The interplay between illness, time use, mobility and social contact behaviour and its relevance for infectious disease transmission.

Kim Van Kerckhove

Promotor: Prof. dr. Niel Hens
Co-Promotor: Prof. dr. John W. Edmunds

Dankwoord

Today a project of seven years has come to an important milestone: My Public Defense. I am proud and delighted to have reached this milestone, but meanwhile, I am convinced that this path wouldnŠt have been the same without an extensive list of wonderful and unique people. I want to take a moment to honour them, starting and ending with two essential men.

First and foremost I am indebted to my promotor, Niel. Niel, thank you for giving me the opportunity to start this PhD and to believe in me at that point and beyond. Throughout the years we had many moments, from office meetings to travels to Antwerp, London, Boston, Paris, ... The harmony between serious (or less serious) scientific meetings and more relaxed trips was lovely. In the more recent years, I experienced and highly appreciated your people oriented approach throughout our various interactions. I recall being surprised how well you know me, even related to things we never discussed. Niel, I am glad I had the chance to work with and learn from you. I am sure that many more will learn from you and that your three girls are blessed with you as their father. John, we met during my master thesis as I visited the London School of Hygiene and Tropical Medicine for the illness project. I am grateful to you for offering me the opportunity to grow into the field of epidemiology and more specifically social contact behaviour. Arriving in London was lovely thanks to your excellent welcoming, at that time together with Ken Eames. In the early years of my PhD, we met more frequently at the London School compared to later on. But when crossing each other's paths, it was always pleasant to meet. Thank you for your valuable input and for being critical, which offered opportunities for me to learn. I would also like to thank you and Ken for your kindness and warmth during all my visits at the London School and your critical questions when results were not as you expected them to be. This helped me grow in being (scientifically) critical. I am grateful for all the people I met in London who made me feel at home. This way I never felt alone. Vittoria, it was a pleasure to work with you, it was lovely, with scientific challenges in a warm and welcoming atmosphere in Paris and in Diepenbeek! Furthermore, I am grateful to the other jury members for taking time and making an effort to provide feedback to an earlier version of this thesis. To all my co-authors, it was a pleasure to work with and learn from you.

I am grateful to the beautiful people I met during the Master of Statistics. Carolina, I missed our geeky way of working, but it gave me a drive. Martin, Chella and Kathy I enjoyed going for lunch together, attending the World evening or meeting at JOSS activities with or without our families. I am grateful I could spend the past years at Censtat, meet a lot of nationalities and to be able contribute to JOSS. I enjoyed meeting each of you for a short hello or when we helped each other out. Furthermore, the “koffiegroep” was a welcoming and enjoyable group that helped me adjust and feel at ease at the university. An, we met for the first time in Vancouver during the JSM. From that moment it was obvious that you would be a warm and loving colleague. At work, it was clear that you were the mentor of the teaching assistants with all your experience and care for each of us! Nele, you helped me to get introduced to social contact data, and we always had interesting discussions related to the topic! An, Philippe, Ruth, Leen, Candida, Thomas, Stijn, Lisa, Steven, Jürgen, Kevin, Pieter and Maren, it was a pleasure to teach with you! Ruth, thank you for your patience with my numerous questions, your loving and friendly personality and your feedback to my Dutch summary! Leen, I am glad we shared similar experiences and that we found each other on those topics or had lots of recognitions ;-)! Lisa, by sharing an office the last few years, I got to know you better. I am happy I got the chance to see you as the loving, warm-hearted and considerate person you are. Thomas, your calmness helped me clear my mind from time to time during the last phase of my PhD, thank you!

To all the people I shared an office with: Girma, Birhanu, Amin, Leacky, Trias, Thomas and Lisa, thank you for the smalltalk or the help with research issues. Next, I would like to thank all SIMID colleagues for the fruitful discussions that were held. Lander, together we cleaned the Flemish data, which was an enormous work, but also led to exciting projects. I enjoyed our fruitful discussions, during meetings or travels. Niel and Philippe, thank you for giving me the opportunity to do the pilot study. It was a pleasant combination of research and organisation and helped me to find my future direction.

Doorheen de jaren zijn Koen en ikzelf gezegend geweest met veel lieve vrienden. Een enorme lijst en onmogelijk om er volledig in te zijn. Ik wil graag ieder van jullie bedanken voor de gezellige avonden, namiddagen, ... waarin we altijd weer plezier konden maken. Maar ook om over andere dingen dan het werk te praten, of soms juist wel het werk trachten uit te leggen. In het afgelopen jaar is de frequentie van bezoekjes echter drastisch verminderd, het is te zeggen waar ik bij was. Ik vind en vond het zeer spijtig om gezellige momenten te (moeten) missen, maar ben blij en trots hier vandaag te kunnen staan. Tegelijk ben ik ook zeer dankbaar voor jullie begrip en dat Koen en de kinderen toch fijne tijden konden beleven of halvelings mochten komen 'kamperen' (en ik goed kon doorwerken :-)). Tevens ben ik dankbaar voor de vele vormen van hulp die we van jullie kregen!. Ondanks de lange lijst wil ik graag toch enkele mensen vernoemen in een willekeurige volgorde. Sam & Evi, jullie zijn een prachtkoppel, het doet ons steeds deugd om met jullie te kunnen uitwisselen over opvoeding, het geloof, een doctoraat en een gezin, en om jullie steun te voelen! Jullie zijn voor ons als familie! Febe, ik ben blij en trots je meter te zijn en je vrolijkheid te ervaren of volgen via je mama en papa! Mieke, ik ben blij dat we (bijna) maandelijkse dates hielden om onze harten te luchten! Sanne & Wouter, de ontelbare ijsjes en de daarbij horende gezellige avonden hebben zeker voor de nodige ontspanning gezorgd! Bart, je bent er steeds weer bij de belangrijke momenten in ons leven en staat altijd klaar met je luisterend oor en begrip! Marie-Jeanne, een dikke knuffel voor jouw warme vriendschap en je oprechtheid! Katrien & Maarten, we hebben gezellige en leutige spelletjes avonden gehad, maar ook praktisch stonden jullie klaar om Fons van school te halen! Jessie & Bart, het doet deugd te kunnen spreken met een koppel waar de man ook zelfstandige is! Paul, je hielp mij toen het nodig was en gaf me een kader om de dingen die gebeuren te plaatsen! Moeke, je hebt altijd voor mij gesupporterd, al was het moeilijk te begrijpen wat ik juist deed. Ik hoop dat je er vandaag een tikkeltje meer van hebt begrepen! Nonke Luc, Tante Linda, Tante Stien en Nonke Danny, jullie toonden elke keer weer jullie oprechte interesse. Het moet niet evident geweest zijn om het proces te volgen, maar ik ben blij dat jullie ook mee-supporterden! Hopelijk was het spreektempo vandaag rustiger dan tijdens verhalen op familiefeesten ;-).

Aan ieder van jullie, vernoemd en niet vernoemd, een hartelijke dank je wel!

Deze thesis zou niet dit boek geworden zijn zonder de praktische hulp van jullie, Eddy en Vivianne! Bedankt dat jullie steeds klaar stonden om Koen en mij bij te staan. Samen met jullie wil ik ook graag de andere 'huybies' bedanken: Marijke, Vincent en Peter voor jullie interesse en de huybies-avonden. Maar ook voor jullie begrip over mijn afwezigheid! Gudrun & Karel, Evelyne & Sonni, Chris & Kathleen, ook aan jullie een welgemeende dank jullie wel voor het supporteren en de fijne babbels tijdens de VK-middagen. Maar ook bedankt voor de ontspannende momenten tussendoor zoals bv. het zwemmen met de kinderen. Evelyne & Sonni, bedankt voor de 'oppeper'-doos, dat ontroerde mij enorm! Mama en papa, zonder jullie had ik hier nooit kunnen staan! Dankzij de kansen die jullie mij gegeven hebben om verder te studeren, naar Londen te kunnen gaan voor mijn master thesis en de opvoeding die jullie mij gaven ben ik tot hier geraakt! Dank jullie wel!

Kinderen tijdens een doctoraat, er zijn vele meningen. Toch kozen wij ervoor, wisten wij veel dat het onze wereld zo door elkaar zou schudden :-). Fons, jouw komst bracht nieuwe dingen in mij naar boven. Hoewel het niet altijd gemakkelijk was ben ik ontzettend blij dat jij er bent! Met je verhalen en grapjes maak je mijn dag altijd weer goed. Je bent duidelijk een zeer attente en lieve jongen zowel voor je zus als voor ons! Julie, jij kan iedereen doen lachen, zo'n vrolijke meid. Je ondeugd straalt vaak genoeg van je af en doen me smelten. Je 'verhalen' worden al duidelijker en met sommige van je zinnen doe je onze harten altijd weer zwellen.
Fons & Julie, ik ben trots jullie mama te mogen zijn. Jullie waren voor mij een drijfveer om een doctoraat af te leveren om trots op te zijn!

Zoals steeds moet de belangrijkste persoon tot het laatste wachten ... Koen, ik weet niet wat ik zonder jou had moeten beginnen! Je hebt afgezien met mij en toch bleef je mij nabij en bleef je mij steunen. Jij bleef altijd in mij geloven, ook als ik dat zelf niet deed en ondanks mijn weerstand hierover! Jij stimuleerde mij op vele wijzen, ging zonder problemen met de kinderen op uitstap zonder dat ik erom moest vragen, stimuleerde mijn verblijf in Averbode, ... Jij was er bij de fijne momenten maar meer nog bij de moeilijke momenten. Jij hoort vandaag ook extra in de bloemetjes gezet te worden want dit is ook jouw verdienste. Helaas kan ik je geen doctoraatstitel geven maar wel de titel van SUPERMAN en, zoals op een T-shirt staat dat je ooit van mij kreeg: SUPERDAD! Lieve schat, ik hou ontzettend veel van je en ben je oneindig dankbaar! En je hebt (altijd) gelijk: alles komt goed!

Kim Van Kerckhove

Publications

The materials presented here, are based on the following publications and reports:

Van Kerckhove, K., Hens, N., Edmunds, J.W., Eames, K.T.D. (2013). The impact of illness on social networks: Implications for transmission and control of influenza. *American Journal of Epidemiology* 178 (11): 1655-62.

Willem, L., **Van Kerckhove, K.**, Chao, D.L., Hens, N., Beutels, P. (2012). A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS One* 7 (11): e48695.

Kifle, Y.W., Goeyvaerts, N., **Van Kerckhove, K.**, Willem, L., Kucharski, A., Faes, C., Leirs, H., Hens, N., Beutels, P. (2016). Correction: Animal ownership and touching enrich the context of social contacts relevant to the spread of human infectious diseases. *PLoS One*, 11 (2): e0148718.

Santermans, E., **Van Kerckhove, K.**, Azmon, A., Edmunds, W.J., Beutels, P., Faes, C., Hens, N. (2017). Structural differences in mixing behaviour informing the role of asymptomatic infection and testing symptom heritability. *Mathematical Biosciences* 285: 43-54.

Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., **Van Kerckhove, K.**, Willem, L., Beutels, P., Hens, N. Household members do not contact each other at random: Implications for infectious disease modelling. In preparation.

De Luca, G., **Van Kerckhove, K.**, Coletti, P., Poletto, C., Bossuyt, N., Hens, N., Colizza, V. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. Submitted to BMC Infectious Diseases.

Van Kerckhove, K., Willem, L., Faes, C., Beutels, P., Hens, N. The dispersion of social contacts and their relevance for the construction of epidemic models. In preparation.

Van Kerckhove, K. *et al.* The impact of influenza-like illness and chronic diseases on social contact patterns relevant for infectious disease transmission - a pilot study. In preparation.

Publications not covered in this dissertation:

Beutels, P., Vandendijck, Y., Willem, L., Goeyvaerts, N., Blommaert, A., **Van Kerckhove, K.**, Bilcke, J., Hanquet, G., Neels, P., Thiry, N., Liesenborgs, J., Hens, N. (2013). Seasonal influenza vaccination: prioritizing children or other target groups? Part II: cost-effectiveness analysis. *KCE Report, Health Technology Assessment*, 204.

Goeyvaerts, N., Willem, L., **Van Kerckhove, K.**, Vandendijck, Y., Hanquet, G., Beutels, P., Hens, N. (2015). Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence. *Epidemics* 13: 1-9.

Hens, N., Abrams, S., Santermans, E., Theeten, H., Goeyvaerts, N., Lernout, T., Leuridan, E., **Van Kerckhove, K.**, Goossens, H., Van Damme, P., Beutels, P. (2015). Assessing the risk of measles resurgence in a highly vaccinated population: Belgium anno 2013. *Eurosurveillance* 20 (1), pii=20998.

