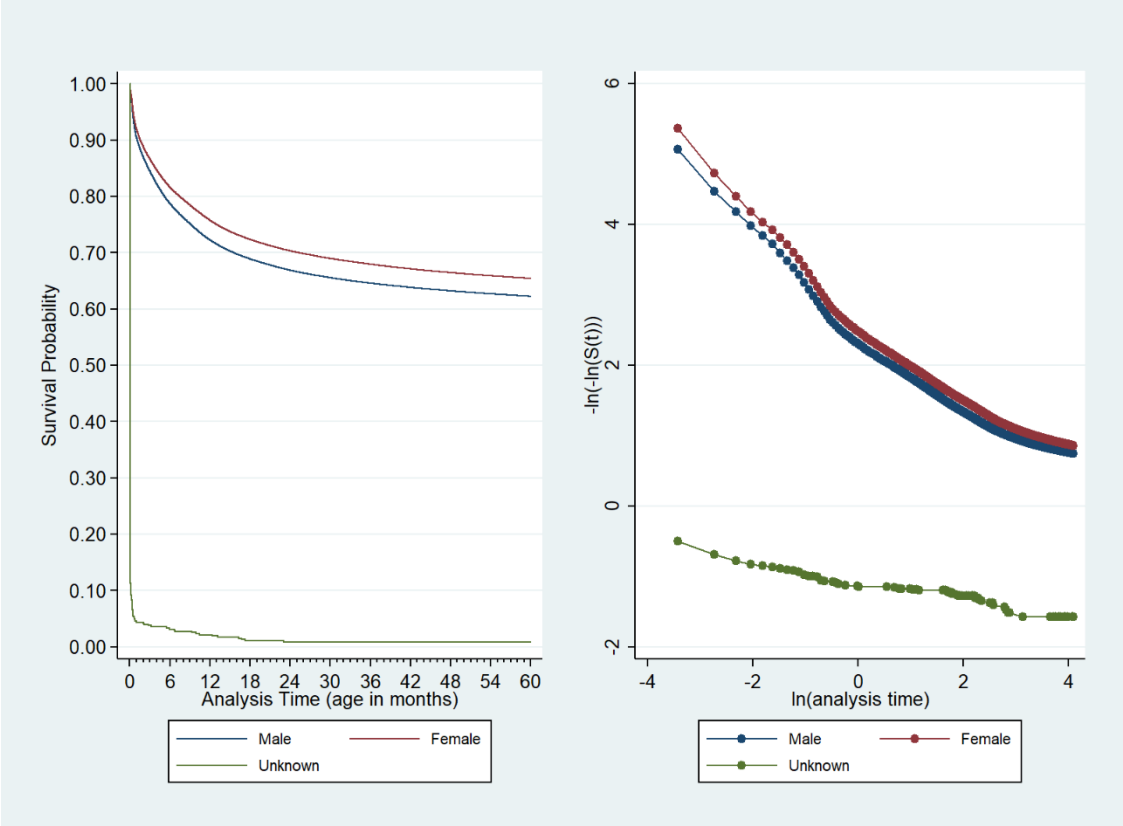
Figure A3. Death Prior to Age 5 per Level of Consanguinity by Sex and Type of Consanguinity, Frequencies and Percentages[1], Colonial Quebec, 1720-1830



Source: RPQA-IMPQ.

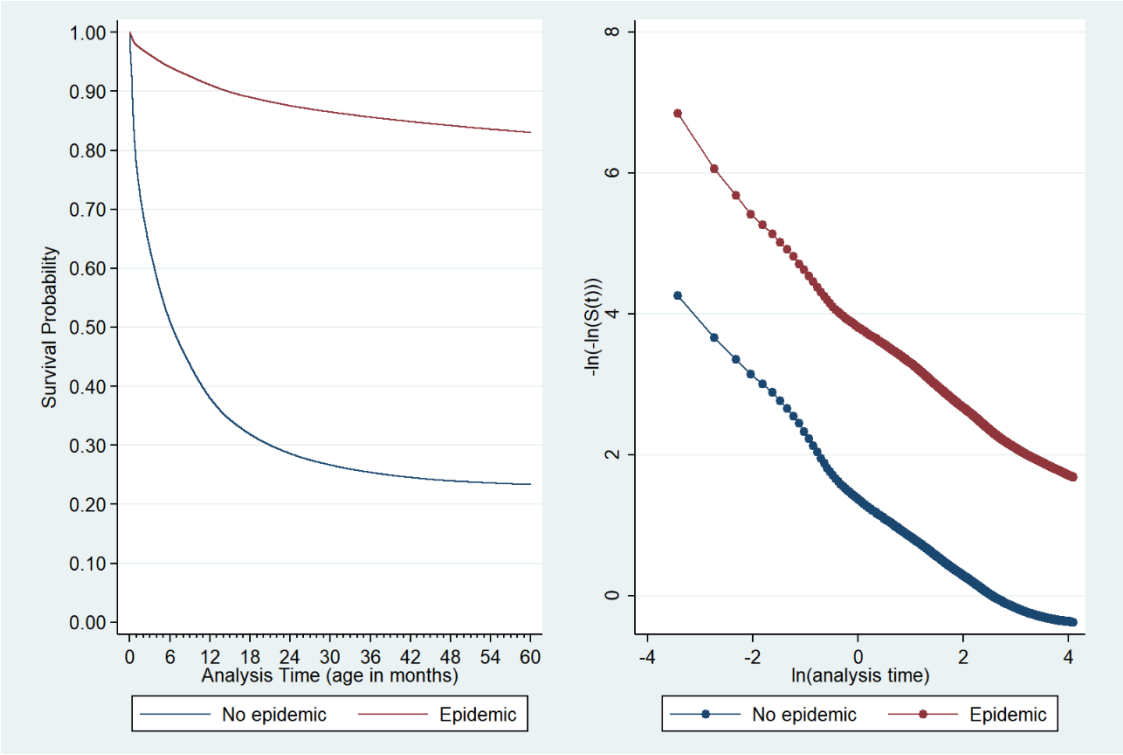
[1]Percentages are based on the number of boys, girls and unknown sexes in each type and level of consanguinity (see Table VI). For example, when observing close consanguinity, 35.15% of (625) strongly consanguineous boys die prior to 5 years old. Note: children who died on their date of birth are included in this table.

Figure A4. Kaplan-Meier Survival Curves by Sex, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

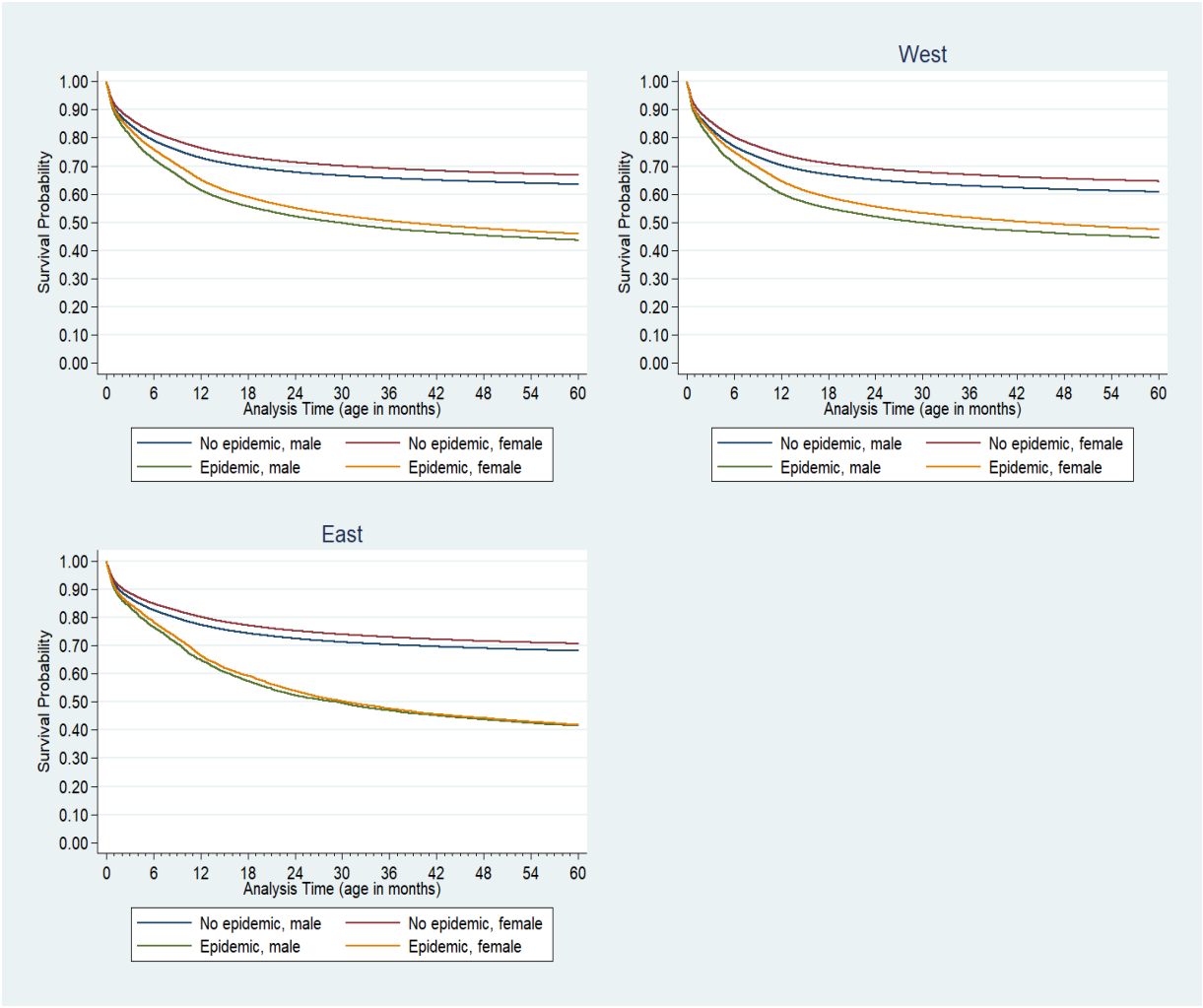
Figure A5. Kaplan-Meier Survival Curves by "Ever lived an Epidemic" Variable, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

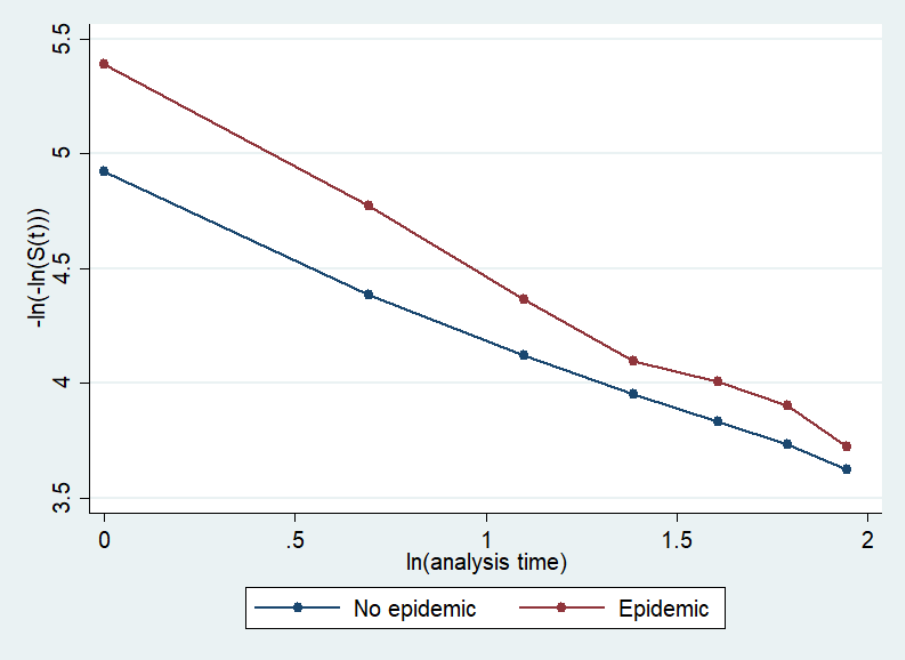
For the epidemic variable calculated as ever having lived an epidemic, children having lived an epidemic have better survival probabilities at all ages. You may think it counterintuitive to see children having lived an epidemic at some time during the observation period at lower risk of dying compared to the children not having experienced an epidemic during the first five years of their life. However, the children in "no epidemic" include children who died prior to having the chance to enter an epidemic. Therefore, it is logical that children who survived the longest lived through an epidemic and have better survival probabilities than children who did not survive long enough to experience an epidemic. This variable is not used in the Cox analysis as it puts children having experienced an epidemic at 2 months old at the same level of risk as children having experienced an epidemic for the first time the day before the end of observation (date at censorship). Therefore, using epidemics as a time varying variable is key in this study.