Contents

List of abbreviations

AIC	Akaike Information Criteria
BC_a CI	Bias-corrected and Accelerated Confidence Interval
BE	Belgium
BIC	Bayesian Information Criteria
CI	Confidence Interval
DE	Germany
DIC	Deviance Information Criterium
ECDC	European Centre for Disease Prevention and Control
EDF	Empirical Distribution Function
ERGM	Exponential-family Random Graph Model
EQ-5D-3L or EQ-5D	Standard Euroqol Questionnaire
FI	Finland
GB	Great Britain
GEE	Generalised Estimating Equations
GLMM	Generalised Linear Mixed Model
HIC	High-income Country
HRQoL	Health-related Quality-of-Life
ILI	Influenza-like Illness
IQR	Inter Quartile Range
IT	Italy
LRT	Likelihood Ratio Test
LU	Luxembourg
MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
NA	Not Applicable
NL	the Netherlands
ODE	Ordinary Differential Equation
OR	Odds Ratio
PL	Poland
RFID	Radio-Frequency Identification
RNC	Relative Number of Contacts
SE	Standard Error

SEIR	Susceptible-Exposed-Infected-Recovered
SIR	Susceptible-Infected-Recovered
UK	United Kingdom
VAS	Visual Analogue Scale
VZV	Varicella-Zoster-Virus
WAIFW	Who-Acquires-Infection-From-Whom
WHO	World Health Organisation
WLS	Weighted Least Squares

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Paper-based social contact surveys

To quantify the possible reduction in the cost and burden of infectious diseases, such as influenza, measles and varicella, to the community by implementing intervention strategies, it is important to use realistic, yet parsimonious models. These models can be mathematical models that aim to reflect the transmission process of these diseases. Social contact behaviour relevant for infectious diseases, collected through social contact surveys, can prove to be valuable to inform the transmissions process. These social contact surveys collect information on who contacts, and possibly infects, whom. Currently, only a few mathematical models use this contact information, and even fewer allow the behaviour to change. As a result, these models are not completely realistic. In this thesis, we quantify changes in social contact behaviour as a result of illness, time use and mobility. Our results can improve the realism of models, which can then better predict the effects of interventions, such as vaccination or isolation, and hence reduce the burden and cost of diseases. In this first section, we elaborate on the link between transmission and social contact surveys, starting from three examples of airborne infectious diseases.

Seasonal influenza is an acute respiratory infection caused by influenza viruses. The symptoms are a sudden onset of fever, cough (usually dry), headache, muscle and joint pain, severe malaise, sore throat and a runny nose. The fever and the other symptoms, except for the cough, usually last a week without the need for medical attention. On the other hand, the cough can be severe and can last for two or more weeks (World Health Organization, 2017a). According to the European Centre for Disease Prevention and Control (ECDC) (2017a), seasonal influenza, which is vaccine preventable, causes approximately 4-50 million symptomatic in the EU/EEA each year with 15,000 - 70,000 European citizens who die every year of causes associated with influenza. Older people, younger children and those with chronic conditions are most at risk of developing serious complications, such as dehydration, infections of the respiratory tract, and worsening of chronic illnesses (Wetenschappelijk Instituut voor Volksgezondheid, 2017), that might result in death.

Measles is another very contagious infectious disease, affecting mostly children. The initial symptoms are high fever, runny nose, bloodshot eyes, and tiny white spots on the inside of the mouth. These spots usually appear 10-12 days after the infection. The most visible symptom, however, is the rash spreading over the body, occurring several days later. The World Health Organization (WHO) (2017b) states that no specific treatment exists and that the recovery period takes 2-3 weeks. In malnourished children and people with reduced immunity, serious complications can present, including blindness, encephalitis, severe diarrhoea, ear infection and pneumonia. According to the WHO, measles is one of the leading causes of death among young children, even though a safe and cost-effective vaccine is available.

Another childhood infection, varicella, is caused by the varicella-zoster virus (VZV) and presents itself at the beginning with cold-like symptoms, which are followed by a raise in temperature and a very itchy, blister-like rash. The rash appears over 3-5 days on the trunk of the body, and the severity varies from person to person. When you previously had varicella, the virus can be reactivated later in life and cause shingles, which begins, in older people, with pain in the area of the affected nerve, followed by a rash of blisters in the affected area. This rash usually lasts seven days, but the pain can last longer. Even though varicella is a mild disease, complications can occur, predominantly in infants, adults, pregnant women and immunocompromised individuals. These complications can include infection of the skin, pneumonia and encephalitis. In pregnant women, varicella can cause miscarriage, congenital varicella syndrome or neonatal varicella. Congenital varicella syndrome can leave the baby with severe, life-long physical and learning disabilities. By avoiding contact with people with varicella or shingles, people can dodge varicella infection. When a person suffers from varicella or shingles, they are advised to stay at home from work or school until all rash blisters have crusted and no new ones appear (ECDC, 2017b).

Each of these examples results in varying degrees of burden in different people, and the complications lead to additional costs and deaths. Furthermore, suggestions are offered to reduce or avoid the infections, including vaccination and isolation. However, when introducing new containment measures, one has to take precautions regarding the consequences. Testing the strategies and quantifying their effect and consequences by executing them directly in the public would be unethical; thus, mathematical models come into play. The aim of these mathematical models is to represent the complexities of reality in a structured way and thus to be able to vary certain aspects related to containment strategies. The models are based on existing information, such as the duration of infection, e.g., 2-3 weeks for measles, and attempt to mimic the transmission of infections. Although some strategies have already been implemented or advised, research to make these mathematical models as realistic as possible is still of interest. Additional research allows for improvements regarding the social contact behaviour employed or in allowing for co-infections; the continuation of research furthermore allows for the study of various containment strategies or the study of potential future outbreaks.

As these mathematical models aim to reflect the transmission process, it is important to understand this process. Brankston *et al.* (2007) define four modes of transmission:

(1) direct contact, when transmission occurs from direct physical contact between an infected and a susceptible individual;

(2) indirect contact, when transmission occurs via an intermediate object;

(3) droplet, transmission occurs via large droplets generated from the respiratory tract of the infected individual during coughing or sneezing or talking. These droplets are propelled a distance of less than 1 m through the air and are deposited on the nasal or oral mucosa of the new host or in their immediate environment. The droplets do not remain suspended in the air;

(4) airborne, transmission occurs via the dissemination of microorganisms that remain suspended in the air for long periods of time. The organisms are widely dispersed by air currents and inhaled by susceptible hosts who may be some distance away from the source patients or individuals.

Based on these definitions, we can describe the transmission of the above examples as being through direct contact or through droplets given that they either require direct contact or that transmission occurs at a close distance through talking, coughing or sneezing. To reflect these transmissions, we use a matrix with information about Who-Acquires-Infection-From-Whom (WAIFW). In the beginning, these WAIFW matrices were assumption- and parameter-driven, wherein the homogeneous assumption was the simplest. This assumption implies that everyone can transmit the infection to anyone else with the same probability, which is often unrealistic. Two decades ago, the first attempt to collect data on social interactions was published by Edmunds *et al.* (1997). These authors requested participants to fill in their contacts during one randomly assigned day with additional information, such as age, gender, and duration. They suggested that these data could be used to obtain realistic WAIFW matrices, and hence the first social contact survey, albeit small, was conducted. Other authors followed their strategy, including Beutels *et al.* (2006); Edmunds *et al.* (2006); Wallinga *et al.* (2006). The latter authors proposed the 'social contact hypothesis', in which the transmission rates are assumed to be proportional to the social contact rates.

The data obtained in these surveys have been shown to be relevant to quantifying the transmission of infectious diseases (Wallinga *et al.*, 2006; Rohani *et al.*, 2010; Melegaro *et al.*, 2011; Goeyvaerts *et al.*, 2010), of which the last two articles use data based on a large-scale social contact survey (Mossong *et al.*, 2008). Hoang *et al.* (In preparation) performed a systematic review of social contact surveys with a focus on surveys using a contact-diary approach. These authors found 71 articles related to 23 surveys. The surveys predominantly took place in developed countries, and the authors denote varying definitions of contacts, different sampling strategies and various collection tools.

Other studies have also quantified social contact behaviour using different instruments, such as proximity sensors (Cattuto *et al.*, 2010; Salathé *et al.*, 2010; Stehlé *et al.*, 2011), or based on time use data (Zagheni *et al.*, 2008; De Cao *et al.*, 2014). Read *et al.* (2012) highlighted research in the area (before 2012) and narratively reviewed the various methodologies and instruments to quantify social interactions relevant to airborne infections. These authors indicate the limitations of social contact diaries, including potential biases in participant recall and reporting, study fatigue when collecting longitudinal information and limitations of the definition of a contact as the reported encounters are only a subset of all potential contacts that permit transmission. Despite these limitations, they indicate that the method has important advantages compared to other methods: it is easy to administer, capable of collecting information retrospectively, applicable in a wide range of settings and, as stated before, able to explain age-based patterns of infections (Wallinga *et al.*, 2006; Rohani *et al.*, 2010; Melegaro *et al.*, 2011; Goeyvaerts *et al.*, 2010; Ogunjimi *et al.*, 2009).

In this thesis, we restrict ourselves to paper-based social contact surveys that aim to record social contact behaviour relevant to the spread of airborne infectious diseases. We aim to quantify changes in social contact behaviour as the result of various factors. In the following section, we explain the basic concepts used throughout this thesis, and in Section 1.3, we give an overview of the thesis.

1.2 Basic concepts

1.2.1 Introduction

Throughout this work, we use different concepts, both epidemiological and statistical. In this section, we also shed some light on the mathematical models that we use when describing the transmission of infectious diseases. Additionally, we aim to set the scene by introducing the notation used in the remainder of this work.

1.2.2 Mathematical models

Vynnycky and White (2010) argue that a model is a simplified representation of the complex truth or a complex phenomenon. As stated before, models are used to avoid ethical issues but are also driven by the fact that research in human populations is difficult and expensive. Mathematical models are one of the kinds of models introduced by Vynnycky and White (2010), and they describe population parameters using symbols and link them using algebraic formulae, usually a set of differential equations. As a result of their representation, they can be analysed using computers, demonstrating their usefulness. The aim of these models is to mimic the real world as closely as possible.

When simplifying the three examples of infectious diseases from before, e.g., seasonal influenza, measles and varicella, we can explain the evolution of the states an individual can go through as follows: initially, all individuals are susceptible and hence prone to infection. When infection is introduced in the population, infected individuals become infectious for a certain period ν^{-1}, and afterwards, they recover and become immune, and thus these individuals do not contribute to the transmission process any longer. Each of these states is described by a letter: S for susceptible, I for infected and R for recovered or immune. Given the serial nature of events, one represents the process by differential equations that describe the flow between states, and the model is called an SIR model. Figure 1.1 shows a flow diagram of this SIR model, wherein the arrows indicate the flow from one state to another. In this figure, we assume that every newborn is born in the susceptible state. In addition to the previous description, we include deaths: μ reflects the death due to natural causes, and α is the infection-related death rate. The parameter λ reflects the transmission rate; the mass-action principle states that the number of new infections at a given time point is proportional to all possible contacts between infectious and susceptible individuals at the previous time point. Based on this assumption, λ is given by βI, with β being the transmission rate, representing the probability per susceptible person of having a contact that leads to a transmission event and is often referred to as the effective contact rate. λ is often referred to as the force of infection, expressing the rate at which susceptible individuals become infected.

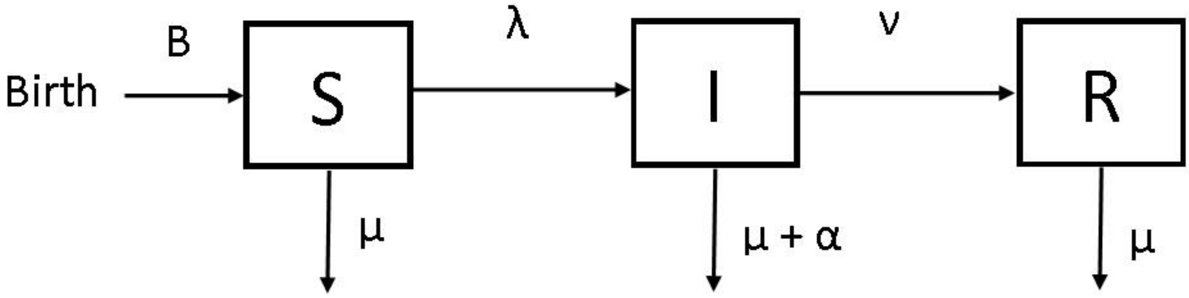


Figure 1.1: Flow diagram for the SIR model. The individuals enter the susceptible state, then move to the infected state (at rate λ), and after recovery, they move into the immune state (at rate ν). Figure adapted from Hens *et al.* (2012).

The dynamics of the flows shown in Figure 1.1 can also be described by a set of partial differential equations:

$$\begin{cases} \frac{\partial S(a,t)}{\partial a} + \frac{\partial S(a,t)}{\partial t} & = -(\lambda(a,t) + \mu(a,t))S(a,t), \\ \frac{\partial I(a,t)}{\partial a} + \frac{\partial I(a,t)}{\partial t} & = \lambda(a,t)S(a,t) - (\nu(a,t) + \alpha(a,t) + \mu(a,t))I(a,t), \\ \frac{\partial R(a,t)}{\partial a} + \frac{\partial R(a,t)}{\partial t} & = \nu(a,t)I(a,t) - \mu(a,t)R(a,t), \end{cases} \tag{1.1}$$

where $S(a,t)$, $I(a,t)$ and $R(a,t)$ are the age- and time-specific numbers of susceptible, infected and recovered individuals, respectively. The boundary conditions are given by $S(0,t) = B(t)$, $I(0,t) = R(0,t) = 0$, thus ruling out the vertical transmission of infection. The initial conditions are given by $S(a,0) = \tilde{S}(a)$, $I(a,0) = \tilde{I}(a)$ and $R(a,0) = \tilde{R}(a)$, where $\tilde{S}(a)$, $\tilde{I}(a)$ and $\tilde{R}(a)$ are prescribed functions of age such that they sum to $N(a,0)$, the age-specific population at time 0. It is often assumed that both the natural and infection-related death rates are independent from time, thus $\mu(a,t) = \mu(a)$ and $\alpha(a,t) = \alpha(a)$, and the recovery rate is often assumed to be constant $\nu(a,t) = \nu$.