Figure A6. Kaplan-Meier Survival Curves by Epidemics and Sex for the Whole Colony and for two Cardinal Directions, East and West, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

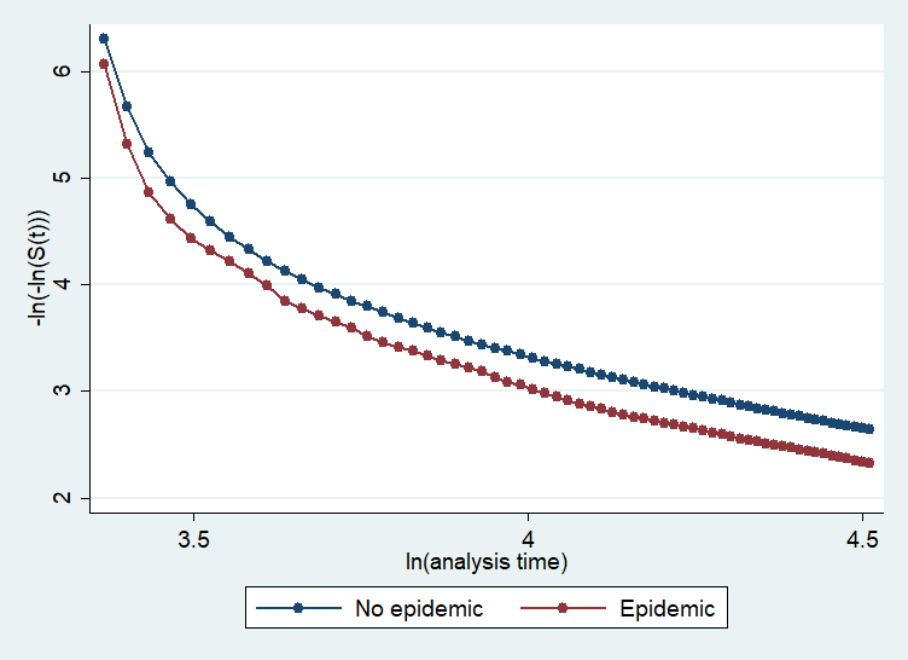
Figure A7. Graphical Test of Proportional Hazards: Epidemic, Distantly Consanguineous Males, 0-7 Days Old, Adjusted for Period and Region-Urban, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

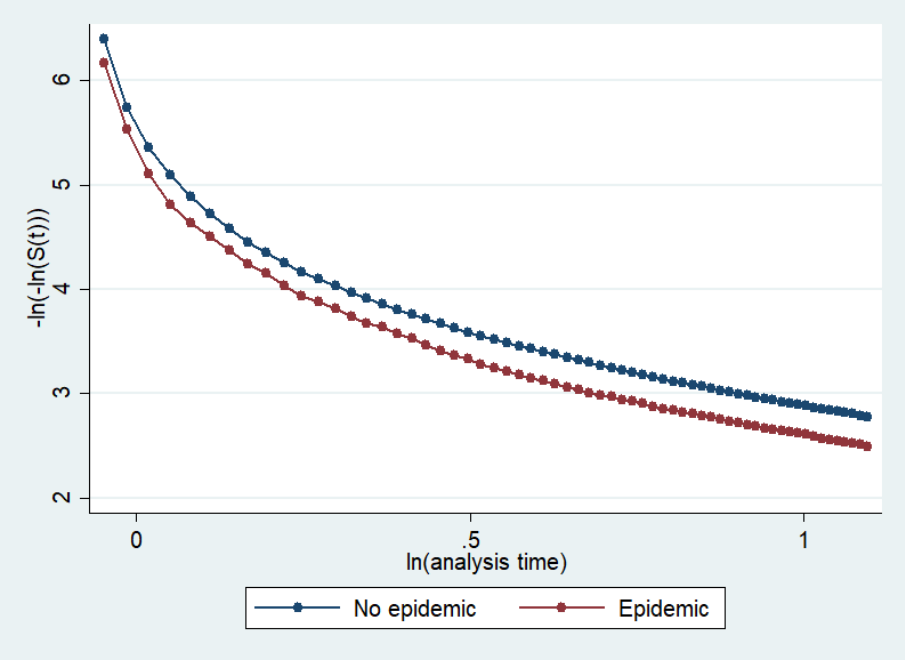
The interaction between epidemic and distant consanguinity is not proportional under Schoenfeld residuals for boys when looking at the early neonatal period. When we observe the log-negative-log functions, we understand why; epidemics have a slightly larger impact on the hazard for distantly consanguineous boys in the first few days after birth compared to the end of the interval (at seven days) (see Figure A7). This slight difference in distance between the curves is a minor violation of the proportional hazards assumption so the HR can be interpreted as an averaged relative effect in the first week of life (Barraclough et collab., 2011).

Figure A8. Graphical Test of Proportional Hazards: Epidemic, Males 1-3 Months Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



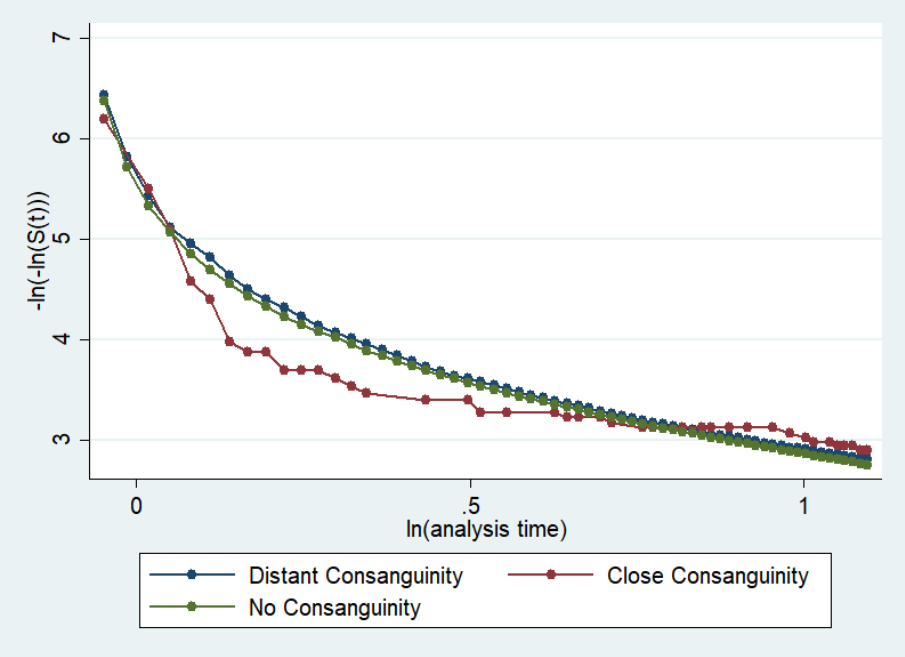
Generated by STATA using RPQA-IMPQ data.

Figure A9. Graphical Test of Proportional Hazards: Epidemic, Females 1-3 Months Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



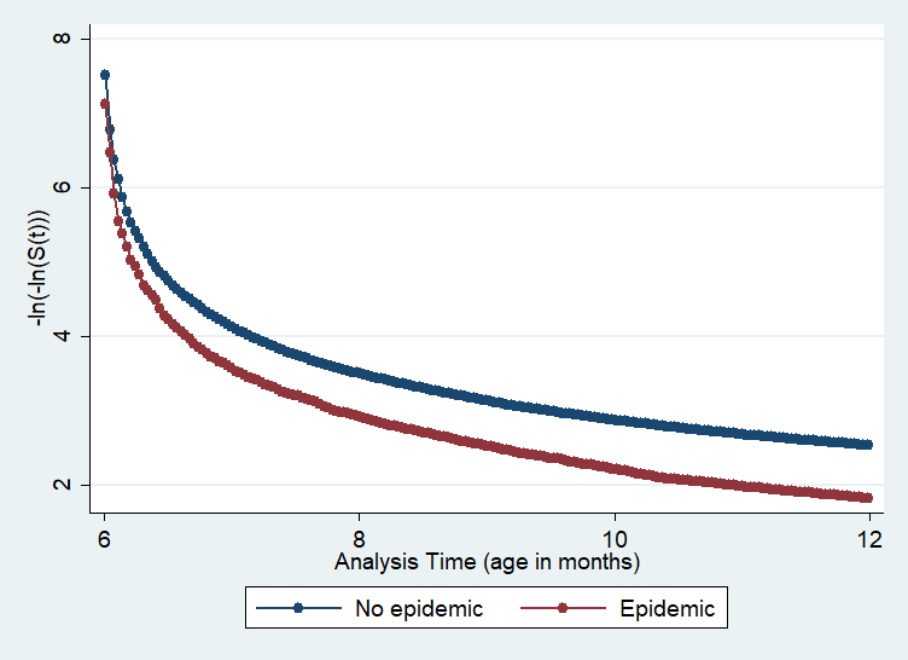
Generated by STATA using RPQA-IMPQ data.

Figure A10. Graphical Test of Proportional Hazards: Consanguinity, Females 1-3 Months Old, Adjusted for Epidemic, Period, Region-Urban, Colonial Quebec, 1720-1830



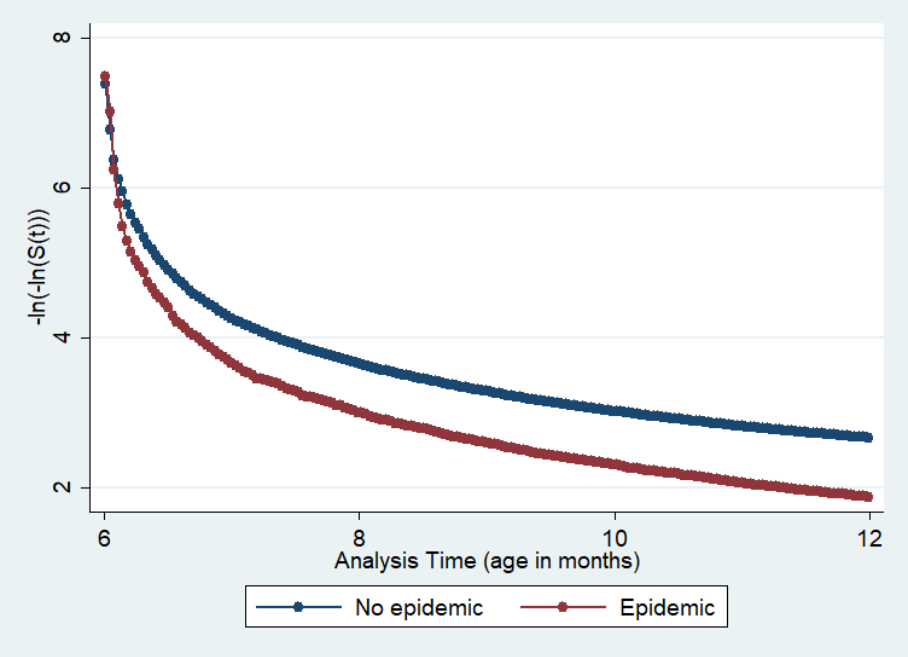
Generated by STATA using RPQA-IMPQ data.

Figure A11. Graphical Test of Proportional Hazards: Epidemic, Male, 6 Months to 1 Year Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



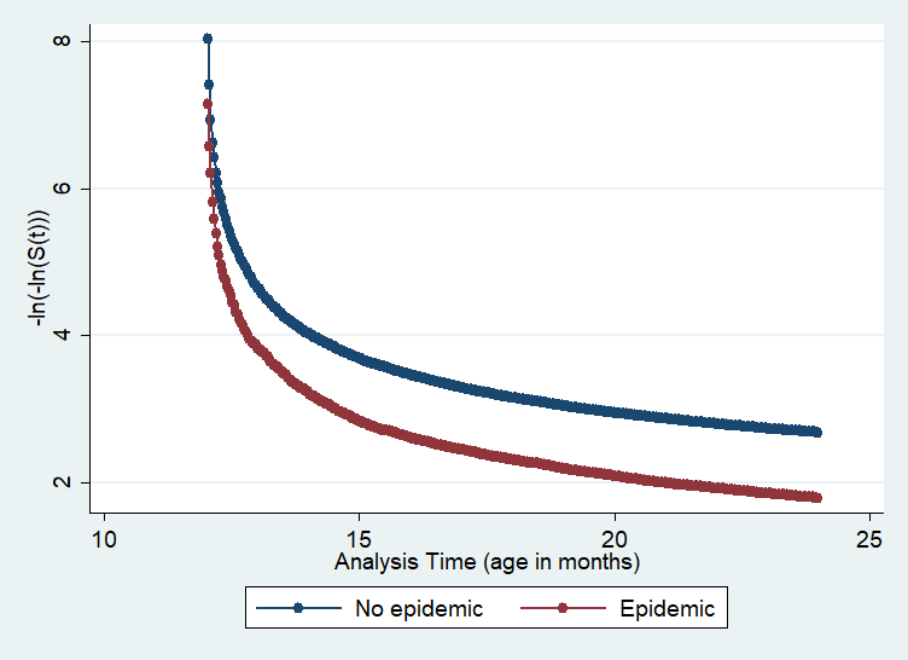
Generated by STATA using RPQA-IMPQ data.

Figure A12. Graphical Test of Proportional Hazards: Epidemic, Females, 6 Months to 1 Year Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



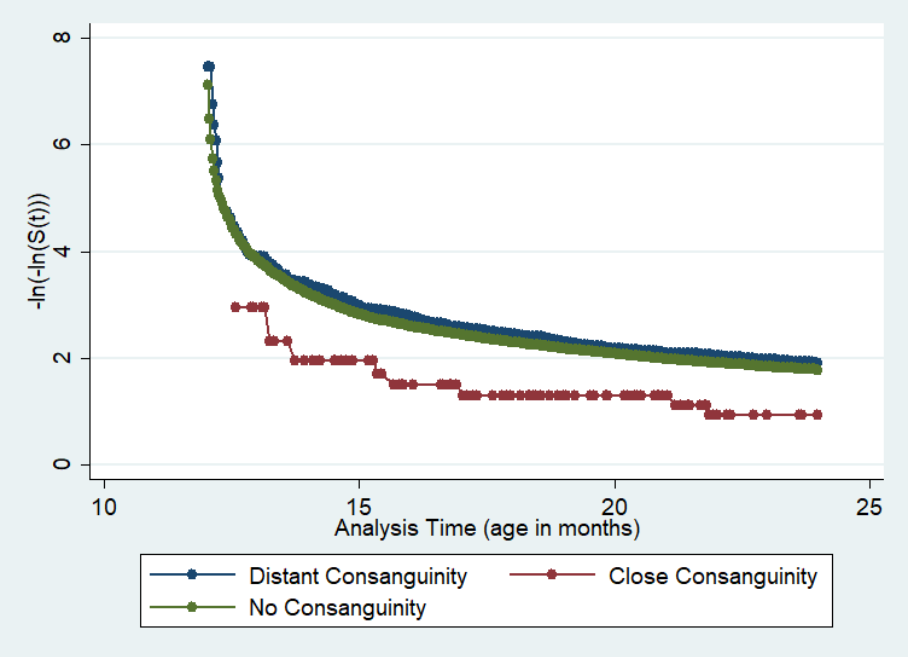
Generated by STATA using RPQA-IMPQ data.