For simplification, one often assumes homogeneity with age, implying that the parameters in Eq. (1.1) are assumed to be constant with age. One refers to this simplification as the age-homogeneous model. However, one can consider different modifications of the simple SIR model: one can include additional compartments, such as the exposed compartment, consisting of individuals that are infected but not yet infectious, or a vaccinated compartment, containing vaccinated individuals. Depending on the choices, this last model can yield a reduced rate of infection with vaccination, or one can also reflect that a proportion of people will be fully protected and hence moved to the recovered compartment.

1.2.3 Epidemiological parameters and their estimation

As pointed out previously, the information from the social contact surveys is used to infer the structure of transmission networks, as these are impossible to observe directly. The implicit assumption in this survey approach is that the characteristics (such as age) of individuals you talk to are similar to the characteristics of individuals you might expose to infection (Wallinga *et al.*, 2006), and that people en masse report accurately and reliably about the average communication pattern in the population (Wasserman and Faust, 1994).

There are relatively few approaches to the statistical analysis of such social contact datasets. A first approach to the statistical estimation of contact rates by age from social contact data involves binning the contact data by age categories and then applying a maximum likelihood estimation of rates using a likelihood function that is constrained such that reciprocity of contacts is guaranteed (Wallinga *et al.*, 2006). We refer to this method as the 'piecewise constant surface approach'. A second approach involves estimating a smooth tensor product spline, without (Mossong *et al.*, 2008) or with (Goeyvaerts *et al.*, 2010) guaranteed reciprocity of contacts. We refer to this method as the 'bivariate smoothing approach'.

Throughout this thesis, we make a few assumptions about the structure of the network: we adapt an egocentric point of view, considering only the first ring of contacts of participants. The resulting data gives a first order approximation of the social network. We assume mass action, and as such, everyone experiences an average social environment given his or her demographic characteristics, such as age. We also assume that we can extrapolate the duration of the time interval over which we measure contact rates to a suitable time interval during which infection can be transmitted. Lastly, we assume that heterogeneities, other than demographics and observed characteristics, are unimportant (Farrington *et al.*, 2001; Hens *et al.*, 2009c).

1.2.3.1 The degree distribution

Denote Y_t as the random variable that represents the number of contacts made by participant t ($t = 1, \ldots, T$), and let $\mathbf{x}_t$ be a vector of covariates expressing participant characteristics, such as age, gender, and household size. A natural approach to modelling the degree distribution conditional on a set of covariates is to use a negative binomial regression model allowing for overdispersion, i.e., allowing the variance to exceed the expected value in a Poisson regression model where the variance equals the mean. We can model the mean m_t as:

$$\log(m_t) = \beta_0 + \mathbf{x}_t'\boldsymbol{\beta}, \tag{1.2}$$

with $\boldsymbol{\beta}$ being the coefficient vector associated with the different covariates and ϕ being the overdispersion parameter, which leads to a negative binomial variance $m_t + m_t^2/\phi$. The use of the log-link function yields an easy derivation of the expected relative number of contacts (RNC) and corresponding 95% confidence intervals (CI) as $\exp(\boldsymbol{\beta})$ and 95% CI limits: $\exp(\hat{\boldsymbol{\beta}} \pm 1.96 \times \text{s.e.}(\hat{\boldsymbol{\beta}}))$. A weighted likelihood approach is considered to ensure representative results by using post-stratification weights w_t^d. These post-stratification weights, also called diary weights, allow for the comparison of the results of different surveys, as using them yields estimates of population-based quantities.

The calculation of these weights depends on the survey due to design aspects and is based on the factors for which the sampling might not have been representative. Common factors include age and household size, day of the week and holiday or regular periods. In a case where each of these factors is present, the weighting takes the following form:

$$\tilde{w}_t^d = \frac{P_{pop}(age_t \cap hhsize_t) P_{pop}(day_t \cap regular_t)}{P_{data}(age_t \cap hhsize_t) P_{data}(day_t \cap regular_t)}, \tag{1.3}$$

$$w_t^d = \frac{\tilde{w}_t^d}{\sum_{l=1}^{N} \tilde{w}_l^d} \times N, \tag{1.4}$$

where N is the total sample size of participants in the considered analysis, and $P_{pop}(.)$ is the population distribution of the quantities within the brackets. $P_{data}(.)$ refers to the observed distribution. Whenever a particular factor is not present for a participant, we remove the related probabilities (from the population as well as from the data) from the weight. Note that the sum in the denominator of Eq. (1.4) includes these last weights. All weights are capped at 2 (unless stated otherwise) to prevent any participant from disproportionately affecting the results.

It is important to note that by using the weights including week, weekend and regular, holiday period information, we assume our results to be representative for one year, whereas often the collection of data occurred over a limited period of the year. In the next section, we describe how contact characteristics such as age can be taken into account when studying the degree distribution.

1.2.3.2 Mixing patterns

Let y_{ijt} be the reported number of contacts made by participant t in age class i with people in age class j during one day for $t = 1, \ldots, T_i$, with T_i being the number of participants in age class i ($i, j = 1, \ldots, K$, with K being the number of age classes). $\mathbf{M}$ denotes the social contact matrix, with elements m_{ij} representing the expected number of contacts with people in age class j during 1 day as reported by a respondent in age class i. In contrast to the piecewise constant surface and the bivariate smoothing approach, we estimate the elements m_{ij} by:

$$\hat{m}_{ij} = \frac{\sum_{t=1}^{T_i} w_t^d y_{ijt}}{\sum_{t=1}^{T_i} w_t^d}. \tag{1.5}$$

Now, the per capita contact rate c_{ij}, which is the per capita rate at which individuals of age class j make contact with persons of age class i per day, are related to the social matrix by:

$$c_{ij} = \frac{m_{ij}}{w_j}, \tag{1.6}$$

where w_j gives the population size in age class j and is based on demographic data. When contacts are expected to be reciprocal, e.g., in cases where participants and contacts experience the same situation, this needs to be taken into account (Wallinga *et al.*, 2006) by imposing $m_{ij}w_i = m_{ji}w_j$, implying that the total number of contacts of age class i with age class j must equal the total number of contacts of age class j with age class i. This is achieved using the following equation:

$$c_{ij} = \frac{m_{ij}w_i + m_{ji}w_j}{2w_iw_j}. \tag{1.7}$$

Given the same notation for symmetrical and asymmetrical matrices, the choice of either matrix in this work is supported by the context. In general, reciprocity is necessary when the selection of the contact matrix is driven by contact properties, e.g., the duration of contacts. Selection based on the patient properties of the contacts, on the other hand, leads to non-reciprocal matrices, e.g., the location of contacts.

1.2.3.3 Social contact matrices and the spread of an epidemic

Based on the estimated transmission rates, we predict how an infectious agent like the influenza virus spreads in a population. This spread can be described using different parameters, mainly derived from the next-generation matrix $\mathbf{G}$ with elements g_{ij} representing the average number of secondary infections in age class i resulting from the introduction of a single infectious individual of age class j into a population (Diekmann *et al.*, 1990). This next-generation matrix describes how infection passes within and between age classes and we define it as

$$g_{ij} = m_{ij}s_iDq, \tag{1.8}$$

where s_i is the fraction of group i that is susceptible to infection, D is the average duration of infectiousness, and q is the disease-specific (proportionality) factor, which potentially measures the infectivity but also expresses differences related to susceptibility or discrepancies between the social contact proxies measured in the contact survey and the true contact rates underlying infectious disease transmission (Goeyvaerts *et al.*, 2010). In this thesis, we mainly consider q to be constant, notwithstanding the existing literature taking age-specific proportionality factors into account (Goeyvaerts *et al.*, 2010; Ogunjimi *et al.*, 2009; Wallinga *et al.*, 2006). An alternative formulation of the next-generation matrix, wherein $s_i = 1$ is assumed, is given by:

$$g_{ij} = c_{ij}w_iDq. \tag{1.9}$$

Using the next-generation matrix, we can estimate the basic reproduction number R_0 as the dominant eigenvalue of $\mathbf{G}$ when the entire population is susceptible (i.e., s_i=1). The basic reproduction number R_0 represents the number of secondary infections when introducing an 'average' infectious individual into a fully susceptible population (Diekmann *et al.*, 1990). This number serves as a threshold value, whereby a value below one results in extinction of the infection, and a value above one can result in an outbreak.

The normalised dominant eigenvector of the next-generation matrix gives the stable age distribution of cases during the initial growth phase of the epidemic (Diekmann *et al.*, 2013) and thus an estimate of the relative incidence across age groups.

The force of infection $\lambda_i(t)$ gives the rate at which a susceptible person of age class i acquires infection at time t. We denote β_{ij} as the age-specific transmission rates from age class j to age class i. These transmission rates can be linked to the social contact rates c_{ij} via the so-called social contact hypothesis, as proposed by Wallinga *et al.* (2006), stating:

$$\beta_{ij} = qc_{ij} \tag{1.10}$$

with q being a proportionality factor as described when the next-generation matrix was introduced.

The force of infection is linked to the age-specific transmission rates through the following equation:

$$\lambda_i(t) = \sum_{j=0}^{K} \beta_{ij} I_j(t), \tag{1.11}$$

with $I_j(t)$ representing the total number of infectious individuals in age class j at time t. Note that we implicitly assume time-independence for the transmission rates. This might be a strong assumption, given the existing literature showing variation over time in social contact patterns (Hens *et al.*, 2009a; Eames *et al.*, 2010, 2011) as well as in infections. Nevertheless, extensions to time-dependent transmission rates are possible and will be discussed when necessary in this thesis. Eq. (1.11) can also be re-expressed in terms of the social contact matrix $\mathbf{M}$ as follows:

$$\lambda_i(t) = \sum_{j=0}^{K} q m_{ij} \iota_j(t), \tag{1.12}$$

with $\iota_j(t)$ being the relative frequency of infectious individuals in age class j at time t.

1.2.4 Statistical inference

Standard regression techniques, such as generalised linear models, assume that observations are independent. Whenever we consider the number of contacts of people when they were ill or when they were healthy, or when we consider the number of contacts at different locations or distances from home, or of different durations, each participant contributes multiple times, and as such, the observations are dependent, and the standard regression techniques are not appropriate. While the standard regression techniques typically leave the consistency of point estimation intact, this is not the case for measures of precision. In the event of a 'positive' clustering effect (i.e., the number of contacts from a participant when ill or healthy are more alike than the number of contacts from other participants), ignoring this aspect of the data will lead to an overestimation of the precision and an underestimation of the standard errors and lengths of confidence intervals. In the first paragraph, we discuss the adaptations of standard regression techniques to account for the clustering effect. In the second paragraph, we discuss the bootstrap inference, a technique that can also be used and is of interest when the primary interest is in confidence intervals.

1.2.4.1 GEE/GLMM

Different strategies for accounting for clustering in a regression setting exist (Molenberghs and Verbeke, 2005) and three model families are generally distinguished for discrete longitudinal data. Here, we will briefly discuss two of these families. In the marginal models, the parameters characterise the marginal expectation (e.g., the marginal number of contacts) of a subset of the outcomes without conditioning on other outcomes. These models typically capture the association structure via association parameters, including correlations, and odds ratios. Maximum likelihood can be unattractive due to excessive computational requirements; hence, alternative methods are proposed, such as generalised estimating equations (GEE). GEE only requires the correct specification of the univariate marginal distributions, provided one is willing to adopt "working" assumptions about the association structure. Furthermore, GEE yields consistent and asymptomatically normal estimators with an empirically corrected variance estimator. For additional alternative methods and more information on the GEE method, we refer to Molenberghs and Verbeke (2005).

The other model family contains the subject-specific models, which include parameters specific to the subject. In these models, the responses are assumed to be independent given the subject-specific parameters. For example, when the response is the number of contacts, the effect of covariates on this response is conditional on the level of the subject-specific effect. Hence, a unit difference in the covariate translates to an appropriate change in the number of contacts, keeping the level of the subject-specific effect fixed. If the interest is in within-subject changes, these models are useful. One can deal in different ways with the subject-specific parameters, amongst which one includes random-effects, for which one can assume the parameter vector to be normally distributed. These models are called generalised linear mixed models (GLMM) (Breslow and Clayton, 1993), implying that the random effects operate linearly at the level of the linear predictor. One can fit these models via the maximum likelihood method with approximation techniques to overcome computational difficulties. For other approaches to introducing subject-specific effects and for more information, we refer to Molenberghs and Verbeke (2005).