Figure A13. Graphical Test of Proportional Hazards: Epidemic, Females 1-2 Years Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



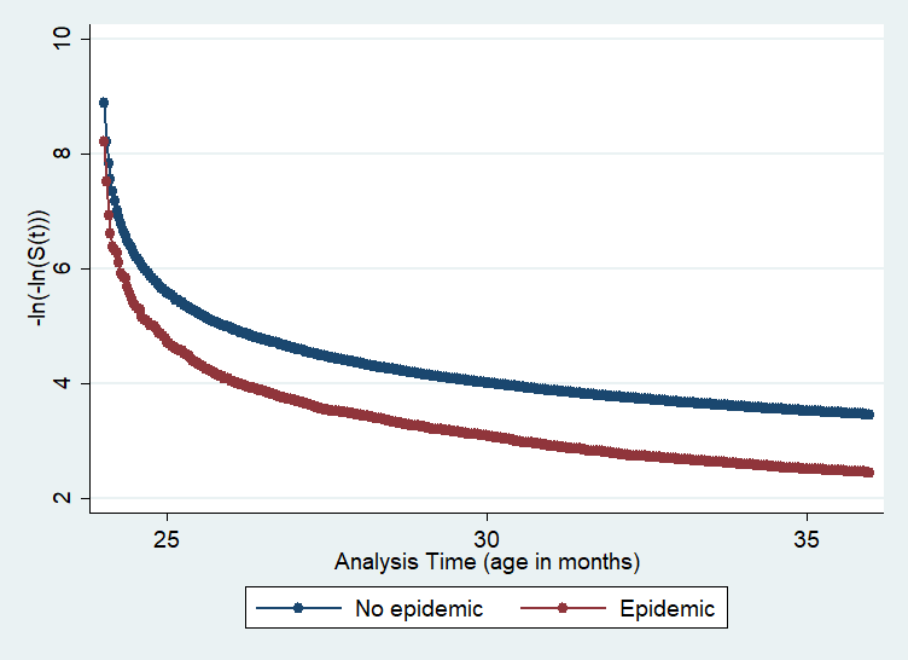
Generated by STATA using RPQA-IMPQ data.

Figure A14. Graphical Test of Proportional Hazards: Consanguinity, Females 1-2 Years Old in an Epidemic, Adjusted for Period, Region-Urban, Colonial Quebec, 1720-1830



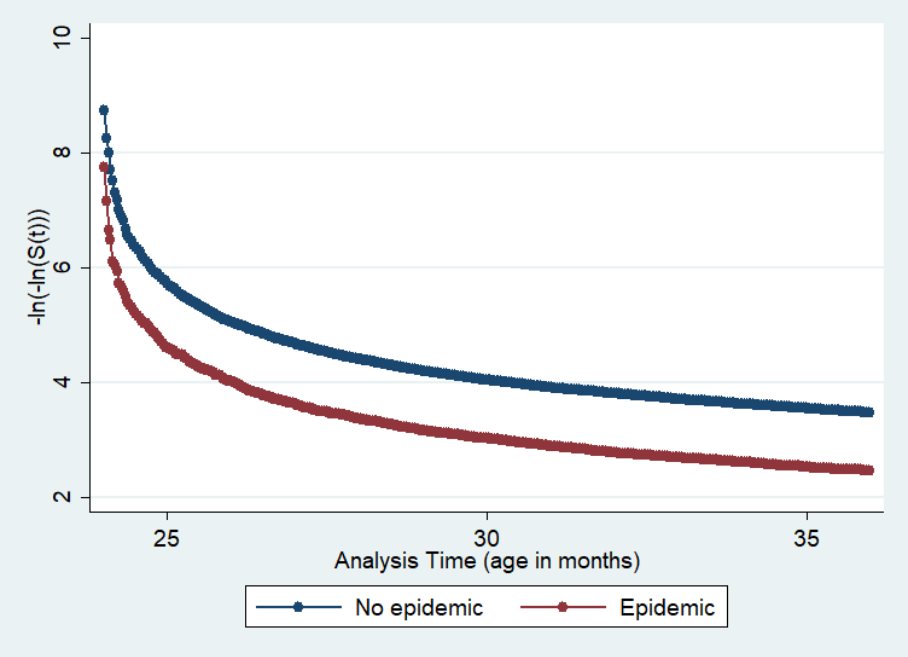
Generated by STATA using RPQA-IMPQ data.

Figure A15. Graphical Test of Proportional Hazards: Epidemic, Males 2-3 Years Old, Adjusted for Consanguinity, Period, Region-Urban, Colonial Quebec, 1720-1830



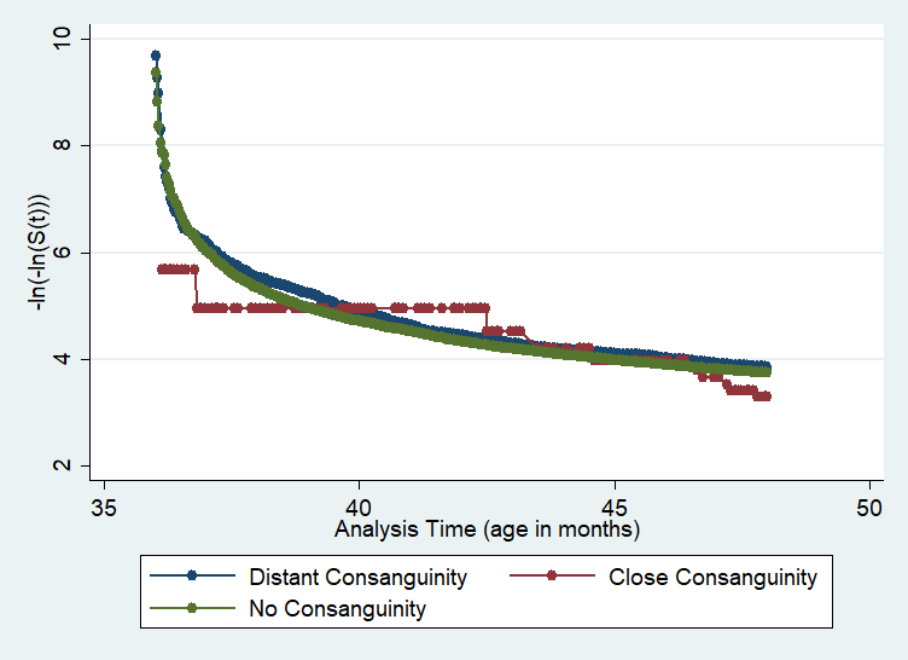
Generated by STATA using RPQA-IMPQ data.

Figure A16. Graphical Test of Proportional Hazards: Epidemic, Distantly Consanguineous Females, 2-3 Years Old, Adjusted for Period, Region-Urban, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

Figure A17. Graphical Test of Proportional Hazards: Consanguinity, Males, 3-4 Years Old, Adjusted for Epidemic, Period, Region-Urban, Colonial Quebec, 1720-1830



Generated by STATA using RPQA-IMPQ data.

Table A1. Consanguinity, Smallpox and Their Interaction, Stratified[1] Cox Models Per Sex, Neonatal Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[0 days, 7 days[	**Male**	Mod. 1S	1.069*	1.691*				310,961	312,569	-70,876	141,755	141,776
		Mod. 2S (NP)			*1.085*				312,569	-70,879	141,760	141,771
		Mod. 3S	1.069*	1.693*	*1.086*				312,569	-70,874	141,755	141,787
		Mod. 4S	1.079*	1.761**	*1.122**	0.819	-		312,569	-70,873	141,755	141,809
		% lost										
[0 days, 7 days[	**Female**	Mod. 1S	1.043	2.085***				299,451	301,022	-54,972	109,948	109,970
		Mod. 2S			1.238***				301,022	-54,971	109,944	109,955
		Mod. 3S	1.043	2.084***	1.238***				301,022	-54,966	109,937	109,969
		Mod. 4S	1.056	2.120***	1.283***	0.793	*0.713*		301,022	-54,965	109,939	109,992
		% lost	-	-								
[7 days, 28 days[	**Male**	Mod. 1S	1.014	1.357*				292,655	297,907	-165,370	330,743	330,764
		Mod. 2S			0.979				297,907	-165,371	330,745	330,755
		Mod. 3S	1.014	1.357*	0.979				297,907	-165,369	330,745	330,777
		Mod. 4S	1.011	1.344†	0.970	1.062	1.265		297,907	-165,369	330,748	330,801
	% lost (6.1%)		4.7%	7.7%								
[7 days, 28 days[	**Female**	Mod. 1S	1.059*	1.263				284,162	289,251	-136,035	272,073	272,094
		Mod. 2S			1.041				289,251	-136,038	272,078	272,088
		Mod. 3S	1.059*	1.264	1.040				289,251	-136,034	272,074	272,106
		Mod. 4S	1.056*	1.239	1.030	1.054	1.503		289,251	-136,034	272,078	272,130
	% lost (5.3%)		4.3%	7.8%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_t}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals.

Table A2. Consanguinity, Smallpox and Their Interaction, Stratified[1] Cox Models Per Sex, Post-Neonatal Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[28 days, 3 months[	**Male**	Mod. 1S	1.016	1.048				273,052	287,815	-165,935	331,875	331,896
		Mod. 2S			1.129***				287,815	-165,929	331,860	331,871
		Mod. 3S	1.016	1.048	1.129***				287,815	-165,929	331,864	331,896
		Mod. 4S	1.023	1.009	1.154***	0.859	1.657		287,815	-165,927	331,865	331,917
	% lost (6.8%)		6.2%	8.7%								
[28 days, 3 months[	**Female**	Mod. 1S	1.045†	0.960				267,915	282,224	-142,507	285,019	285,040
		Mod. 2S			1.044				282,224	-142,509	285,019	285,030
		Mod. 3S	1.045†	0.960	1.044				282,224	-142,507	285,019	285,051
		Mod. 4S	1.044†	*0.964*	1.041	1.020	0.910		282,224	-142,507	285,023	285,076
	% lost (5.8%)		5.4%	6.9%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals.

Table A3. Consanguinity, Smallpox and Their Interaction, Stratified[1] Cox Models Per Sex, Weaning Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[3 months, 6 months[	**Male**	Mod. 1S	1.029	1.027				253,076	272,031	-147,153	294,309	294,331
		Mod. 2S			1.243***				272,031	-147,134	294,270	294,281
		Mod. 3S	1.029	1.027	1.243***				272,031	-147,133	294,273	294,304
		Mod. 4S	1.026	1.045	1.235***	1.043	0.672		272,031	-147,133	294,276	294,329
		% lost (7.4%)	6.9%	7.4%								
[3 months, 6 months[	**Female**	Mod. 1S	1.007	1.026				250,606	269,662	-122,357	244,718	244,739
		Mod. 2S			1.220***				269,662	-122,344	244,689	244,700
		Mod. 3S	1.006	1.026	1.220***				269,662	-122,344	244,693	244,725
		Mod. 4S	1.007	1.040	1.223***	0.986	0.749		269,662	-122,344	244,697	244,750
		% lost (6.5%)	6.2%	6.1%								
[6 months, 1 year[	**Male**	Mod. 1S	1.009	1.4656*				235,199	270,552	-163,411	326,826	326,847
		Mod. 2S (NP)			*1.718***				270,552	-163,245	326,492	326,502
		Mod. 3S (NP)	1.009	1.471*	*1.718***				270,552	-163,242	326,490	326,521
		Mod. 4S (NP)	1.009	1.428*	*1.718***	0.996	1.459		270,552	-163,242	326,493	326,546
		% lost (7.2%)	6.5%	6.8%								
[6 months, 1 year[	**Female**	Mod. 1S	*1.013*	1.173				235,446	270,876	-143,990	287,984	288,005
		Mod. 2S (NP)			*1.923***				270,876	-143,761	287,525	287,535
		Mod. 3S (NP)	*1.013*	1.166	*1.923***				270,876	-143,761	287,528	287,559
		Mod. 4S (NP)	*1.018*	1.287	*1.946***	0.9369	0		270,876	-143,757	287,524	287,577
		% lost (6.1%)	5.6%	6.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_t}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals.

Table A4. Consanguinity, Smallpox and Their Interaction, Stratified[1] Cox Models Per Sex, Weaning Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[1 year, 2 years[	**Male**	Mod. 1S	0.942*	0.804				214,918	280,194	-132,019	264,043	264,064
		Mod. 2S (NP)			*2.124****				280,194	-131,709	263,421	263,431
		Mod. 3S	0.943*	0.800	*2.124****				280,194	-131,706	263,417	263,449
		Mod. 4S (NP)	0.939*	0.804	*2.113****	1.039	0.954		280,194	-131,706	263,421	263,474
		% lost (8.7%)	8.0%	11.6%								
[1 year, 2 years[	**Female**	Mod. 1S	0.931**	0.835				217,072	283,223	-128,247	256,498	256,519
		Mod. 2S (NP)			*2.261****				283,223	-127,882	255,766	255,777
		Mod. 3S (NP)	0.930**	0.827	*2.261****				283,223	-127,877	255,760	255,792
		Mod. 4S	0.934**	0.732	*2.270****	0.961	2.107		283,223	-127,876	255,762	255,815
		% lost (7.9%)	7.4%	9.1%								
[2 years, 3 years[	**Male**	Mod. 1S	0.950	1.118				197,411	260,417	-55,224	110,451	110,472
		Mod. 2S (NP)			*2.710****				260,417	-54,945	109,892	109,903
		Mod. 3S	0.948	1.114	*2.710****				260,417	-54,944	109,894	109,925
		Mod. 4S	0.962	1.286	2.765***	0.884	-		260,417	-54,942	109,893	109,946
		% lost (8.3%)	7.3%	6.3%								
[2 years, 3 years[	**Female**	Mod. 1S	0.952	0.592				198,948	263,038	-54,717	109,438	109,459
		Mod. 2S			2.847***				263,038	-54,407	108,817	108,827
		Mod. 3S	0.953	0.592	2.847***				263,038	-54,405	108,817	108,848
		Mod. 4S	0.970	0.585	2.905***	0.864	1.097		263,038	-54,404	108,819	108,871
		% lost (8.5%)	7.8%	7.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals.

Table A5. Consanguinity, Smallpox and Their Interaction, Stratified[1] Cox Models Per Sex, Weaning Period, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[3 years, 4 years[	**Male**	Mod. 1S	1.045	*1.580*				189,001	249,781	-32,490	64,983	65,004
		Mod. 2S			2.78***				249,781	-32,313	64,629	64,639
		Mod. 3S	1.045	*1.595*	2.78***				249,781	-32,312	64,630	64,661
		Mod. 4S	1.028	*1.442*	2.72***	1.139	1.920		249,781	-32,311	64,633	64,685
		% lost (4.3%)	4.1%	4.0%								
[3 years, 4 years[	**Female**	Mod. 1S	0.926†	0.517				189,589	250,912	-33,769	67,541	67,562
		Mod. 2S			2.850***				250,912	-33,571	67,145	67,155
		Mod. 3S	0.927†	0.515	2.850***				250,912	-33,569	67,144	67,175
		Mod. 4S	0.915†	0.449	2.809***	1.096	2.060		250,912	-33,568	67,147	67,199
		% lost (4.7%)	4.8%	2.9%								
[4 years, 5 years[	**Male**	Mod. 1S	0.994	2.080*				183,028	242,949	-22,504	45,011	45,032
		Mod. 2S			2.771***				242,949	-22,381	44,764	44,774
		Mod. 3S	0.993	2.082*	2.771***				242,949	-22,379	44,764	44,795
		Mod. 4S	0.952	1.339	2.610***	1.3627*	5.0945*		242,949	-22,374	44,759	44,811
		% lost (3.1%)	3.4%	4.1%								
[4 years, 5 years[	**Female**	Mod. 1S	0.942	1.965*				182,625	242,536	-22,717	45,439	45,460
		Mod. 2S			3.116***				242,536	-22,552	45,107	45,117
		Mod. 3S	0.940	1.964*	3.117***				242,536	-22,550	45,106	45,137
		Mod. 4S	0.972	2.334**	3.241***	*0.770*	-		242,536	-22,547	45,102	45,143
		% lost (3.6%)	4.0%	2.7%								

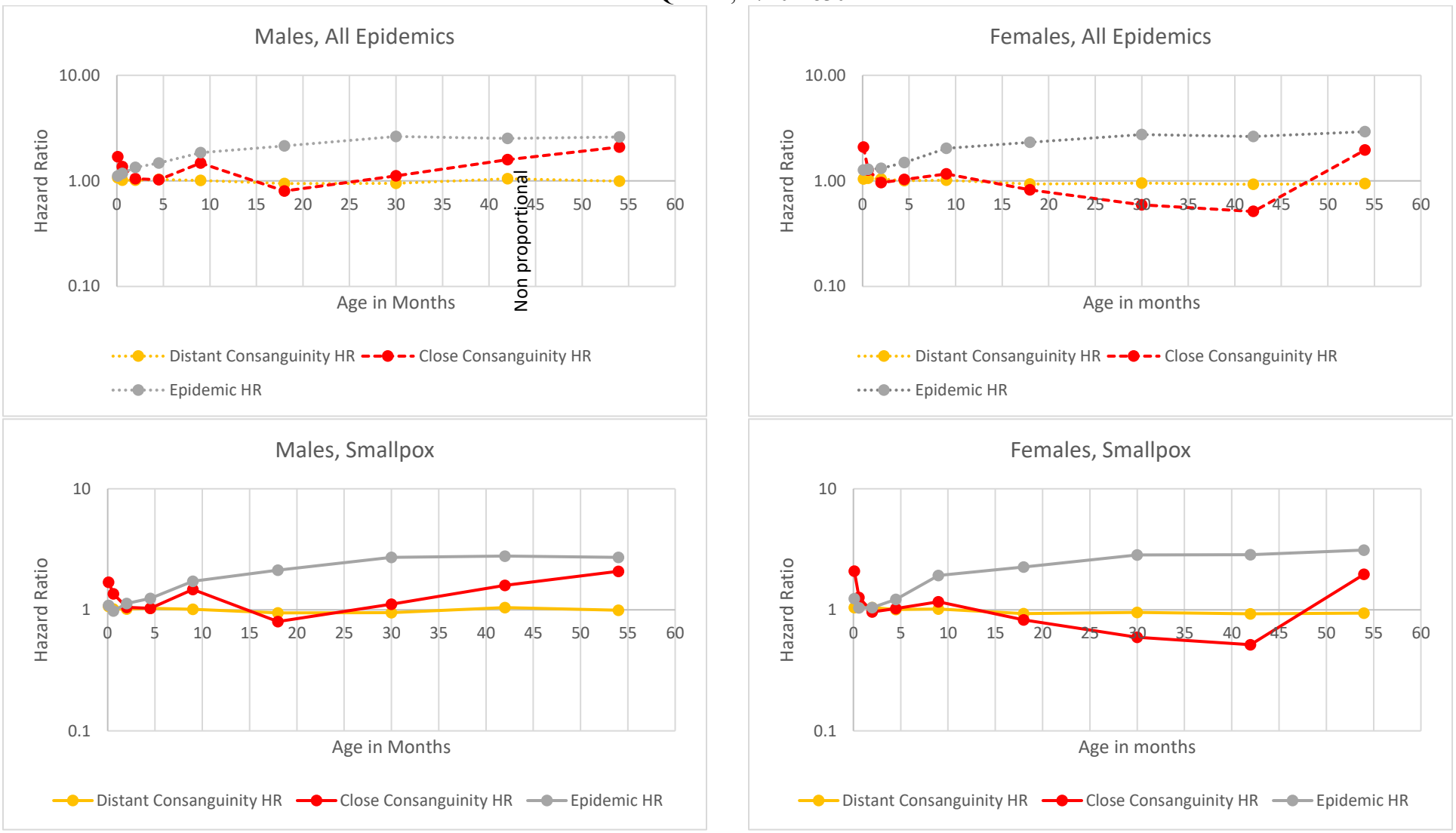
Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_t}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals.

Figure A18. Plot of Hazard Ratios, Consanguinity and Epidemic, Model 3 and 3S of Stratified[1] Cox Models Per Sex, 0 to 5 years old, Colonial Quebec, 1720-1830



Generated by Excel with RPQA-IMPQ data. [1] Stratified by period, region-urban.
No consanguinity is the reference for distant and close consanguinity; No epidemic is the reference for the upper graphs (All Epidemics); No smallpox epidemic is the reference for the lower graphs (Smallpox).

ANNEX II

Adding Control variables to the Stratified Cox Models

Next, the stratified Cox models include other variables known to be studied with child mortality: the birth interval between the child at study and their previous sibling, the fate of the previous sibling (if sibling died prior to one year old), the mother's age at birth, the rank of the child, and the number of great grandparents found in the database as a proxy to settlement. All the models for each age group are presented below (Table A9 – Table A17). Only variables of interests in the study have their hazard ratio presented in those tables, thus consanguinity, epidemic, and their interaction. As described in the notes under each table, Model 5 additionally controls for the birth interval, Model 6 controls for the fate of the previous sibling, Model 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child, and Model 8 controls for all the variables in Model 7, plus the number of great grandparents.

Section 4.2.1 showed that effects were often insignificant in the interaction variable due to low sample sizes. Thus, two types of models are presented: the models including the interaction variable of consanguinity and epidemics (A), and models without that interaction variable (B). These control variables only serve to interpret any changes in the hazard ratios of consanguinity, epidemics or their interaction compared to the section 4.2.1.

In infancy, Models 7B and 8B have the best AIC and BIC statistics for each sex. These are the models with the most control variables when the interaction variable between epidemic and consanguinity is omitted. Boys also have Model 8A among the lowest AICs prior to one year old. This is the model including all control variables as well as the interaction variable between epidemic and consanguinity. For girls, this model (8A) is only among the top two lowest AIC statistics in the six months to one year age group. However, even though Model 8A is sometimes among the better models according to the AIC statistic, the interaction hazard ratios are never significant at the 95% confidence level. The HR for living an epidemic and having distant consanguinity is only significant at the 90% confidence level in the early neonatal period (0 days to 7 days). These retained models shown below for the neonatal period suggest

that epidemics affect infants regardless of their consanguinity level, as including the interaction variable often does not allow for better models. The neonatal period shows that hazard ratios do not vary much in between the models, therefore, only Model 8A are shown for the remaining age groups of infancy.

Table A6. Complete Stratified[1] Cox Models, Neonatal Period, Colonial Quebec, 1720-1830

VARIABLES		Males [0 days, 7 days[			Females [0 days, 7 days[			Males [7 days, 28 days[			Females [7 days, 28 days[		
		Mod. 8A	Mod. 7B	Mod. 8B	Mod. 8A	Mod. 7B	Mod. 8B	Mod. 8A	Mod. 7B	Mod. 8B	Mod. 8A	Mod. 7B	Mod. 8B
Consanguinity (REF=No Consanguinity)	Distant Consang.	1.083 *	1.060 †	1.067 *	1.059	1.034	1.043	1.007	1.003	1.011	1.050 *	1.048 *	1.048 *
	Close Consang.	1.752 **	1.662 *	1.677 *	2.130 ***	2.040 ***	2.063 ***	1.339 †	1.332 †	1.344 †	1.217	1.260	1.260
Epidemic (REF=No Epidemic)	Epidemic	1.158 **	1.115 *	1.115 *	1.318 ***	1.273 ***	1.273 ***	1.170 ***	1.181 ***	1.181 ***	1.301 ***	1.297 ***	1.297 ***
Interaction Epidemic X	Distant Consang.	*0.776* †			0.783			1.066			0.974		
	Close Consang.	-			*0.572*			1.092			1.634		
Mother Birth Interval (REF= >35m)	<15m	1.512 ***	1.505 ***	1.512 ***	1.379 ***	1.374 ***	1.378 ***	1.653 ***	1.648 ***	1.653 ***	1.766 ***	1.768 ***	1.766 ***
	15-20m	0.958	0.955	0.959	0.859 **	0.856 **	0.859 **	1.123 ***	1.119 ***	1.123 ***	1.179 ***	1.179 8***	1.179 ***
	21-29m	0.872 **	0.870 **	0.872 **	0.756 ***	0.754 ***	0.756 ***	0.874 ***	0.872 ***	0.874 ***	0.895 **	0.896 **	0.895 **
	30-35m	0.932	0.931	0.932 7	0.785 ***	0.784 ***	0.785 ***	0.854 ***	0.853 ***	0.854 ***	0.858 **	0.859 **	0.858 **
	N/A or Unknown	1.256 *	1.256 *	1.257 *	1.327 **	1.327 **	1.327 **	0.749 ***	0.747 3***	0.749 ***	0.778 ***	0.779 ***	0.778 ***
Fate of previous sibling (REF=Survived 1st Year)	Died <12m	1.296 ***	1.298 ***	1.296 ***	1.330 ***	1.332 ***	1.330 ***	1.525 ***	1.525 ***	1.525 ***	1.503 ***	1.503 ***	1.503 ***
	N/A or Unknown	1.038	1.043	1.038	0.967	0.970	0.967	0.975	0.978	0.975	0.999	0.997	0.999
Mother's Age at Birth (REF= 20 to 34 y.o.)	<20y.o	1.173 ***	1.171 ***	1.172 ***	1.202 ***	1.200 ***	1.201 ***	1.007	1.007	1.007	1.017	1.017	1.017
	>35y.o.	1.156 ***	1.160 ***	1.156 ***	1.219 ***	1.222 ***	1.219 ***	1.094 ***	1.097 ***	1.094 ***	1.084 ***	1.083 ***	1.084 ***
	Unknown	0.965	0.996	0.965	1.027	1.040	1.027	0.872 ***	0.877 **	0.872 ***	0.994	0.984	0.994
Birth Rank (REF=First Born, >13 are omitted)	2-3	0.737 ***	0.741 1***	0.737 ***	0.762 ***	0.765 ***	0.762 ***	0.358 ***	0.359 ***	0.358 ***	0.358 ***	0.359 ***	0.359 ***
	4-5	0.683 ***	0.686 ***	0.683 ***	0.735 ***	0.738 ***	0.735 ***	0.390 ***	0.391 ***	0.390 ***	0.389 ***	0.389 ***	0.389 ***
	6-7	0.771 ***	0.773 ***	0.771 ***	0.795 ***	0.798 ***	0.795 ***	0.471 ***	0.472 ***	0.471 ***	0.476 ***	0.476 ***	0.476 ***
	8-9	0.806 ***	0.809 ***	0.806 ***	0.872 *	0.873 *	0.871 *	0.562 ***	0.563 ***	0.562 ***	0.592 ***	0.593 ***	0.593 ***
	10-12	0.872 *	0.872 *	0.871 *	0.868 *	0.870 *	0.868 *	0.701 ***	0.702 ***	0.701 ***	0.717 ***	0.717 ***	0.717 ***
Number great grandparents (REF=0-2)	3-5	0.849		0.850	1.084		1.084	1.229 *		1.229 *	1.143		1.129
	6-8	0.780 *		0.780 *	0.950		0.950	1.048		1.048	1.143		1.129

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. See tables in Annex for the model statistics.