1.2.4.2 Bootstrap inference

Bootstrap inference is a re-sampling method for assessing the distributional properties of an estimator, including bias, variability, quantiles and percentiles (Efron and Tibshirani, 1993; Davison and Hinkley, 1997). When standard inference techniques are complex or unavailable, they are even more useful. Nonetheless, they can be applied to standard approximations as well.

We denote $y_1, \ldots, y_n$ as a sample from an unknown underlying distribution F. The interest is in the estimation and inference about the parameter θ. A key principle is to replace the unknown F with a known estimate $\hat{F}$ by generating and simulating bootstrap data, denoted as $y_1^*, \ldots, y_n^*$, from $\hat{F}$ to imitate the original data generation mechanism ($y_1, \ldots, y_n$ from F). Based on the bootstrap data, we can estimate the parameter of interest $\hat{\theta}^*$. We perform this process repeatedly, e.g., $B = 1000$ times, which results in B copies of the parameter of interest $\hat{\theta}_1^*, \ldots, \hat{\theta}_n^*$, which allows for the observation of distributional characteristics. Efron and Tibshirani (1993) state that in many cases, the bootstrap approach has been shown to be an appropriate approximation of the true distribution and that it might even be superior to typical approximations. Furthermore, these authors state that there are settings in which the standard application of the bootstrap principle fails. In this thesis, we consider both parametric and nonparametric bootstrap inferences.

In a nonparametric bootstrap, we generate bootstrap samples $y_1^*, \ldots, y_n^*$ from the empirical distribution function (EDF) $\hat{F}_{EDF}$. This comes down to sampling from the original data $y_1, \ldots, y_n$ with replacement. When doing this, we put a weight of $1/n$ on each of the observed y_i. In a parametric bootstrap, we assume a particular parametric distribution $\hat{F}$, based on estimated parameters, from which to bootstrap data. It is important to choose the appropriate distribution in this case, as misspecification of the generating distribution may lead to substantial bias. Efron and Tibshirani (1993) and Davison and Hinkley (1997) give a more in-depth description of the bootstrap methods.

One way of describing the distributional characteristics of the bootstrap sample is via a confidence interval or in this case, a bootstrap CI. The simplest type is the normal bootstrap CI, in which the quantiles from the simulated normal distribution replace the normal critical points. Another approach is the percentile bootstrap CI, which is given by

$$[\hat{\theta}^*_{(\alpha/2)}, \hat{\theta}^*_{(1-\alpha/2)}] \tag{1.13}$$

and is predominantly used in this thesis. However, we also consider an improved version which employs the bias-corrected and accelerated intervals (BC$_a$ CI) to reduce bias. The calculation of these intervals is also based on percentiles but is slightly different as the percentiles depend on two parameters, the acceleration and the bias-correction. The first parameter describes the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter θ and needs to be estimated as well, for example, via jackknife values of $\hat{\theta}$. The BC$_a$ method has two advantages: the first is that it respects transformation, implying that a transformation of the parameter of interest θ can be directly applied to the boundaries of a BC$_a$ interval. The other advantage is accuracy, which is higher in this case. Efron and Tibshirani (1993) expand more on this method and its advantages.

1.3 Overview of the thesis

From the first part of this chapter, it is clear that social contact surveys have become increasingly popular. In 2005-2006, a large-scale social contact study, the POLYMOD Study, was performed in eight European countries (Mossong *et al.*, 2008), allowing for the quantification of social contact behaviour more rigorously. We briefly introduce and describe the POLYMOD Study in Chapter 2 (Section 2.2). Various authors used these data to quantify temporal differences (week vs. weekend and regular vs. holiday periods) (Hens *et al.*, 2009a,b), while other studies were set up to quantify the effect of school closure on social contact behaviour (Eames *et al.*, 2010, 2011). In this thesis, we go beyond these aspects by studying changes in social contact behaviour as a result of various aspects, including illness, weather, animal ownership and distance.

Read *et al.* (2012) discuss different challenges related to social contact surveys in general. One of them is "the uncertainty in how contact patterns may change in response to infection or control" (Read *et al.*, 2012). These authors indicate the importance of characterising this influence when modelling epidemics. Similarly, Eames *et al.* (2010) performed a specifically designed survey to gain insights. In this thesis, we use these data to quantify the change in social contact behaviour as a result of illness. We therefore introduce the English Illness Survey, as established by Eames *et al.* (2010), in Chapter 2 (Section 2.3). In Chapter 3, we describe and quantify the resulting changes and discuss the consequences of using a simple method, as described in Van Kerckhove *et al.* (2013). I contributed to this paper as a first author by data cleaning, data quality check, statistical analysis and in the writing of the paper. In Santermans *et al.* (2017), we extend the simple model in a follow-up study (Section 3.2) using adaptations of the SEIR (Susceptible-Exposed-Infected-Recovered) model to incorporate symptomatic and asymptomatic infections and by keeping track of the status of the infector. As a result, we challenge the symptom heritability assumption. I contributed to this paper by discussing the set-up of the models, critical appreciation of the analysis and results and of the paper. After the initial survey, there remains a lack of knowledge about the impact of other infections as well as the influence of chronic conditions on social contact behaviour. Given that the first study was conducted based on a convenience sample, the results might not be generalizable to the whole population. I, therefore, conducted a pilot study to gain insight into the feasibility of a representative study to quantify the change in social contact behaviour as a result of seasonal influenza and/or chronic conditions. I introduce this pilot study in Chapter 2 (Section 2.5), and report the results and conclusions in Chapter 3 (Section 3.3). It is good to keep in mind that setting up and performing such a study takes a considerable amount of time, disproportionate to the number of pages used to describe it. I was involved in each part of the process of this last study from the design to the analysis and writing.

In 2010-2011, a new social contact survey was conducted in Flanders (Belgium), which, in part, has been described by Willem *et al.* (2012) and Kifle *et al.* (2016), that incorporated several novelties compared to the POLYMOD Survey, allowing to answer a large number of different research questions. A considerable amount of time was necessary to obtain a clean dataset, and as a result, we can investigate a variety of research topics. As part of this thesis, we present some of the work in Chapter 4. Each of the aspects discussed in that chapter is novel compared to existing methods. We address the influence of weather conditions on social contact behaviour by comparing the mean number of contacts as well as the R_0 ratios (Section 4.2) and give advice on aspects to be taken into consideration when setting up a survey to study the interplay between weather and social contact behaviour. Furthermore, we discuss the factors influencing animal ownership or animal touching and their impacts on the number of contacts, as well as a method to quantify the probability of a major zoonotic outbreak in different age classes (Section 4.3). As the Flemish Survey consisted of two parallel surveys, of which one surveyed complete households, we are able to quantify the within-household networks by using Exponential-family Random Graph Models (ERGMs) (Section 4.4). In this section, we also apply a two-level mixing model to challenge the commonly made random mixing assumption within households. I contributed to the work in Chapter 4 by data cleaning, data quality checking and critical revision of the work given the knowledge I acquired when cleaning and checking the data.

Read *et al.* (2012) depicted a lack of studies measuring social contact behaviour with geographical space. Based on the Flemish Survey, we can assess the combination of social contact behaviour and geography. We present the analyses concerned with social contact behaviour and geographical aspects in Chapter 5. In this chapter, we present two approaches to this aspect, the first relating social contact behaviour to distances from home, and the second part (Section 5.2) implementing social contact behaviour in a spatially explicit meta-population model. Here, we further attempt to link the model results to the infection risk and quantify the impact of school closures using holidays as well an extension of existing holidays. I conducted the analysis and wrote the work of the first part of this chapter and contributed to the last part of the chapter by providing data and critical appreciation of the results and the article.

Chapter 2

Social Contact Surveys

2.1 Introduction

In this chapter, we give an introduction to each of the paper-based contact surveys used or conducted as part of this thesis. To begin, we describe the common elements in each of these surveys. Section 2.2 describes the first large-scale social contact survey, POLYMOD, in more detail, Section 2.3 elucidates the details of the first social contact survey to quantify the differences in contact behaviour due to infection. In Section 2.4, we describe a social contact survey conducted in Flanders in 2010-2011, and in Section 2.5, we describe a pilot study, ICSOC, to quantify changes in social contact behaviour due to acute and chronic illnesses.

These types of surveys aim to collect detailed information about the social contact behaviour of individuals throughout the day. To do so, participants fill in a questionnaire during one (randomly) assigned day. These surveys consist of at least two main parts: one part queries the background information of the participant, and the other queries his/her contact behaviour.

The background information contains the following elements: the age and gender of the participant together with the household composition. This last question asks the number of household members as well as the age and gender of each household member. In the contact diary part, the participant enters the details of each person he/she met during the assigned day on one single line. The common aspects queried are age or age range (when the exact age was unknown to the participant), gender, location, duration, frequency and whether physical contact was involved. The common locations are home, work, school, leisure, transport and other. The common options for duration are less than 5 minutes, 5 to 15 minutes, 15 minutes to 1 hour, 1 to 4 hours or over 4 hours, except in the English Illness Survey (Section 2.3), in which the boundary of 15 minutes was changed to 10 minutes, leading to two deviating duration options: 5 to 10 minutes and 10 minutes to 1 hour. The common options for the frequency of contacts are daily or almost daily, about once or twice a week, about once or twice a month, less than once a month and first time. The Flemish Survey (Section 2.4) and the ICSOC Survey (Section 2.5) slightly adapted the options, except for the first and last one, to a few times a week, a few times a month and a few times a year or less. The definition of a contact deviates slightly between the surveys, but all exclude contacts via phone and internet.

2.2 The POLYMOD Survey anno 2005-2006

This section is based on the published work of others: "Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine, **5**, *381-391.".*

The POLYMOD Study (Mossong *et al.*, 2008), which aimed at improving public health policy in Europe through the modelling and economic evaluation of interventions for the control of infectious diseases, was the first large-scale study to quantify social contact patterns. It was set up in eight European countries: Belgium (BE), Germany (DE), Finland (FI), Great Britain (GB), Italy (IT), Luxembourg (LU), the Netherlands (NL) and Poland (PL). According to commonly agreed upon quota, sampling methodologies and diary designs, countries organised the survey at their level leading to slight variations by country. The researchers implemented a recruitment procedure to obtain a broadly representative sample of the population regarding geographical spread, age and gender for most of the countries. In Belgium, Italy and Luxembourg, the researchers recruited participants via random digit dialling using land line telephones; in Great Britain, Germany and Poland, recruitment was conducted through face-to-face interviews; and in the Netherlands and Finland, population registers were the basis for recruitment. In the Netherlands, the researchers linked the study to a larger serological study on vaccine-preventable infections. In Germany, the study was part of a multi-theme survey.

Given the expected critical role of children and adolescents in the spread of infectious agents, the researchers oversampled them. Only one member of a household was allowed to participate. When the participant was a young child, a parent or guardian filled in the diary on behalf of the child, while older children, whose parents gave parental consent, obtained a diary in adapted language to be filled in on their own. The researchers distributed the contact diaries from May 2005 until September 2006. Deviations by country existed due to languages and choices made, more details about which deviations are described in Mossong *et al.* (2008).

The questionnaires comprised the two parts discussed in Section 2.1, the background information and the contact information. The background questionnaire requested additional information about the employment status and the level of education completed. The contact diary part was supposed to be filled in on a randomly assigned day between 5 A.M. and 5 A.M. the following morning. A contact is defined as either skin-to-skin contact such as a kiss or a handshake (referred to as a physical contact) or a two-way conversation with three or more words in the physical presence of another person but no skin-to-skin contact (referred to as a non-physical contact). The contact dairy consisted of all parts described in the previous section; when the participant met a person multiple times throughout the day, the information had to be included in one line by summing the duration over the day and tick all relevant locations. For Belgium, Germany, Finland and the Netherlands, participants with professional contacts exceeding 10 (20 for BE) did not include those contacts in the contact diary, as requested.

To ease later comparisons, we briefly describe the highlights of the results, mainly for Belgium and Great Britain. In Mossong *et al.* (2008), diary weights are calculated for each country specifically based on the joint distribution of age and household size data from census data for the year 2000 published by Eurostat (http://epp.eurostat.ec.europa.eu/).

Mossong *et al.* (2008) report on a total of 7,290 diaries, 267 from the Netherlands and 1,328 from Germany. For Belgium, they retrieved 750 diaries, and for Great Britain 1,012. The average number of contacts in Belgium was 11.84 (sd 9.85) compared to 11.74 (sd 7.67) contacts on average in Great Britain. Mossong *et al.* (2008) estimated the social contact patterns via a negative binomial model on the aggregated number of contacts over 5-year age intervals using a tensor product spline as a smooth interaction term. They observe a similarity in the general contact pattern structure across the different countries with a strong diagonal element reflecting a like-with-like, a.k.a., assortative mixing, and also observe two parallel secondary diagonals starting at about 30-35 years representing children-to-adult mixing at a lower rate compared to the diagonal. Adults mixing with other adults resulted in a wider contact "plateau".

Mossong *et al.* (2008) ignore the question about the number of work-related contacts, leading to a possible reduction in contacts for Belgium, Germany, Finland and the Netherlands. Hens *et al.* (2009b) impute the omitted work contacts based on certain assumptions for Belgium and showed that the average number of contacts increased to 15.7 contacts. They furthermore report on the results from the full Belgian survey in which participants intentionally filled in the diary on two days (a weekday and a weekend day). The authors observe a decrease of about 19% in daily contact frequency for children and adolescents in holiday periods, and they find similar observations for adults on the weekend. Hens *et al.* (2009a) impute the excess work contacts for each of the selected countries and hence improved the comparability of the results between the countries. They compare contact surfaces between weekdays and the weekend as well as between holidays and regular periods and report a decrease from the week to the weekend between 12% to 26% in the basic reproduction number and a reduction of 10% to 45% from regular to holiday periods.

2.3 The English Illness Survey anno 2009-2010

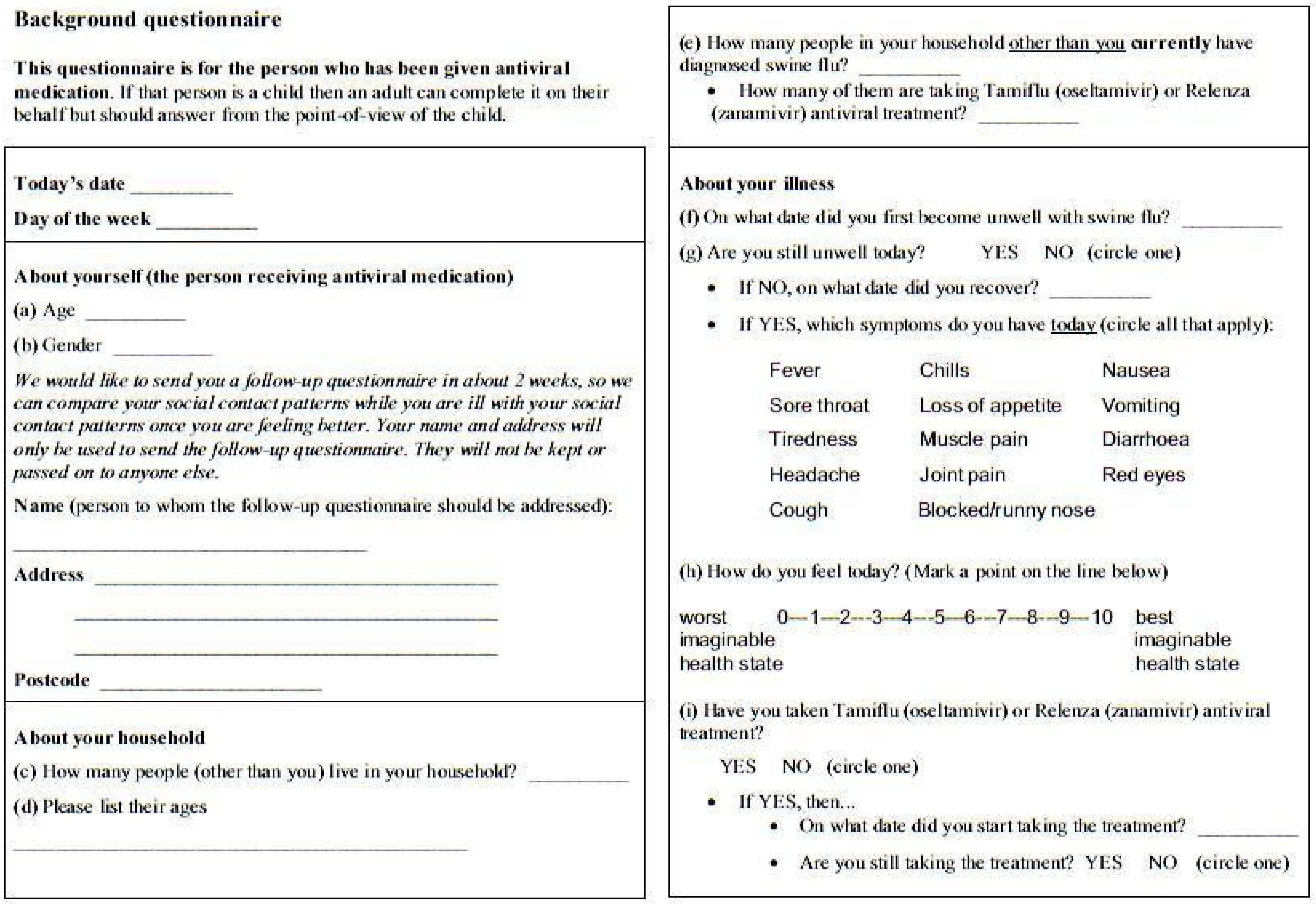
This section is based on the published work: "Van Kerckhove, K., Hens, N., Edmunds, J.W., Eames, K.T.D. (2013). The impact of illness on social networks: implications for transmission and control of influenza. American Journal of Epidemiology, **178***, 1655-1662.".*

This survey, referred to as the Illness Survey for the remainder of this thesis, was carried out during the A/H1N1pdm influenza epidemic in England in the autumn and winter of 2009-2010 and is described in detail in Eames *et al.* (2010). During the epidemic, participants were recruited into the study through questionnaire packs distributed with antiviral medication at 31 antiviral distribution centres throughout England. Participants were asked to complete the questionnaire while they were symptomatic with influenza-like illness (ILI). For children below 16 years of age, either the children could fill in the diary by themselves or a parent/guardian could fill it in on their behalf. Participants were asked to return the questionnaire in a prepaid envelope and to supply their address. Two weeks later (by which time participants were expected to have recovered from their illness), participants were sent a similar, follow-up questionnaire.

The questionnaires comprised the two main parts described in Section 2.1. The background survey included the common demographic factors but also queried information about the participant's illness (such as the onset date and severity) as shown on pages 17 & 18. The social contact diary was similar to the contact diary described in

Section 2.1, with slight deviations in the options for the duration of a contact as discussed in Section 2.1. Contacts were defined as either talking face-to-face or skin-to-skin contact (e.g., a handshake, a kiss, contact sports). As was the case in the POLYMOD Survey for Great Britain, the number of entries in the contact diary was limited to 33.

During the survey we aimed to obtain 2 contact diaries for each participant: one completed when the participant was symptomatic and one completed after he or she had recovered. We will refer to persons with symptoms of ILI as "ill" or "symptomatic" and to persons without symptoms of ILI as "healthy" or "asymptomatic".



Background questionnaire

This questionnaire is for the person who has been given antiviral medication. If that person is a child then an adult can complete it on their behalf but should answer from the point-of-view of the child.

Today's date __________

Day of the week __________

About yourself (the person receiving antiviral medication)

(a) Age __________

(b) Gender __________

We would like to send you a follow-up questionnaire in about 2 weeks, so we can compare your social contact patterns while you are ill with your social contact patterns once you are feeling better. Your name and address will only be used to send the follow-up questionnaire. They will not be kept or passed on to anyone else.

Name (person to whom the follow-up questionnaire should be addressed):

____

Address ____

____

____

Postcode ______________________

About your household

(c) How many people (other than you) live in your household? __________

(d) Please list their ages

____

(e) How many people in your household other than you **currently** have diagnosed swine flu? __________

- How many of them are taking Tamiflu (oseltamivir) or Relenza (zanamivir) antiviral treatment? __________

About your illness

(f) On what date did you first become unwell with swine flu? __________

(g) Are you still unwell today? YES NO (circle one)

- If NO, on what date did you recover? __________
- If YES, which symptoms do you have today (circle all that apply):

Fever	Chills	Nausea
Sore throat	Loss of appetite	Vomiting
Tiredness	Muscle pain	Diarrhoea
Headache	Joint pain	Red eyes
Cough	Blocked/runny nose	

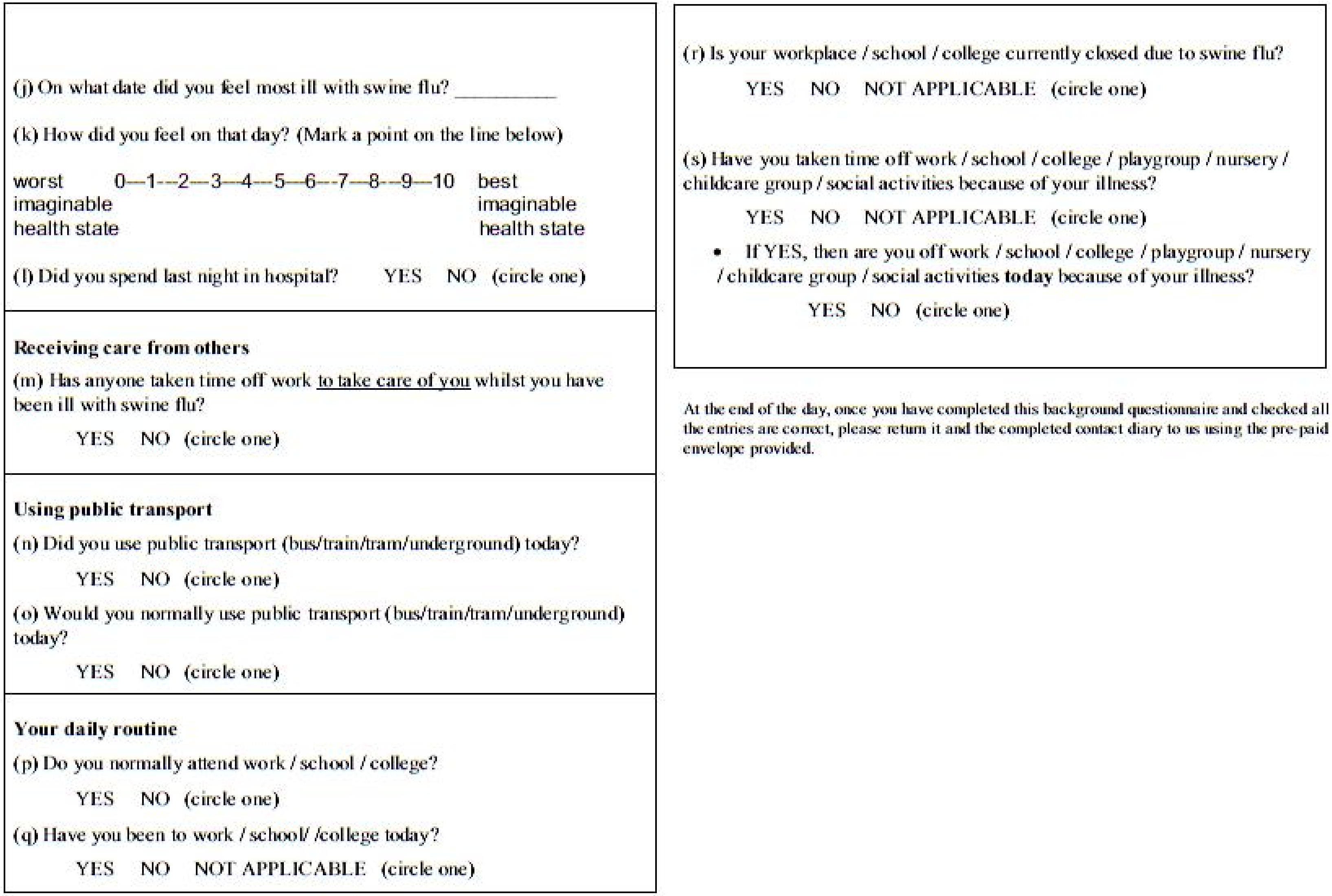
(h) How do you feel today? (Mark a point on the line below)

worst imaginable health state 0—1—2--3—4--5—6--7—8--9—10 best imaginable health state

(i) Have you taken Tamiflu (oseltamivir) or Relenza (zanamivir) antiviral treatment?

YES NO (circle one)

- If YES, then...
  - On what date did you start taking the treatment? __________
  - Are you still taking the treatment? YES NO (circle one)



(j) On what date did you feel most ill with swine flu? __________

(k) How did you feel on that day? (Mark a point on the line below)

worst imaginable health state 0—1---2—3—4—5—6---7—8--9—10 best imaginable health state

(l) Did you spend last night in hospital? YES NO (circle one)

Receiving care from others

(m) Has anyone taken time off work to take care of you whilst you have been ill with swine flu?

YES NO (circle one)

Using public transport

(n) Did you use public transport (bus/train/tram/underground) today?

YES NO (circle one)

(o) Would you normally use public transport (bus/train/tram/underground) today?

YES NO (circle one)

Your daily routine

(p) Do you normally attend work / school / college?

YES NO (circle one)

(q) Have you been to work / school/ /college today?

YES NO NOT APPLICABLE (circle one)

(r) Is your workplace / school / college currently closed due to swine flu?

YES NO NOT APPLICABLE (circle one)

(s) Have you taken time off work / school / college / playgroup / nursery / childcare group / social activities because of your illness?

YES NO NOT APPLICABLE (circle one)

- If YES, then are you off work / school / college / playgroup / nursery / childcare group / social activities **today** because of your illness?

  YES NO (circle one)

At the end of the day, once you have completed this background questionnaire and checked all the entries are correct, please return it and the completed contact diary to us using the pre-paid envelope provided.