HR *in italic* = Non proportional variable by Schoenfeld residuals. However, figures at the end of this Annex show that they are all minor violations (may view the HR as an average effect). Proportionality is not verified for control variables.

Table A7. Complete Stratified[1] Cox Models (8A), Post-Neonatal Infancy, Colonial Quebec, 1720-1830

		[28 days, 3 months[	[28 days, 3 months[	[3 months, 6 months[	[3 months, 6 months[	[6 months, 1 year[	[6 months, 1 year[
		Male	**Female**	**Male**	**Female**	**Male**	**Female**
VARIABLES		Mod. 8A	Mod. 8A	Mod. 8A	Mod. 8A	Mod. 8A	Mod. 8A
Consanguinity (REF=No Consanguinity)	Distant Consang.	1.021	1.032	1.022	1.004	1.014	1.018
	Close Consang.	1.027	*0.988*	1.077	0.983	1.464*	1.303
Epidemic (REF=No Epidemic)	Epidemic	1.374***	1.313***	1.469***	1.495***	*1.856***	*2.048****
Interaction Epidemic x	Distant Consang.	0.906	1.042	1.045	0.928	0.973	0.952
	Close Consang.	1.268	0.583	0.457	1.567	1.474	0.217
Mother Birth Interval (REF= >35m)	<15m	1.6398***	1.7603***	1.7683***	1.9629***	1.4767***	1.5773***
	15-20m	1.1689***	1.2319***	1.3525***	1.4609***	1.2658***	1.3734***
	21-29m	0.8837***	0.9332†	0.9779	1.0959*	1.0323	1.0819*
	30-35m	0.8250***	0.9637	0.9589	1.0455	0.9762	1.0313
	N/A or Unknown	0.6122***	0.7261***	0.7748***	0.7502***	0.7376***	0.7623***
Fate of previous sibling (REF=Survived 1st Year)	Died <12m	1.3274***	1.2792***	1.1549***	1.1600***	1.0968***	1.0899***
	N/A or Unknown	1.0195	0.8621***	0.8515***	0.8714**	0.9849	1.0278
Mother's Age at Birth (REF= 20 to 34 y.o.)	<20y.o	1.0970**	1.0405	0.9546	0.9643	0.951	1.0445
	>35y.o.	1.0177	0.9857	0.9957	1.0268	0.978	1.0209
	Unknown	0.9213*	0.9523	0.9230*	0.8829**	0.9352†	0.9332†
Birth Rank (REF=First Born, >13 are omitted)	2-3	0.4292***	0.3877***	0.4616***	0.4119***	0.5547***	0.5808***
	4-5	0.4822***	0.4580***	0.5133***	0.4591***	0.5828***	0.6133***
	6-7	0.5548***	0.5038***	0.5700***	0.5323***	0.6375***	0.6754***
	8-9	0.6022***	0.5892***	0.6240***	0.5716***	0.6924***	0.7555***
	10-12	0.7404***	0.7261***	0.7718***	0.7219***	0.7998***	0.8033***
Number great grandparents (REF=0-2)	3-5	1.0911	1.1023	0.9548	1.0786	1.0578	1.0367
	6-8	0.92	1.0362	0.8437*	0.9919	0.8957	0.9129

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. See tables in Annex for the model statistics.

HR *in italic* = Non proportional variable by Schoenfeld residuals. However, figures at the end of this Annex show that they are all minor violations (may view the HR as an average effect). Proportionality is not verified for control variables.

In the remaining early childhood age groups, Model 6B is always among the models with the best AIC and/or BIC statistics for both boys and girls. This model omits the interaction variable between epidemic and consanguinity and only includes the death of the previous sibling prior to one year as a control. This is particular as including infant mortality of the previous sibling gives the best overall Cox models for child mortality in the models past infancy. Even if the child at study survived the period that their previous sibling potentially didn't, the previous sibling's death as an infant still helps predict the child at study's mortality risk. Whereas other reproductive variables such as birth interval, rank and mother's age at birth do not seem to aid in choosing the best model for childhood mortality past infancy. Model 6A is similarly very often among the models with the best AIC and/or BIC statistics for boys and girls after infancy. It is also the model with the fate of the previous sibling as a control variable, but this model includes the interaction variable as well. Since the interaction variable is significant for close consanguinity and epidemics from one to two years for girls, for distant consanguinity and epidemics from two to three years for girls, and for close consanguinity and epidemics from four to five years for boys (at least at the 90% confidence level), it is logic that a model including the interaction is retained for its predictability of childhood mortality. Like infancy, estimates are very similar across some models. Consequently, Model 6A, with the interaction variable, and 8B, without it, are presented in the tables below for early childhood age groups past infancy. The hazard ratios for the control variables in Model 8A and 8B are similar. If there are estimates that are different from one another, then the difference is apparent in the variables of interest (consanguinity, epidemic and their interaction), which have estimates available in Tables A9 to A17.

Table A8. Complete Stratified[1] Cox Models, 1 to 5 years old, Colonial Quebec, 1720-1830

		[1 year, 2 years[		[1 year, 2 years[		[2 years, 3 years[		[2 years, 3 years[		[3 years, 4 years[		[3 years, 4 years[		[4 years, 5 years[		[4 years, 5 years[	
		Male		**Female**		**Male**		**Female**		**Male**		**Female**		**Male**		**Female**	
VARIABLES		Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B	Mod. 6A	Mod. 8B
Consanguinity (REF=No Consanguinity)	Distant Consang.	0.94 *	0.96 †	0.94 *	0.94 *	0.95	0.96	0.98	0.96	1.02	1.05	0.93	0.93	0.95	1.00	0.97	0.94
	Close Consang.	0.81	0.81	0.63 †	0.84	1.12	1.13	0.60	0.60	*1.45*	*1.59*	0.46	0.51	1.35	2.15 *	2.37 **	1.97 *
Epidemic (REF=No Epidemic)	Epidemic	*2.13* **	2.15 ***	*2.33* **	2.32 ***	2.63 ***	*2.64* ***	2.81 ***	2.74 ***	2.46 ***	2.53 ***	2.63 ***	2.63 ***	2.46 ***	2.60 ***	3.02 ***	2.92 ***
Interaction Epidemic X	Distant Consang.	1.04		0.94		0.87		0.82 †		1.17		1.00		1.36 *		0.81	
	Close Consang.	0.85		3.15 **		-		1.00		1.82		1.84		4.98 *		-	
Mother Birth Interval (REF= >35m)	<15m		1.15 ***		1.08		1.1		1.12		0.91		1.07		0.98		0.91
	15-20m		1.15 ***		1.11 **		1.10		1.16 **		0.98		0.99		1.01		0.98
	21-29m		1.12 ***		1.07		1.06		1.10		0.97		1.01		0.95		0.93
	30-35m		1.11 *		0.97		1.02		1.07		0.95		1.05		0.81		1.03
	N/A or Unknown		0.94		0.81 **		1.01		0.87		0.83		1.00		1.05		0.88
Fate of previous sibling (REF=Survived 1st Year)	Died <12m	1.12 ***	1.10 ***	1.06 ***	1.05 *	1.13 ***	1.12 ***	1.04	1.03	1.07	1.08 *	1.02	1.02	1.08	1.07	1.13 **	1.14 **
	N/A or Unknown	0.90 ***	0.99	0.88 ***	1.01	0.94	0.98	0.94	1.06	1.05	1.19 *	0.92	1.02	1.00	0.93	1.00	1.18
Mother's Age at Birth (REF= 20 to 34 y.o.)	<20y.o		0.98		1.02		1.02		1.00		0.94		0.95		0.98		1.13
	>35y.o.		1.07 **		1.00		1.01		1.02		1.01		0.98		0.91		1.09
	Unknown		0.99		1.09 *		1.11		1.11		1.00		0.94		0.85		0.89
Birth Rank (REF=First Born >13 are omitted)	2-3		0.88 *		0.87 **		0.97		0.85 *		0.98		1.09		0.92		1.12
	4-5		0.93		0.92		1.03		0.97		1.00		1.16		1.03		1.25
	6-7		0.96		0.94		1.04		0.94		0.98		1.11		1.03		1.15
	8-9		0.93		0.94		1.01		0.94		0.98		1.10		1.13		1.07
	10-12		0.99		0.97		0.99		0.99		0.98		1.11		1.07		1.02
Number great grandparents (REF=0-2)	3-5		0.99		0.87		1.11		0.91		0.90		0.99		1.76 *		1.43
	6-8		0.81 **		0.77 ***		0.94		0.85		0.83		0.94		1.40		1.31

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. See tables in Annex for the model statistics.

HR *in italic* = Non proportional variable by Schoenfeld residuals. However, figures at the end of this Annex show that they are all minor violations (may view the HR as an average effect). Proportionality is not verified for control variables.

Table A9. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 0 to 7 days, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[0 days, 7 days[	**Male**	Mod. 5A	1.071*	1.754**	1.156**	0.785†	-	310,961	313,184	-70,586	141,190	141,286
		Mod. 6A	1.081*	1.739**	1.155**	*0.777†*	-	310,961	313,184	-70,684	141,381	141,456
		Mod. 7A	1.075*	1.737**	1.158**	*0.777†*	-	310,961	313,184	-70,451	140,941	141,155
		Mod. 8A	1.083*	1.752**	1.158**	*0.776†*	-	310,961	313,184	-70,447	140,937	141,172
		Mod. 5B	1.057†	1.676*	1.115*			310,961	313,184	-70,589	141,194	141,279
		Mod. 6B	1.067*	1.662*	1.113*			310,961	313,184	-70,686	141,383	141,436
		Mod. 7B	1.060†	1.662*	1.115*			310,961	313,184	-70,453	140,943	141,135
		Mod. 8B	1.067*	1.677*	1.115*			310,961	313,184	-70,449	140,939	141,152
		% lost	-	-								
[0 days, 7 days[	**Female**	Mod. 5A	1.044	2.079***	1.313***	0.790	*0.565*	299,451	301,688	-54,700	109,419	109,526
		Mod. 6A	1.056	2.117***	1.315***	0.788	*0.597*	299,451	301,688	-54,791	109,596	109,670
		Mod. 7A	1.05	2.106***	1.318***	0.783	*0.572*	299,451	301,688	-54,578	109,196	109,408
		Mod. 8A	1.059	2.130***	1.318***	0.783	*0.572*	299,451	301,688	-54,574	109,192	109,426
		Mod. 5B	1.029	2.012***	1.269***			299,451	301,688	-54,701	109,418	109,503
		Mod. 6B	1.041	2.057***	1.271***			299,451	301,688	-54,792	109,594	109,648
		Mod. 7B	1.034	2.040***	1.273***			299,451	301,688	-54,579	109,195	109,386
		Mod. 8B	1.043	2.063***	1.273***			299,451	301,688	-54,576	109,191	109,404
		% lost	-	-								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See figure at the end of this Annex.

Table A10. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 7 to 28 days, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[7 days, 28 days[	**Male**	Mod. 5A	0.996	1.361*	1.166***	1.072	0.980	292,655	299,820	-164,683	329,387	329,493
		Mod. 6A	1.007	1.330†	1.168***	1.063	0.970	292,655	299,820	-164,656	329,325	329,400
		Mod. 7A	0.999	1.327†	1.170***	1.067	1.081	292,655	299,820	-163,261	326,562	326,774
		Mod. 8A	1.007	1.339†	1.170***	1.066	1.092	292,655	299,820	-163,246	326,537	326,770
		Mod. 5B	1.000	1.360*	1.179***			292,655	299,820	-164,684	329,383	329,468
		Mod. 6B	1.011	1.328†	1.179***			292,655	299,820	-164,656	329,322	329,375
		Mod. 7B	1.003	1.332†	1.181***			292,655	299,820	-163,261	326,558	326,749
		Mod. 8B	1.011	1.344†	1.181***			292,655	299,820	-163,247	326,533	326,746
	% lost (6.1%)		4.7%	7.7%								
[7 days, 28 days[	**Female**	Mod. 5A	1.041†	1.193	1.289***	0.982	1.552	284,162	291,163	-135,377	270,775	270,880
		Mod. 6A	1.059*	1.208	1.298***	0.971	1.661	284,162	291,163	-135,426	270,865	270,939
		Mod. 7A	1.050*	1.217	1.301***	0.974	1.635	284,162	291,163	-134,221	268,482	268,694
		Mod. 8A	1.050*	1.217	1.301***	0.974	1.634	284,162	291,163	-134,220	268,484	268,717
		Mod. 5B	1.040†	1.232	1.287***			284,162	291,163	-135,378	270,771	270,856
		Mod. 6B	1.056*	1.251	1.294***			284,162	291,163	-135,426	270,862	270,915
		Mod. 7B	1.048*	1.260	1.297***			284,162	291,163	-134,222	268,479	268,669
		Mod. 8B	1.048*	1.260	1.297***			284,162	291,163	-134,221	268,481	268,693
	% lost (5.3%)		4.3%	7.8%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

Table A11. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 28 days to 3 months, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[28 days, 3 months[	**Male**	Mod. 5A	1.011	1.037	1.363***	0.907	1.232	273,052	293,157	-165,287	330,593	330,699
		Mod. 6A	1.021	1.019	1.368***	0.906	1.165	273,052	293,157	-165,495	331,004	331,078
		Mod. 7A	1.012	1.016	1.374***	0.907	1.261	273,052	293,157	-164,576	329,192	329,404
		Mod. 8A	1.021	1.027	1.374***	0.906	1.268	273,052	293,157	-164,560	329,165	329,398
		Mod. 5B	1.004	1.056	1.345***			273,052	293,157	-165,288	330,591	330,676
		Mod. 6B	1.014	1.033	1.349***			273,052	293,157	-165,496	331,002	331,055
		Mod. 7B	1.005	1.037	1.355***			273,052	293,157	-164,577	329,190	329,381
		Mod. 8B	1.014	1.048	1.356***			273,052	293,157	-164,561	329,162	329,374
	% lost (6.8%)		6.2%	8.7%								
[28 days, 3 months[	**Female**	Mod. 5A	1.025	*0.971*	1.305***	1.048	0.571	267,915	287,831	-141,972	283,964	284,070
		Mod. 6A	1.041†	*0.976*	1.313***	1.032	0.624	267,915	287,831	-142,172	284,357	284,431
		Mod. 7A	1.029	*0.984*	1.312***	1.042	0.583	267,915	287,831	-141,286	282,612	282,823
		Mod. 8A	1.032	*0.988*	1.313***	1.042	0.583	267,915	287,831	-141,284	282,612	282,845
		Mod. 5B	1.028	*0.943*	1.313***			267,915	287,831	-141,972	283,961	284,045
		Mod. 6B	1.043†	0.954	1.319***			267,915	287,831	-142,172	284,354	284,407
		Mod. 7B	1.032	0.957	1.320***			267,915	287,831	-141,286	282,608	282,799
		Mod. 8B	1.035	0.962	1.320***			267,915	287,831	-141,284	282,609	282,820
	% lost (5.8%)		5.4%	6.9%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See figures at the end of this Annex.

Table A12. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 3 to 6 months, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[3 months, 6 months[	**Male**	Mod. 5A	1.014	1.088	1.466***	1.036	0.427	253,076	278,859	-146,630	293,281	293,386
		Mod. 6A	1.024	1.065	1.474***	1.040	0.439	253,076	278,859	-146,898	293,810	293,884
		Mod. 7A	1.015	1.067	1.469***	1.046	0.455	253,076	278,859	-146,202	292,444	292,655
		Mod. 8A	1.022	1.077	1.469***	1.045	0.457	253,076	278,859	-146,193	292,431	292,663
		Mod. 5B	1.017	1.038	1.473***			253,076	278,859	-146,631	293,278	293,362
		Mod. 6B	1.027	1.019	1.481***			253,076	278,859	-146,899	293,808	293,860
		Mod. 7B	1.019	1.024	1.478***			253,076	278,859	-146,203	292,441	292,631
		Mod. 8B	1.026	1.033	1.478***			253,076	278,859	-146,194	292,428	292,639
	% lost (7.4%)		6.9%	7.4%								
[3 months, 6 months[	**Female**	Mod. 5A	0.997	0.966	1.497***	0.930	1.708	250,606	276,761	-121,903	243,827	243,932
		Mod. 6A	1.011	0.972	1.505***	0.926	1.744	250,606	276,761	-122,125	244,264	244,337
		Mod. 7A	1.000	0.976	1.494***	0.928	1.569	250,606	276,761	-121,406	242,853	243,063
		Mod. 8A	1.004	0.983	1.495***	0.928	1.567	250,606	276,761	-121,403	242,851	243,083
		Mod. 5B	0.992	1.016	1.483***			250,606	276,761	-121,904	243,824	243,908
		Mod. 6B	1.005	1.025	1.490***			250,606	276,761	-122,126	244,261	244,314
		Mod. 7B	0.994	1.021	1.480***			250,606	276,761	-121,407	242,850	243,039
		Mod. 8B	0.998	1.027	1.480***			250,606	276,761	-121,404	242,848	243,059
	% lost (6.5%)		6.2%	6.1%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, ŧp<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

Table A13. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 6 to 12 months, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[6 months, 1 year[	**Male**	Mod. 5A	1.005	1.444*	*1.856***	0.970	1.438	235,199	283,486	-162,947	325,914	326,020
		Mod. 6A	1.011	1.419*	*1.859***	0.969	1.400	235,199	283,486	-163,068	326,149	326,223
		Mod. 7A	1.005	1.448*	*1.857***	0.973	1.468	235,199	283,486	-162,718	325,476	325,687
		Mod. 8A	1.014	1.464*	*1.856***	0.973	1.474	235,199	283,486	-162,703	325,450	325,682
		Mod. 5B	1.002	1.497**	*1.850***			235,199	283,486	-162,948	325,911	325,995
		Mod. 6B	1.008	1.468*	*1.852***			235,199	283,486	-163,068	326,146	326,199
		Mod. 7B	1.002	1.504**	*1.851***			235,199	283,486	-162,719	325,473	325,663
		Mod. 8B	1.012	1.521**	*1.850***			235,199	283,486	-162,704	325,447	325,658
	% lost (7.2%)		6.5%	6.8%								
[6 months, 1 year[	**Female**	Mod. 5A	1.008	1.279	*2.046***	0.955	0.216	235,446	284,664	-143,456	286,931	287,037
		Mod. 6A	1.017	*1.277*	*2.049***	0.953	0.226	235,446	284,664	-143,594	287,203	287,276
		Mod. 7A	1.010	*1.288*	*2.048***	0.952	0.218	235,446	284,664	-143,256	286,551	286,763
		Mod. 8A	1.018	1.303	*2.048***	0.952	0.217	235,446	284,664	-143,248	286,539	286,772
		Mod. 5B	1.004	*1.156*	*2.028***			235,446	284,664	-143,458	286,931	287,016
		Mod. 6B	1.012	*1.160*	*2.031***			235,446	284,664	-143,596	287,203	287,255
		Mod. 7B	1.005	*1.165*	*2.029***			235,446	284,664	-143,258	286,552	286,742
		Mod. 8B	1.012	*1.178*	*2.029***			235,446	284,664	-143,250	286,539	286,751
	% lost (6.1%)		5.6%	6.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, ŧp<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See figures at the end of this Annex

Table A14. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 1 to 2 years, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[1 year, 2 years[	**Male**	Mod. 5A	0.9391*	0.818	*2.135***	1.038	0.857	214,918	304,209	-131,601	263,222	263,328
		Mod. 6A	0.9392*	0.815	*2.137***	1.037	0.848	214,918	304,209	-131,593	263,199	263,274
		Mod. 7A	0.9410*	0.815	*2.137***	1.037	0.840	214,918	304,209	-131,559	263,157	263,370
		Mod. 8A	0.9528†	0.828	*2.137***	1.036	0.844	214,918	304,209	-131,538	263,120	263,353
		Mod. 5B	0.9429*	0.804	2.146***			214,918	304,209	-131,601	263,218	263,303
		Mod. 6B	0.9431*	0.801	2.148***			214,918	304,209	-131,593	263,196	263,249
		Mod. 7B	0.9448*	0.800	2.147***			214,918	304,209	-131,559	263,154	263,345
		Mod. 8B	0.9566†	0.813	2.147***			214,918	304,209	-131,538	263,116	263,329
	% lost (8.7%)		8.0%	11.6%								
[1 year, 2 years[	**Female**	Mod. 5A	0.9368**	0.6311†	2.335***	0.942	3.158**	217,072	308,468	-127,733	255,485	255,592
		Mod. 6A	0.9376*	0.6296†	*2.337***	0.941	3.152**	217,072	308,468	-127,744	255,501	255,575
		Mod. 7A	0.9392*	0.6327†	*2.336***	0.940	3.167**	217,072	308,468	-127,714	255,468	255,680
		Mod. 8A	0.9484*	0.641	2.335***	0.940	3.167**	217,072	308,468	-127,702	255,448	255,682
		Mod. 5B	0.9303**	0.828	*2.322***			217,072	308,468	-127,736	255,488	255,573
		Mod. 6B	0.9310**	0.826	*2.324***			217,072	308,468	-127,747	255,504	255,557
		Mod. 7B	0.9325**	0.830	2.323***			214,918	304,209	-127,717	255,470	255,662
		Mod. 8B	0.9416*	0.842	2.322***			214,918	304,209	-127,705	255,450	255,663
	% lost (7.9%)		7.4%	9.1%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

HR *in italic* = Non proportional variable by Schoenfeld residuals. See figures at the end of this Annex

Table A15. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 2 to 3 years, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[2 years, 3 years[	**Male**	Mod. 5A	0.966	1.307	2.689***	0.871	-	197,411	282,973	-54,912	109,842	109,937
		Mod. 6A	0.948	1.118	2.633***	0.871	-	197,411	282,973	-54,914	109,845	109,929
		Mod. 7A	0.968	1.301	2.692***	0.871	-	197,411	282,973	-54,899	109,839	110,050
		Mod. 8A	0.977	1.317	2.693***	0.870	-	197,411	282,973	-54,894	109,832	110,065
		Mod. 5B	0.948	1.118	2.633***			197,411	282,973	-54,914	109,845	109,929
		Mod. 6B	0.967	1.300	*2.691***			197,411	282,973	-54,906	109,825	109,899
		Mod. 7B	0.950	1.112	*2.636***			197,411	282,973	-54,902	109,840	110,030
		Mod. 8B	0.959	1.125	*2.637***			197,411	282,973	-54,897	109,834	110,045
	% lost (8.3%)		7.3%	6.3%								
[2 years, 3 years[	**Female**	Mod. 5A	0.978	0.597	2.812***	0.821†	0.998	198,948	287,404	-54,378	108,777	108,883
		Mod. 6A	0.979	0.595	2.813***	0.821†	1.000	198,948	287,404	-54,383	108,781	108,855
		Mod. 7A	0.981	0.598	2.814***	0.820†	0.997	198,948	287,404	-54,364	108,767	108,979
		Mod. 8A	0.986	0.603	2.814***	0.820†	0.995	198,948	287,404	-54,362	108,768	109,000
		Mod. 5B	0.953	0.597	2.739***			198,948	287,404	-54,380	108,777	108,861
		Mod. 6B	0.954	0.595	2.741***			198,948	287,404	-54,385	108,780	108,833
		Mod. 7B	0.956	0.597	2.741***			198,948	287,404	-54,366	108,767	108,957
		Mod. 8B	0.961	0.602	2.740***			198,948	287,404	-54,364	108,767	108,979
	% lost (8.5%)		7.8%	7.0%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

Table A16. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 3 to 4 years, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[3 years, 4 years[	**Male**	Mod. 5A	1.024	*1.449*	2.460***	1.167	1.800	189,001	271,753	-32,327	64,673	64,778
		Mod. 6A	1.023	*1.447*	2.461***	1.167	1.820	189,001	271,753	-32,326	64,666	64,739
		Mod. 7A	1.025	*1.440*	2.461***	1.167	1.807	189,001	271,753	-32,322	64,685	64,895
		Mod. 8A	1.030	*1.450*	2.461***	1.167	1.803	189,001	271,753	-32,321	64,687	64,918
		Mod. 5B	1.046	*1.592*	2.526***			189,001	271,753	-32,328	64,671	64,755
		Mod. 6B	1.045	*1.591*	2.528***			189,001	271,753	-32,327	64,664	64,716
		Mod. 7B	1.047	*1.582*	2.528***			189,001	271,753	-32,323	64,683	64,872
		Mod. 8B	1.053	*1.593*	2.528***			189,001	271,753	-32,322	64,685	64,895
	% lost (4.3%)		4.1%	4.0%								
[3 years, 4 years[	**Female**	Mod. 5A	0.926	0.455	2.633***	1.000	1.837	189,589	274,229	-33,574	67,168	67,273
		Mod. 6A	0.927	0.456	2.632***	1.001	1.837	189,589	274,229	-33,577	67,167	67,241
		Mod. 7A	0.926	0.454	2.632***	1.001	1.842	189,589	274,229	-33,571	67,183	67,393
		Mod. 8A	0.929	0.456	2.632***	1.001	1.842	189,589	274,229	-33,571	67,186	67,417
		Mod. 5B	0.926†	0.513	2.635***			189,589	274,229	-33,574	67,164	67,249
		Mod. 6B	0.927†	0.514	2.634***			189,589	274,229	-33,577	67,164	67,216
		Mod. 7B	0.926†	0.512	2.635***			189,589	274,229	-33,571	67,179	67,368
		Mod. 8B	0.929	0.515	2.634***			189,589	274,229	-33,571	67,182	67,393
	% lost (4.7%)		4.8%	2.9%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, †p<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

Table A17. Consanguinity, Epidemics and Their Interaction, Stratified[1] and Controlled[2] Cox Models Per Sex, 4 to 5 years, Colonial Quebec, 1720-1830