A total of 301 participants returned a contact diary with their first questionnaire. A total of 174 participants returned a contact diary with their second questionnaire. Each of the diaries was double entered and compared to perform data cleaning. One participant did not report his/her age and hence does not contribute to the analysis. Not all participants were ill when they completed the first questionnaire, and not all had recovered by the time they completed their second questionnaire. A total of 140 participants returned two contact diaries, including one completed when they were ill and one completed when they were healthy; these 280 contact diaries make up the paired sample. A total of 283 participants were ill when they completed their first contact diary; these 283 contact diaries when participants were ill, make up the ill sample. A total of 159 participants had either recovered when they completed their second contact diary (n=151) or were not ill when they completed their first contact diary and did not complete a second contact diary (n=8); these 159 contact diaries, completed when participants were healthy, make up the healthy sample. A summary of the participants' demographic information is shown in Table 2.1.

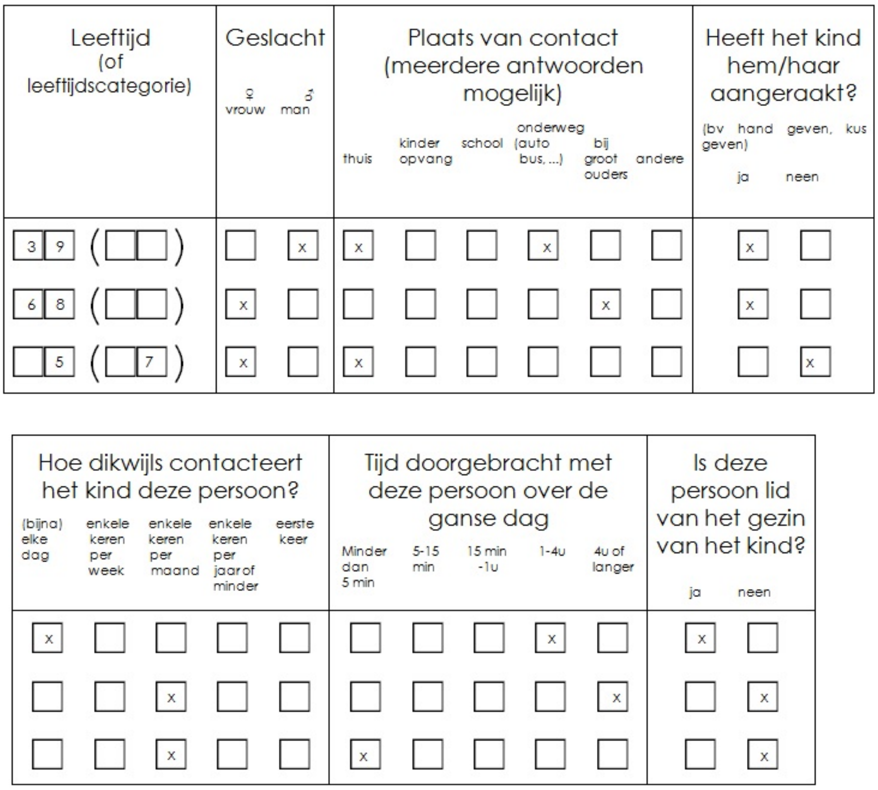
Table 2.1: Study population demographic summary based on a survey performed in 2009-2010 in England in comparison to the population demographics of the POLYMOD Survey and the United Kingdom (UK) population. Table from Van Kerckhove *et al.* (2013).

	Paired (140)	**Ill (284)**	**Healthy (159)**	**POLYMOD (1012)**	**UK population**
Female (%)	63	64	59	52	50.87
Age 0-3 (%)	6.43	6.36	5.66	7.51	4.91
Age 4-10 (%)	7.86	7.07	7.55	14.03	7.82
Age 11-21 (%)	5.71	7.77	7.55	21.44	13.93
Age 22-45 (%)	35	34.63	33.33	29.35	33.33
Age 46-64 (%)	37.14	34.63	37.11	22.04	23.74
Age $\geq$ 65 (%)	7.86	9.54	8.81	5.63	16.27
Mean hh size	1.8	1.90	1.76	3.31	2.4

2.4 The Flemish Contact Survey anno 2010-2011

In 2010-2011, Niel Hens and Philippe Beutels conducted two parallel social contact surveys in Flanders, Belgium. One of them focused on recruiting single individuals; thus, individuals from the same household were not allowed. The other survey focused on recruiting full households, focusing on households with at least one child aged 12 years or less. We refer to the first survey as the individual survey, and to the latter as the household survey. The recruitment for the individual survey took place between September 2010 and February 2011. Specific quotas for age, geographical spread and gender are used for sampling to obtain a representative sample. However, the researchers oversampled children and elderly to increase the information available concerning them. The researchers designed the survey to sufficiently sample participants during school holiday periods (1-7 Nov, 11 Nov, 27 Dec - 9 Jan) to make proper comparisons between holidays and regular periods. The recruitment for the household survey took place between October 2010 and July 2011. Here, sampling was performed using a stratified sampling approach to ensure representativeness concerning geographical spread, weekday-weekend distribution and the gender of the youngest child. All participants living more than 50% of the time in the household were defined as household members and were recruited to take part in the survey. For both surveys, the recruitment was conducted via random digit dialling to landlines and mobile phones. During this recruitment, participants gave a verbal consent before participation. When the participant gave a positive answer, (s)he received his (her) questionnaires via the postal service, including a pre-stamped envelope with which to send the questionnaires back. By not filling in the questionnaire and diary, and/or by not sending it back, participants could still refuse participation. The first page of the questionnaire explained that the answers of participants would be used anonymously for scientific purposes. Hence, by filling in and sending back the questionnaires, written consent was given. For dependent participants (e.g. children), the next of kin, caretakers or guardians could give verbal consent with implicit written confirmation on the participants' behalf.

Both surveys used the same definition for a contact, e.g. a two-way conversation at less than a 3-meter distance or a physical contact involving skin-to-skin touching (either with or without conversation). They both used similar questionnaires, and those were similar to the format discussed in Section 2.1. An adaptation of the questions posed in the contact diary was the request to indicate whether a contact was a household member or not. Furthermore, different types of booklets were made depending on the age of the participant: for children (0-12 years), for adults between 12-60 years of age and for the elderly (> 60 years) for the individual survey and two type of booklets for the household survey distinguishing between participants below 12 years of age and those older than 12 years. Depending on the age of the participant, different options for the locations were present in the contact diary; for example, the location of grandparents for children (see Figure 2.1) and family for the elderly, which are not present in either of the other diaries.



Leeftijd (of leeftijdscategorie)	Geslacht		Plaats van contact (meerdere antwoorden mogelijk)						Heeft het kind hem/haar aangeraakt? (bv hand geven, kus geven)	
	♀ vrouw	♂ man	thuis	kinder opvang	school	onderweg (auto bus, ...)	bij groot ouders	andere	ja	neen
3 9 (☐☐)	☐	☒	☒	☐	☐	☒	☐	☐	☒	☐
6 8 (☐☐)	☒	☐	☐	☐	☐	☐	☒	☐	☒	☐
☐ 5 (☐ 7)	☒	☐	☒	☐	☐	☐	☐	☐	☐	☒

Hoe dikwijls contacteert het kind deze persoon?					Tijd doorgebracht met deze persoon over de ganse dag					Is deze persoon lid van het gezin van het kind?	
(bijna) elke dag	enkele keren per week	enkele keren per maand	enkele keren per jaar of minder	eerste keer	Minder dan 5 min	5-15 min	15 min -1u	1-4u	4u of langer	ja	neen
☒	☐	☐	☐	☐	☐	☐	☐	☒	☐	☒	☐
☐	☐	☒	☐	☐	☐	☐	☐	☐	☒	☐	☒
☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☐	☒

Figure 2.1: Example of how to fill in the contact diary for children aged less than 12 years for the Flemish Survey. The first column of the top part indicates the age (or age category) followed by gender (female & male). Next is the location of contacts, which allows for multiple locations (left to right: home, kindergarten, school, travel, at grandparents, other), and the last column requests whether touching was involved (yes or no). On the bottom part, the first column asks for the frequency of contact (from left to right: (almost) every day, a few times a week, a few times a month, a few times a year or less, first time), the middle column is concerned with the duration of the contact over the entire day (left to right: less than 5 minutes, 5-15 minutes, 15 minutes - 1 hour, 1-4 hours, longer than 4 hours) and the final column indicates whether the person is part of the household of the child (yes or no).

This survey considers new aspects compared to the general questions in the background surveys (Section 2.1) and compared to the POLYMOD Survey. The general background questions were extended by information related to educational background, which can deviate by age. More extensively, two different health questionnaires, information about the time use of the participants and information related to animal ownership and touching were added. The educational background questions deviated for children as the educational level for the mother was requested in this type of booklet. For the elderly, the education question asked information about possible former employment. For children, the researchers also asked information about childcare, whereas, for the elderly, they requested information about living situations, e.g. at home or in an elderly home. The elderly were also requested to provide information related to their children and possible grandchildren. For the adults between 12 and 60 years of age, as was done in the Belgium survey in the POLYMOD Project, the researchers asked whether participants made many professional contacts and to estimate the amount. In case the estimate exceeded 20, the professional contacts were to be left out from the contact dairy. However, the questionnaire requested information about the age classes of these professional contacts. We explain the other three additional parts in more detail in the next sections.

Data for both surveys were single-entered in an electronic database and double-checked manually. The researchers did not collect physical samples as part of this study, and the ethical committee of the Antwerp University Hospital approved the study protocol.

2.4.1 Health questionnaires

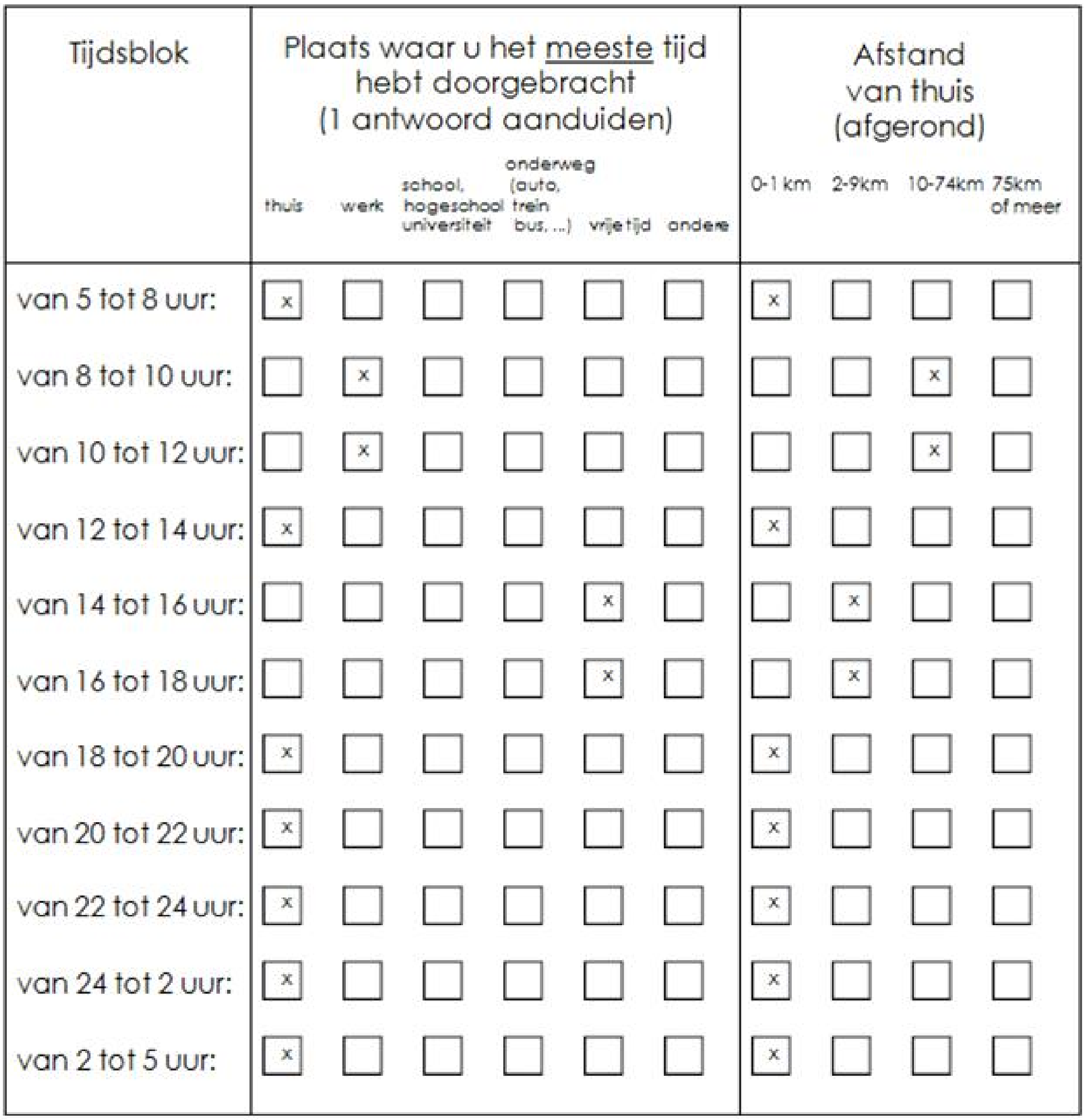
Each of the different types of diaries (by age) contained the standard Euroqol questionnaire (EQ-5D-3L) and a Visual Analogue Scale (VAS). The first part considers different questions related to the health of individuals. The questions are grouped by themes such as mobility, self-care, daily activities, pain, complaints, fear and depression. Participants indicated whether they suffered from any problem related to the topic or whether they had mild issues or severe issues with them. The Visual Analogue Scale is similar to a thermometer in which 0 stands for the worst imaginable health condition and 100 represents the best imaginable health status. Bilcke *et al.* (2017) used these data to obtain health-related quality-of-life (HRQoL) values for a broader age range as compared to the most commonly used values. Using different aspects of the background questionnaire, they showed that the HRQoL values largely depend on age and experience with severe disease. Furthermore, children have the highest probability to be in perfect health. Proxies for 0-2-year-old children who were not in perfect health report EQ-5D and VAS scores as low as those of the elderly. The authors found that the following factors associated with health-related quality-of-life: smoking behaviour, educational attainment, pet ownership, working or having worked in health care, and potentially household size and elderly living on their own (yes/no). They did not find a significant association with gender, living in a single-parent home, the educational attainment of mothers, alcohol consumption for elderly, having (grand-)children or the frequency of seeing them. Bilcke *et al.* (2017) highlight that estimating the HRQoL for children and developing statistical tools capturing particular features of these measures are important areas for further research.

2.4.2 Animal ownership and touching

Regarding animal ownership, participants reported if their household owned one or more living animals, and in case they did, they indicated what animals they owned. The options included cat, dog, horse, chicken, turkey, sheep, pigeon, pig, cow, fish or other, which could be specified. Multiple options were possible. The survey furthermore asked whether the participant touched one or more living animals on the assigned day, regardless of whether they owned the animal. When this was the case, they received the same options as those for the first question. Similar to the animal ownership question, multiple answers were allowed. After data collection, for clarity of presentation, the animals are grouped into four classes: pets (cat, dog, fish), livestock (horse, sheep, pig, cow), poultry (chicken, turkey, pigeon) and “other”.

2.4.3 Time use questionnaire

In the time use survey (Figure 2.2), participants filled in the one location at which they spent most of their time and the distance from home for this location (0-1 km, 2-9 km, 10-74 km or 75 km or more) for pre-specified time blocks. The time blocks were of 1-hour length, except in the morning (2-5h, 5-8h) and in the evening (between 20h and 2h, 2 hours time blocks). The locations were the same as the possible locations for the contacts and thus deviated depending on the age of the participant.



Tijdsblok	Plaats waar u het meeste tijd hebt doorgebracht (1 antwoord aanduiden)						Afstand van thuis (afgerond)			
	thuis	werk	school, hogeschool universiteit	onderweg (auto, trein bus, ...)	vrije tijd	andere	0-1 km	2-9km	10-74km	75km of meer
van 5 tot 8 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 8 tot 10 uur:	☐	☒	☐	☐	☐	☐	☐	☐	☒	☐
van 10 tot 12 uur:	☐	☒	☐	☐	☐	☐	☐	☐	☒	☐
van 12 tot 14 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 14 tot 16 uur:	☐	☐	☐	☐	☒	☐	☐	☒	☐	☐
van 16 tot 18 uur:	☐	☐	☐	☐	☒	☐	☐	☒	☐	☐
van 18 tot 20 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 20 tot 22 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 22 tot 24 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 24 tot 2 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐
van 2 tot 5 uur:	☒	☐	☐	☐	☐	☐	☒	☐	☐	☐

Figure 2.2: Example of how to fill in the time use data for participants aged between 12 and 60 years for the Flemish Survey. The first column indicates the different time blocks, the second column refers to the location at which a participant spent most of his time (from left to right: home, work, school, travel, leisure, other) and the last column requests the distance from home with categories 0-1 km, 2-9 km, 10-74 km or 75 km or more.

2.5 The Pilot Study ICSOC anno 2016-2017

Between January 2017 and April 2017, we conducted a pilot study in Belgium, more precisely, in the provinces of Limburg and Antwerpen. The aim of the pilot study was to quantify changes in social contact behaviour as a result of acute infection, such as seasonal influenza, and as a result of chronic diseases. A previous study, the Illness Survey, as described in Section 2.3, already aimed to quantify changes as a result of 2009 A/H1N1pdm influenza. Notwithstanding the importance of this first study, due to the recruitment method, no representative sample of the population was obtained. Since the English Illness Survey took place during the influenza A/H1N1pdm epidemic, it does not describe changes as a result of seasonal influenza.

Inspired by the above aspects, our aim is to set up a large-scale survey during a ‘regular’ influenza season to quantify these changes using a representative sample. Furthermore, given the lack of information on the changes induced in social contact behaviour as a result of chronic conditions, we aim to quantify these as well. As influenza results in complications for people with chronic infections, this information can be valuable (European Centre for Disease Prevention and Control, 2017a). Due to these novel aspects, we conducted a pilot study to test the feasibility of the design as well as that of the diaries and to gain insight into possible effects.

2.5.1 Study design

We contacted general practitioners in our networks to distribute contact diary questionnaires to a selection of their patients. The general practitioners selected based upon the patient having an acute (influenza-like) illness or a chronic disease combined with age-stratified quotas to represent the population age distribution. Furthermore, they aimed to balance the distribution between genders. We sought to recruit ten general practitioners and set the quota to recruit 25 patients with an ILI and 13 patients with a chronic condition per general practitioner or practice, thus aiming for a total of 250 patients with an ILI and 130 patients with a chronic condition. The general practitioners asked their patients whether they would like to participate and gave them a questionnaire, an informed consent form and a prepaid envelope. The informed consent form (which was also available in the diary as the first two pages) contained information related to the aim of the study and the rights of the participant. Participants with an ILI could write down their postal address at the end of the questionnaire and received a new questionnaire - supposedly - two weeks later, assuming they were recovered. Participants with only a chronic condition were asked to fill in the questionnaire and send it back. They received no follow-up questionnaire unless the participant presented with an ILI as well as a chronic condition.

The questionnaires were based on the questionnaires as described in Section 2.1 but in a format closely resembling the questionnaires from the Flemish Survey (Section 2.4), albeit slightly reduced. We removed the questions concerning time use and animal ownership and touching but kept the health score questionnaire. Furthermore, we translated and added the health-related questions as used in the Illness Survey (Section 2.3). One additional aspect included here was a question requesting a summary of the social contacts from the two days preceding the visit to the doctor. This question may provide insight into changes in the number of social encounters during an illness in a different manner. The follow-up questionnaires were to be filled in on the same day of the week as the first questionnaire and avoided holiday periods. The first aspect cancels the possible weekday effect, which is of importance given the small study, and the latter aspect avoided the additional holiday effect but resulted in a longer period than two weeks between questionnaires. Both aspects allow us to focus solely on the changes as a result of infection.

The study was approved by the ethical committee of Hasselt University and was carried out in the context of a bachelor thesis in medicine. Four students in medicine contacted the potential general practitioners and supplied them with questionnaires. They governed contact with the general practitioners and performed a small survey at the end of the study to question the feasibility of the study from the perspectives of the general practitioners. The survey requested whether the general practitioner remembered to distribute booklets, the time they spent when asking a patient to participate, the amount of explanation given, the feasibility of the quota, the preparedness of patients to participate, the reasons for refusal, other aspects taken into account to select the patients (e.g., educational background, health status, and language), the willingness to join in a similar large-scale study and an open question for suggestions for improvement.

2.5.2 Data entry & collection

Data were double-entered in Qualtrics by the students. The Qualtrics sheets resembled the booklets, and the entered data were saved in a datasheet. Using the compare procedure from SAS and checking important aspects in the variables, such as the number of contacts, date of filling in, and the date of the visit to the doctor, we cleaned the data.

Of the ten recruited general practitioners eight actively distributed questionnaires. We retrieved a total of 99 booklets of which 21 from participants with a chronic condition and 78 booklets from participants with an ILI. Fifty-three booklets were first booklets and 25 follow-up questionnaires.

Driven by the different conditions under study, we create various subgroups for comparison, distinguishing between chronic conditions or not, an ILI or not and paired information or not. As a result, we have the following groups:

1. *Chronic only* ($n = 23$) - which contains booklets from participants whom solemnly suffer from a chronic condition, including 4 from the paired sample;
2. *ILI only* ($n = 31$) - which contains booklets from all participants whom only suffer from an ILI, including 10 from the paired sample;
3. *Chronic and ILI* ($n = 14$) - which contains booklets from the participants suffering from both a chronic condition and an ILI, including 4 from the paired sample;
4. *Healthy* ($n = 25$) - which contains all booklets from participants when they were healthy, which also includes the booklets from those participants who were healthy when they filled in the second questionnaire ($n = 10$);
5. *Paired sample* ($n = 14$) - which contains booklets from all participants who reported when feeling ill with an ILI and/or a chronic condition during the first time and were recovered from the ILI when they filled in the second time.

Note that whenever a participant returned two questionnaires with both of them during illness (health), only the first (second) contributes to the ILI, with or without a chronic condition, (healthy) group. Furthermore, five booklets are not used in the further analysis as they are incomplete or inconsistent, leading to results based on 82 booklets. Notwithstanding the aim to use similar diaries as the Flemish Survey anno 2010-2011, a discrepancy occurred for the location of adults which in this study does not include the leisure option. We should keep this difference in mind for future surveys.

Chapter 3

The Interplay Between Illness and Social Contact Behaviour

We start this chapter by describing the impact of illness on social contact behaviour based on the English Illness Survey (Section 2.3). Based on these results, we conducted two follow-up studies. In the first one, we developed a mathematical model allowing contact behaviour to change during the course of the disease from asymptomatic to symptomatic. Furthermore we also take the (a)symptomatic status of the infector into account. In the second study, we describe the results of a pilot study as a follow-up to the initial English Illness Survey.

3.1 The impact of illness on social contact behaviour

This section is based on the published work: "Van Kerckhove, K., Hens, N., Edmunds, J.W., Eames, K.T.D. (2013). The impact of illness on social networks: implications for transmission and control of influenza. American Journal of Epidemiology, **178***, 1655-1662.".*

3.1.1 Introduction

Knowledge of social networks is vital when seeking to understand and predict the spread of infectious diseases in human populations. In Chapter 1 we eluded on the various approaches to gain knowledge in social contact networks. In Section 2.2 we elaborated on results based on the POLYMOD Survey related to temporal aspects. These temporal variations in social contact behaviour have also been observed by others (Eames *et al.*, 2010, 2011, 2012). Epidemiologic models can include these changes in contact patterns appropriately given the increasing availability of data describing them. However, there remained a critical gap in the knowledge in 2010 - namely, the impact of illness on social contact patterns. If, as might be expected, people modify their behaviour when they are ill, the value to mathematical models of social mixing data collected predominantly from healthy persons is questionable. If illness results in people's taking time off from work or school, avoiding social gatherings or changing their social behaviour in other ways, the behaviour of ill persons would be poorly described by commonly collected social contact data. If asymptomatic infections are common or if transmission takes place before symptoms appear, then possible changes in social behaviour are less important.

To fill in the critical gap about the impact of illness on social contact patterns, the Illness Study (Section 2.3) aimed at collecting contact behaviour of persons when symptomatic and when asymptomatic. We, in Van Kerckhove *et al.* (2013), analyse these data and demonstrate that not only does the number of social contacts change when a person is ill, so does the distribution of contacts across age groups. We show that this has a major impact on the basic reproduction number (R_0) (Diekmann *et al.*, 1990) and the expected age distribution of cases in the population. It is suspected that the 2009 A/H1N1pdm epidemic resulted in a significant number of infections that displayed no symptoms or mild symptoms. Both serological data and modelling work indicated that patients with ILI who sought medical attention were a small fraction of the total number of persons with infections in the United Kingdom (Eames *et al.*, 2012; Baguelin *et al.*, 2011). By comparing the observed age distribution of cases with that predicted by the measured contact patterns of asymptomatic and symptomatic persons, we can address one of the more intractable problems of influenza epidemiology: What contribution do asymptomatic persons make to overall influenza transmission? The findings have important implications for public health since the proportion of transmission resulting from symptomatic persons determines how effective treatment with antiviral agents or home isolation will be in limiting the spread of the disease.

3.1.2 Methods

3.1.2.1 Comparisons

We compare social mixing patterns reported by ill participants with those reported by healthy participants, investigating the number of contacts and the distribution of contacts between age groups. We base the age categorization for children and young adults on the United Kingdom schooling system; 3 age classes are chosen to allow for likely family, household, and workplace arrangements at different ages. The resulting age classes were ≤3, 4-10, 11-21, 22-45, 46-64, and ≥65 years. We make several comparisons: First, we use the paired contact diaries to assess the impact of ILI symptoms on social mixing behaviour at the individual level; and second, we compare unpaired contact diaries (completed by ill or healthy participants) with POLYMOD data from Great Britain as a historical control (Section 2.2).

The contact diaries used in this study had space for encounters with 33 different people to be recorded. Ten participants also reported additional encounters. In the results we present, we cap the number of encounters at 33. Sensitivity analysis showed no substantial deviations when using all reported contacts.