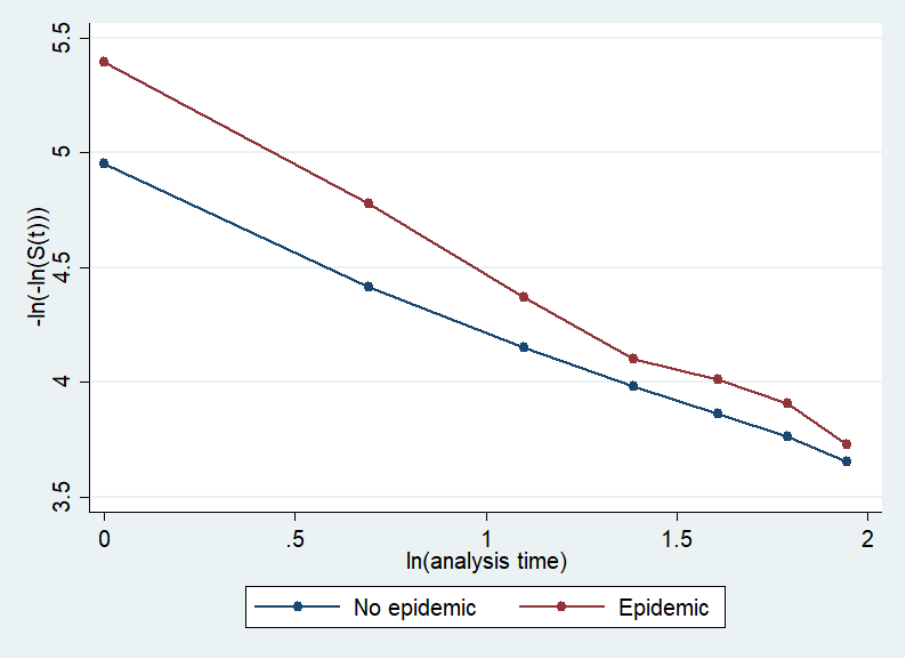
Models stratified by Period Region-urban			**CONSANGUINITY** (REF=No Consanguinity)		**EPIDEMIC** (REF=No Epidemic)	**INTERACTION**						
Age interval	Sex	Model	Distant Consang.	Close Consang.	Epidemic	Epidemic x Distant Consang.	Epidemic x Close Consang.	N subjects	N obs	Log Likelihood	AIC	BIC
[4 years, 5 years[	**Male**	Mod. 5A	0.947	1.352	2.458***	1.3589*	5.080*	183,028	265,454	-22,371	44,762	44,866
		Mod. 6A	0.948	1.345	2.460***	1.3581*	4.982*	183,028	265,454	-22,373	44,759	44,833
		Mod. 7A	0.944	1.363	2.458***	1.3570*	5.099*	183,028	265,454	-22,363	44,766	44,976
		Mod. 8A	0.953	1.382	2.459***	1.3570*	5.082*	183,028	265,454	-22,358	44,760	44,991
		Mod. 5B	0.992	2.103*	2.602***			183,028	265,454	-22,375	44,767	44,851
		Mod. 6B	0.993	2.087*	2.603***			183,028	265,454	-22,377	44,765	44,817
		Mod. 7B	0.989	2.122*	2.601***			183,028	265,454	-22,368	44,771	44,960
		Mod. 8B	0.998	2.150*	2.603***			183,028	265,454	-22,363	44,765	44,975
	% lost (3.1%)		3.4%	4.1%								
[4 years, 5 years[	**Female**	Mod. 5A	0.972	2.374**	3.018***	0.806	-	182,625	265,932	-22,543	45,107	45,212
		Mod. 6A	0.971	2.366**	3.018***	0.807	-	182,625	265,932	-22,542	45,099	45,172
		Mod. 7A	0.973	2.371**	3.018***	0.807	-	182,625	265,932	-22,532	45,103	45,302
		Mod. 8A	0.975	2.381**	3.017***	0.807	-	182,625	265,932	-22,531	45,106	45,337
		Mod. 5B	0.942	1.966*	2.921***			182,625	265,932	-22,546	45,109	45,192
		Mod. 6B	0.941	1.959*	2.922***			182,625	265,932	-22,545	45,100	45,153
		Mod. 7B	0.942	1.964*	2.921***			182,625	265,932	-22,535	45,106	45,295
		Mod. 8B	0.945	1.972*	2.921***			182,625	265,932	-22,534	45,107	45,317
	% lost (3.6%)		4.0%	2.7%								

Source: RPQA-IMPQ. ***p<0.001, **p<0.01, *p<0.05, ŧp<0.1

[1] Stratified by period, region-urban. [2]Mod. 5 controls for the birth interval, Mod. 6 controls for the fate of the previous sibling, Mod. 7 controls for the birth interval, fate of the previous sibling, mother's age at birth, and rank of child. Mod. 8 controls for all the variables in Mod.7, plus the number of great grandparents. N obs. = the number of observations including the multiple episodes of epidemics.

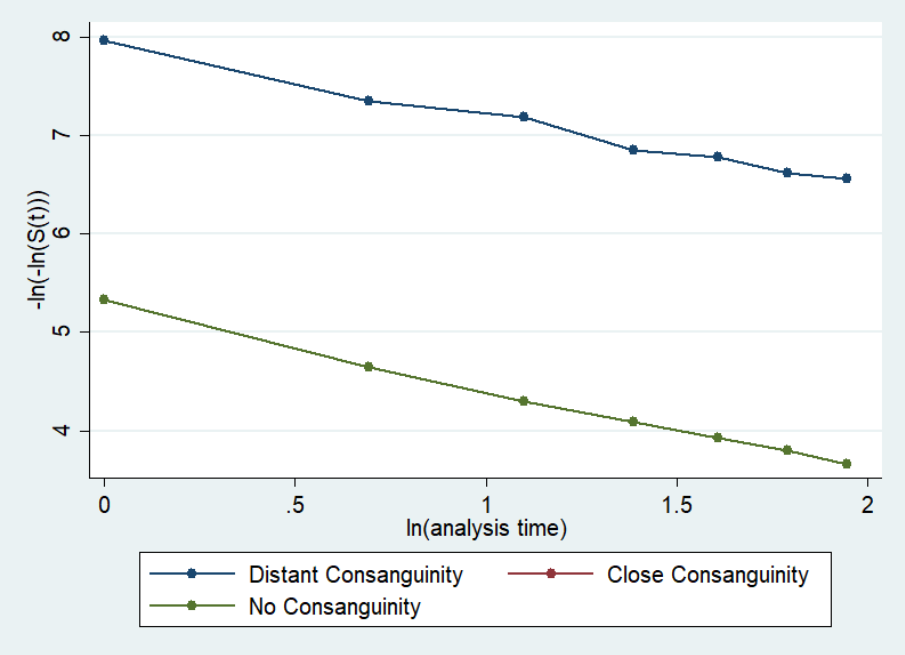
% lost (REF%) = The proportion of individuals lost in the last age interval per consanguinity level: $\frac{[\text{N subjects}]_{t-1} - [\text{N subjects}]_{t}}{[\text{N subjects}]_{t-1}}$.

Figure A19. Graphical Test of Proportional Hazards: Epidemic, Distantly Consanguineous Males, 0-7 Days Old, Adjusted for Every Control Variable



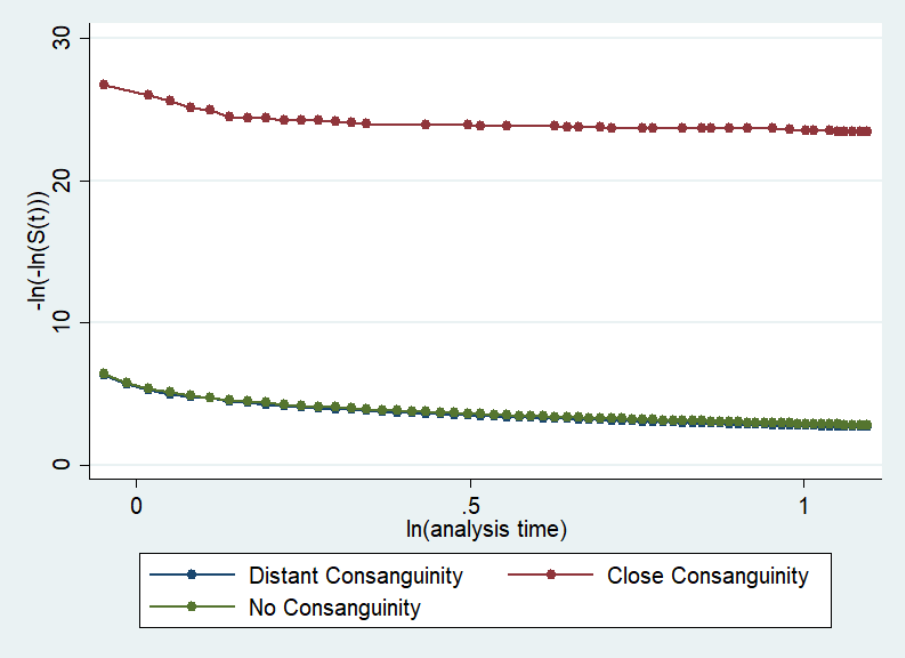
Generated by STATA using RPQA-IMPQ data.

Figure A20. Graphical Test of Proportional Hazards: Epidemic, Closely Consanguineous Females, 0-7 Days Old, Adjusted for Every Control Variable



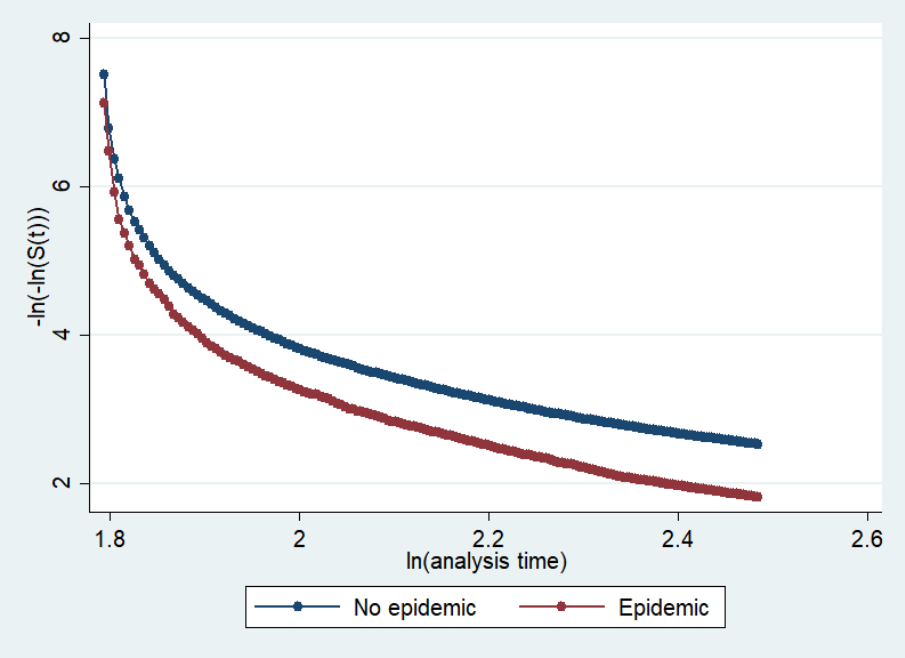
Generated by STATA using RPQA-IMPQ data.

Figure A21. Graphical Test of Proportional Hazards: Consanguinity, Females, 28 Days to 3 Months Old, Adjusted for Every Control Variable and Epidemic



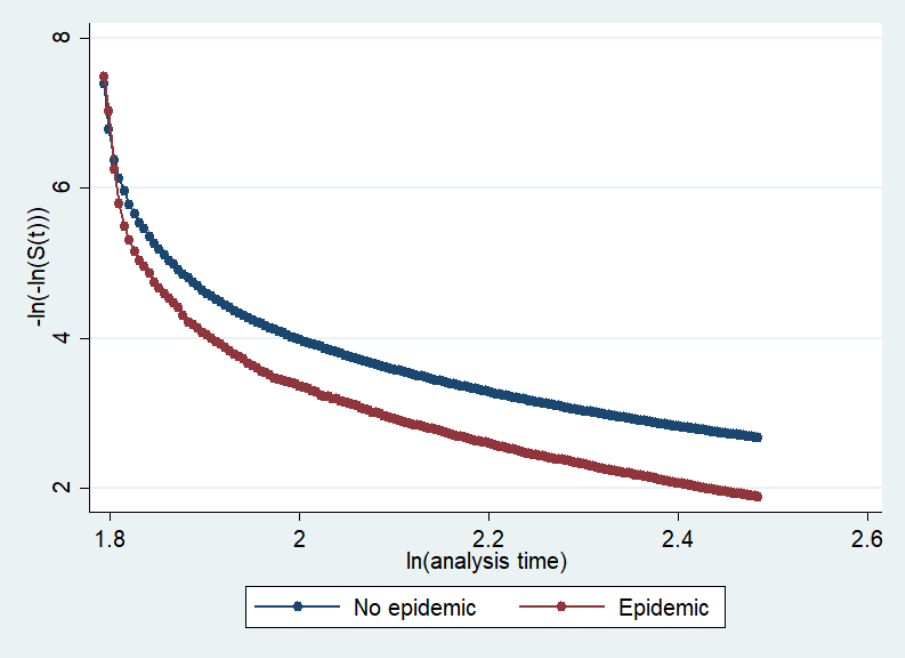
Generated by STATA using RPQA-IMPQ data.

Figure A22. Graphical Test of Proportional Hazards: Epidemic, Males, 6 to 12 Months Old, Adjusted for Every Control Variable and Consanguinity



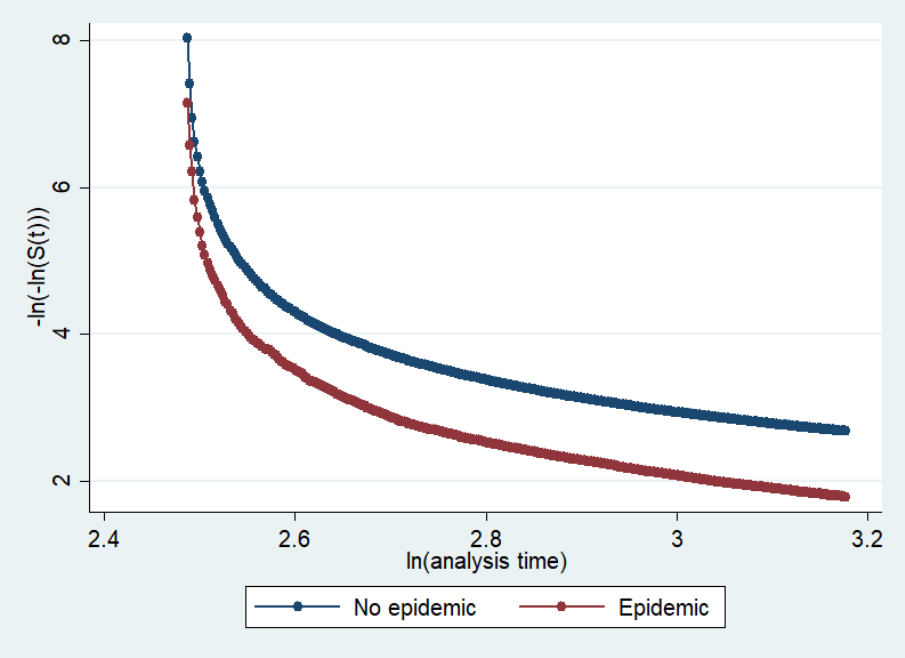
Generated by STATA using RPQA-IMPQ data.

Figure A23. Graphical Test of Proportional Hazards: Epidemic, Females, 6 to 12 Months Old, Adjusted for Every Control Variable and Consanguinity



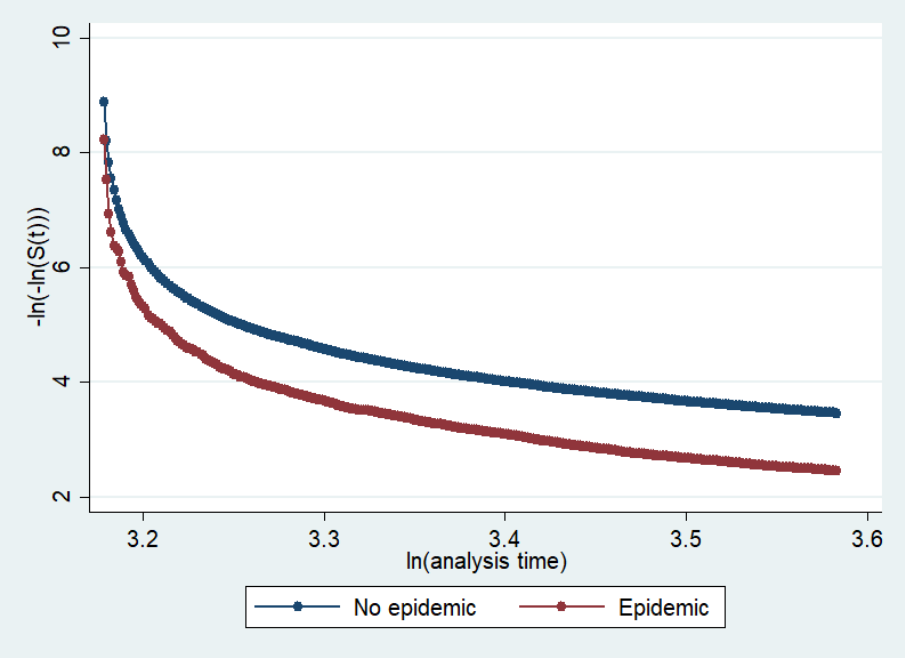
Generated by STATA using RPQA-IMPQ data.

Figure A24. Graphical Test of Proportional Hazards: Epidemic, Females, 1 to 2 Years Old, Adjusted for Consanguinity and Fate of Previous Sibling



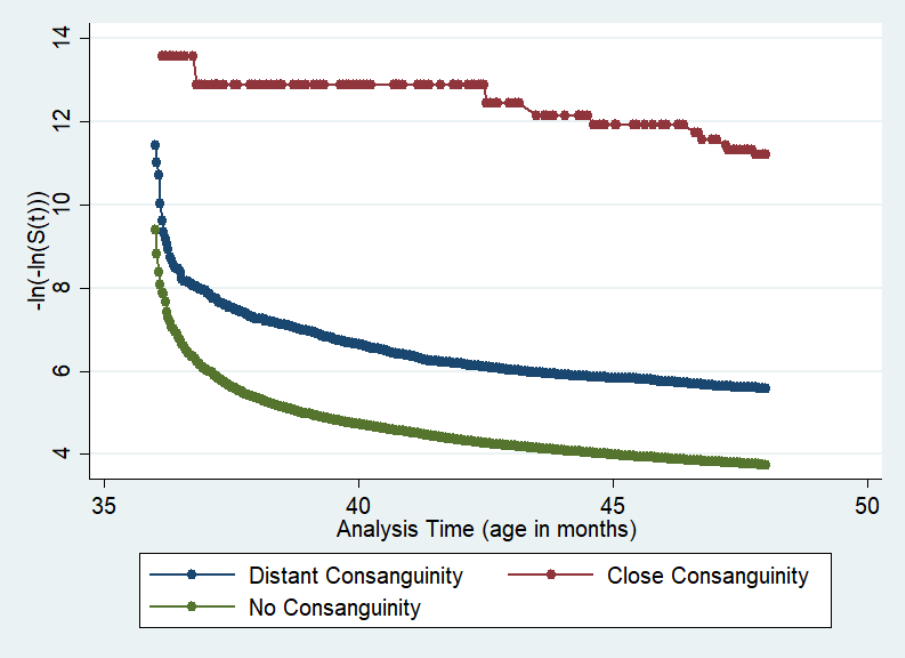
Generated by STATA using RPQA-IMPQ data.

Figure A25. Graphical Test of Proportional Hazards: Epidemic, Males, 2 to 3 Years Old, Adjusted for Every Control Variable and Consanguinity



Generated by STATA using RPQA-IMPQ data.

Figure A26. Graphical Test of Proportional Hazards: Consanguinity, Male, 3 to 4 Years Old, Adjusted for Every Control Variable and Epidemic



Generated by STATA using RPQA-IMPQ data.

ANNEX III

Additional models: Cox Models Stratified by Sex

Table A18. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models, 0 Day to 28 Days, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban Sex	[0 days, 7 days[				[7 days, 28 days[			
	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Consanguinity								
Long-Term Consanguinity	1.058*		1.058*	1.073**	1.034*		1.034*	1.032
Close Consanguinity	1.868***		1.870***	1.938***	1.314*		1.317*	1.300*
No Consanguinity	1.000		1.000	1.000	1.000		1.000	1.000
Epidemic								
No Epidemic		1.000	1.000	1.000		1.000	1.000	1.000
Epidemic		1.176***	1.176***	1.219***		1.225***	1.225***	1.219***
Interaction								
Epidemic x Long-Term Consanguinity				0.782*				1.029
Epidemic x Close Consanguinity				0.381				1.243
N subjects	610,412				576,817			
N obs	614,872				590,983			
Log Likelihood (model)	-125,848	-125,847	-125,838	-125,834	-301,405	-301,368	-301,364	-301,363
AIC	251,700	251,697	251,681	251,678	602,814	602,739	602,733	602,737
BIC	251,723	251,708	251,715	251,734	602,837	602,750	602,767	602,793

Source: RPQA-IMPQ.

[1] Stratified by period, region-urban, and sex. N obs. = the number of observations including the multiple episodes of epidemics.

NPC=Non proportional close consanguinity by Schoenfeld residuals. NP=Non proportional model by Schoenfeld residuals.

Table A19. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models, 28 Days to 6 Months, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban Sex	[28 days, 3 months[				[3 months, 6 months[			
	NPC Mod. 1	Mod. 2	NPC Mod. 3	NPC Mod. 4	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Consanguinity								
Long-Term Consanguinity	1.029		1.029	1.031	1.019		1.018	1.019
Close Consanguinity	1.007		1.006	1.01	1.027		1.027	1.026
No Consanguinity	1.000		1.000	1.000	1.000		1.000	1.000
Epidemic								
No Epidemic		1.000	1.000	1.000		1.000	1.000	1.000
Epidemic		1.330***	1.330***	1.336***		1.482***	1.482***	1.484***
Interaction								
Epidemic x Long-Term Consanguinity				0.971				0.989
Epidemic x Close Consanguinity				0.951				1.012
N subjects	540,967				503,682			
N obs	580,988				555,620			
Log Likelihood (model)	-308,443	-308,356	-308,355	-308,355	-269,510	-269,355	-269,354	-269,354
AIC					539,024	538,712	538,714	538,718
BIC					539,046	538,723	538,748	538,774

Source: RPQA-IMPQ.

[1] Stratified by period, region-urban, and sex. N obs. = the number of observations including the multiple episodes of epidemics.

NPC=Non proportional close consanguinity by Schoenfeld residuals. NP=Non proportional model by Schoenfeld residuals.

Table A20. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models, 6 Months to 2 Years, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban Sex	[6 months, 1 year[				[1 year, 2 years[			
	Mod. 1	NP Mod. 2	NP Mod. 3	NP Mod. 4	Mod. 1	NP Mod. 2	NP Mod. 3	NP Mod. 4
Consanguinity								
Long-Term Consanguinity	1.011		1.01	1.014	0.937***		0.937***	0.938***
Close Consanguinity	1.320*		1.318*	1.352*	0.82		0.813	0.717
No Consanguinity	1.000		1.000	1.000	1.000		1.000	1.000
Epidemic								
No Epidemic		1.000	1.000	1.000		1.000	1.000	1.000
Epidemic		1.935***	1.935***	1.947***		2.233***	2.233***	2.233***
Interaction								
Epidemic x Long-Term Consanguinity				0.963				0.989
Epidemic x Close Consanguinity				0.755				2.039*
N subjects	470,645				431,990			
N obs	568,150				612,677			
Log Likelihood (model)	-307,402	-306,804	-306,801	-306,801	-260,266	-259,424	-259,415	-259,413
AIC	614,808	613,610	613,609	613,612	520,537	518,849	518,836	518,837
BIC	614,830	613,622	613,642	613,668	520,559	518,860	518,870	518,893

Source: RPQA-IMPQ.

[1] Stratified by period, region-urban, and sex. N obs. = the number of observations including the multiple episodes of epidemics.

NPC=Non proportional close consanguinity by Schoenfeld residuals. NP=Non proportional model by Schoenfeld residuals.

Table A21. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models, 2 Years to 4 Years, Colonial Quebec, 1720-1830

Models stratified by Period	[2 years, 3 years[				[3 years, 4 years[			
Region-urban					NPC		NPC	NPC
Sex	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Consanguinity								
Long-Term Consanguinity	0.951*		0.951*	0.972	0.984		0.985	0.974
Close Consanguinity	0.842		0.842	0.932	0.996		0.996	0.908
No Consanguinity	1.000		1.000	1.000	1.000		1.000	1.000
Epidemic								
No Epidemic		1.000	1.000	1.000		1.000	1.000	1.000
Epidemic		2.686***	2.686***	2.751***		2.582***	2.582***	2.549***
Interaction								
Epidemic x Long-Term Consanguinity				0.845*				1.081
Epidemic x Close Consanguinity				0.331				1.693
N subjects	396,359				378,590			
N obs	570,380				545,982			
Log Likelihood (model)	-109,942	-109,316	-109,314	-109,311	-66,262	-65,912	-65,912	-65,911
AIC	219,887	218,635	218,634	218,631	132,528	131,826	131,830	131,832
BIC	219,910	218,646	218,668	218,688	132,550	131,837	131,863	131,888

Source: RPQA-IMPQ.

[1] Stratified by period, region-urban, and sex. N obs. = the number of observations including the multiple episodes of epidemics.

NPC=Non proportional close consanguinity by Schoenfeld residuals. NP=Non proportional model by Schoenfeld residuals.

Table A22. Consanguinity, Epidemics and Their Interaction, Stratified[1] Cox Models, 4 Years to 7 Years, Colonial Quebec, 1720-1830

Models stratified by Period Region-urban Sex	[4 years, 5 years[				[5 years, 7 years[			
	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 1	Mod. 2	Mod. 3	Mod. 4
Consanguinity								
Long-Term Consanguinity	0.968		0.967	0.959	1.02		1.018	1.037
Close Consanguinity	2.019***		2.022***	1.890*	0.984		0.98	0.882
No Consanguinity	1.000		1.000	1.000	1.000		1.000	1.000
Epidemic								
No Epidemic		1.000	1.000	1.000		1.000	1.000	1.000
Epidemic		2.762***	2.762***	2.738***		2.722***	2.722***	2.770***
Interaction								
Epidemic x Long-Term Consanguinity				1.054				0.8738
Epidemic x Close Consanguinity				1.449				1.6676
N subjects	365,653				355,115			
N obs	531,386				673,522			
Log Likelihood (model)	-45,221	-44,934	-44,930	-44,929	-56,751	-56,401	-56,401	-56,400
AIC	90,446	89,870	89,865	89,869	113,507	112,804	112,808	112,810
BIC	90,469	89,881	89,899	89,925	113,529	112,816	112,842	112,867

Source: RPQA-IMPQ. [1] Stratified by period, region-urban, and sex. N obs. = the number of observations including the multiple episodes of epidemics.

NPC=Non proportional close consanguinity by Schoenfeld residuals. NP=Non proportional model by Schoenfeld residuals.