The contact diaries used in this study give a number of different ways of defining an epidemiologically meaningful encounter. Here in common with previous studies (Goeyvaerts *et al.*, 2010; Ogunjimi *et al.*, 2009), we focus on 2 possibilities: First, we analyze all encounters recorded in the diary; and second, we analyze only those encounters that included physical (skin-to-skin) contact (see Table 3.A.1 and Figures 3.A.1 and 3.A.2).

3.1.2.2 Estimating transmission rates

The social contact rate matrices are estimated as described in Section 1.2.3.2. The survey did not involve representative sampling by age group nor day of the week. Hence the contact diary weights account for those two factors as explained in Section 1.2.3.1 whenever the age of the participant and the day on which contacts were recorded are known. The population data are data from UK (Eurostat Population Statistics 2008).

Since people's mixing patterns when ill (symptomatic) are different from those present when they are healthy (asymptomatic), we can calculate matrices $\mathbf{M^S}$ and $\mathbf{G^S}$ based on behaviour when symptomatic and matrices $\mathbf{M^A}$ and $\mathbf{G^A}$ based on behaviour when asymptomatic. However, note that in this case the calculation of $\mathbf{G}$ is slightly deviating from the method in Section 1.2.3.3 and is $g_{ij} = m_{ji}s_iDq$ as the infection is expected to spread starting from the participant who was ill.

3.1.2.3 Comparing contact matrices

While it is possible to display contact patterns graphically using the social contact matrix $\mathbf{M}$, it is not immediately clear how best to quantify epidemiologically relevant differences between contact matrices.

Here we compare contact patterns using the basic reproduction number, R_0, defined as in Section 1.2.3.3. Following the method of (Hens *et al.*, 2009a), we consider the ratio of estimates of R_0 derived from different mixing matrices $\mathbf{G^S}$ and $\mathbf{G^A}$, which is equivalent to comparing the dominant eigenvalues of $\mathbf{M^S}$ and $\mathbf{M^A}$. Under the null hypothesis of equal contact matrices and assuming q to be constant, this ratio is expected to equal 1.

For each comparison, we assess the significance of any deviation from the null hypothesis by calculating 95% confidence intervals based on a paired nonparametric bootstrap (Efron and Tibshirani, 1993; Section 1.2.4.2). We also calculate the expected age distribution of incidence in the population during the exponential phase, as given by the leading eigenvector of the next generation matrix (Diekmann *et al.* (1990); Vynnycky and White (2010); Section 1.2.3.3).

3.1.2.4 Contribution of symptomatic cases to influenza transmission

We seek to estimate the proportion of cases generated by asymptomatic individuals, η. To do this, we assume that a proportion ϕ of infectious individuals are asymptomatic (i.e., have mixing patterns described by $\mathbf{M^A}$), with the remaining proportion $(1-\phi)$ being symptomatic (i.e., having mixing patterns described by $\mathbf{M^S}$). We assume ϕ to be the same for all age groups. To obtain an estimate for ϕ, we consider $\mathbf{v^{mixed}}$, the leading eigenvector of the adapted next generation operator $\mathbf{G^{mixed}} = \phi q_A\mathbf{G^A} + (1-\phi)q_S\mathbf{G^S}$. Here, q_A and q_S represent the transmissibility for asymptomatic and symptomatic persons, respectively. Multiplying $\mathbf{G^{mixed}}$ by a constant does not change the (normalised) eigenvector; hence $\mathbf{v^{mixed}}$ is also the leading eigenvector of

$$\frac{\mathbf{G^{mixed}}}{\phi q_A} = \mathbf{G^A} + \left(\frac{1-\phi}{\phi}\right) q_R\mathbf{G^S}, \tag{3.1}$$

where $q_R = (q_S/q_A) \in (0,\infty)$. We define $\tilde{\phi} = [(1-\phi)/\phi]q_R \in [0,\infty)$, from which ϕ can be deduced if q_R is known. The susceptibility of the population (s_i) in the calculation of $\mathbf{G^A}$ and $\mathbf{G^S}$ is obtained from a serological study conducted using samples from England taken during the year before the pandemic (Miller *et al.*, 2010). We compare the predicted distribution of incidence across age groups using $\mathbf{G^{mixed}}$ to the general practitioners' consultation data from England and Wales from the early part of the epidemic (weeks 23-28 in 2009) (Public Health England, 2010). During this phase of the epidemic, the number of cases was growing approximately exponentially; hence, the expected distribution of incidence is given by $\mathbf{v^{mixed}}$, the leading eigenvector of $\mathbf{G^{mixed}}$. This analysis uses slightly different age groups (≤4, 5-14, 15-44, 45-64, and ≥65 years) to match those in the general practitioners' data. The optimal value of $\tilde{\phi}$ is estimated as the value leading to the smallest least-squares difference between the observed and estimated age distributions of cases, using all contacts, physical contacts, or contacts of a duration longer than 1 hour (these last 2 categories being included for consideration because more intimate and longer-duration contacts have previously been shown to be closely related to transmission patterns (Goeyvaerts *et al.*, 2010; Ogunjimi *et al.*, 2009; Melegaro *et al.*, 2011)). We note that the best-fitting value of $\tilde{\phi}$ is independent of the value of q_R. The proportion of infections generated by asymptomatic persons is given by

$$\eta = \frac{\sum \mathbf{G^A v^{mixed}}}{\sum \mathbf{G^A v^{mixed}} + \tilde{\phi} \sum \mathbf{G^S v^{mixed}}}. \tag{3.2}$$

The estimate of η is independent of q_R; that is, the fraction of infections caused by asymptomatic persons is independent of the relative transmissibility in symptomatic and asymptomatic persons. Having calculated the best-fitting value for $\tilde{\phi}$, then if we are able to independently measure ϕ we can calculate q_R and vice versa.

To compare the change in the number of contacts in different social settings between the ill and healthy conditions, one should take into account the individual-specific heterogeneity induced by the paired information. Using generalised linear mixed models with a negative binomial distribution (Section 1.2.4.1; Molenberghs and Verbeke (2005)), we investigate the effect of the social setting as well as of the duration of contacts.

3.1.3 Results

The general information we use, such as symptoms, the severity of illness, and a preliminary analysis of the number of contacts (but not the age distribution of contacts), has been presented previously (Eames *et al.*, 2010). They observed that when healthy, people reported an average of 14.9 (95% CI: $[11.2, 20.1]$) contacts each day and, when ill, they reported 3.8 (95% CI: $[3.2, 4.6]$) contacts daily.

Here, we observe that ill people altered whom they contacted (Figure 3.1). As expected, a large proportion of encounters reported by participants when they were healthy were with people of a similar age as themselves, with particularly high numbers of contacts occurring among schoolchildren and young adults. Indeed, the contact patterns of participants when healthy are very similar to those observed in the POLYMOD Study (see Table 3.A.1 and Figures 3.A.1 and 3.A.2). In contrast, the strong like-with-like pattern of social mixing is not evident in encounters reported when participants were unwell. We note that the contact patterns displayed in Figure 3.1 may be asymmetric, since, for example, people contacted by ill participants may have been either ill or healthy. As such this is a condition of the participant as discussed in Section 1.2.3.2.

That illness results in a particular reduction in the number of contacts between people of similar ages can be understood when considering the social setting of encounters (Figure 3.2, Table 3.1). When ill, participants reported having far fewer contacts in the work/school and leisure settings - both settings where like-with-like mixing between age groups would be expected to take place. In contrast, illness has little discernible effect on the number of contacts made at home. Although it is tempting to assume that home contacts provide an appropriate proxy for contacts made when ill, we did not find this to be the case: Comparing the reported contacts made at home with the reported contacts made during illness, we found an R_0 ratio significantly different from 1 (R_0 ratio=1.42, 95% CI: [1.01, 2.53]). Turning to the duration of encounters, we see in Figure 3.2 that in most settings there was a tendency for participants to report briefer encounters when they were ill, albeit non-significant.

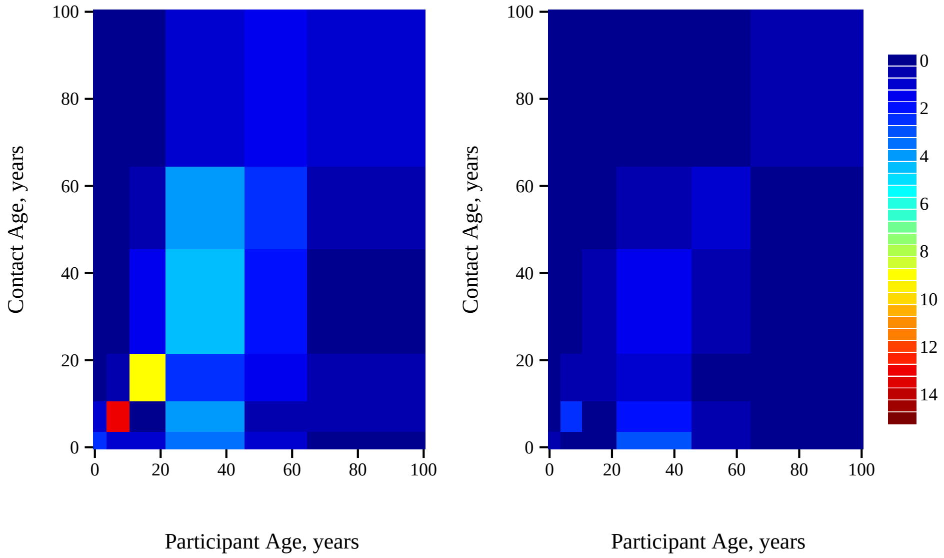


Figure 3.1: Daily numbers of social contacts made between age groups during the 2009 A/H1N1pdm influenza epidemic, England, 2009-2010. Contact patterns were reported by study participants who completed a contact diary on 2 occasions, once when they were asymptomatic (left, n=140) and once when they were symptomatic (right, n=140). Colors represent the mean number of encounters reported by each participant with contacts in each age group. There was less like-with-like mixing reported by symptomatic participants. Figure from Van Kerckhove *et al.* (2013).

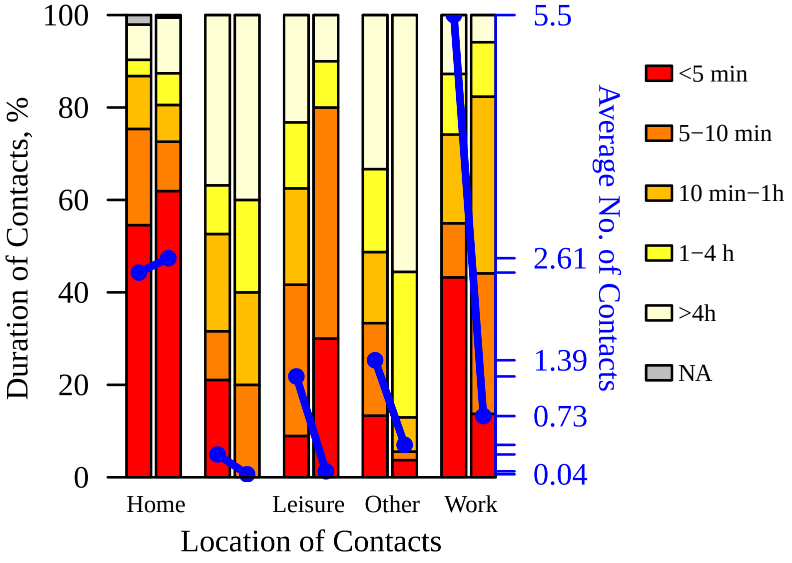


Figure 3.2: Numbers and durations of encounters in different social settings during the 2009 A/H1N1pdm influenza epidemic for asymptomatic and symptomatic study participants in the paired data set (n=140), England, 2009-2010. The blue points show the average number of encounters in each setting (right-hand axis), and the colors show the distribution of encounter durations (left-hand axis). For each pair of bars, the left-hand bar represents contacts made when asymptomatic, and the right-hand bar represents contacts made when symptomatic. In every location except home, a drop in the average number of contacts can be observed. NA, not applicable. Figure from Van Kerckhove *et al.* (2013).

Table 3.1: Relative change in the number of contacts (RNC) when ill for different social settings based on a generalised linear mixed model (n=140). Results are based on a study performed in 2009-2010 in England. At all locations, except home, a significant decrease in the number of contacts is observed when ill. Table from Van Kerckhove *et al.* (2013).[a,b]

Setting	RNC	95% CI
Home	1.01	0.68, 1.34
Work	0.09	0.04, 0.14
Travel	0.13	0.01, 0.26
Leisure	0.06	0.01, 0.11
Other	0.25	0.11, 0.39

[a] Results were based on a generalised linear mixed model

[b] At all locations except home, a significant decrease in the number of contacts was observed when people were ill.

The altered patterns of social encounters when people are symptomatic result in substantial reductions in the basic reproduction number, R_0. When calculated using the symptomatic contact matrix, R_0 is only one-quarter of the value it takes when calculated using the asymptomatic contact matrix (Table 3.2). These changes also affect the expected distribution of infection during the early stages of an outbreak. As Figure 3.3 shows, the contact patterns of healthy participants would result in a concentration of incidence among children of primary-school age. However, the contact patterns of ill participants led to a greater concentration of incidence in young adults. This effect emerged because children reported far fewer contacts with other children when symptomatic. Because like-with-like mixing was less evident in ill persons, mixing among adults and between adults and children played a greater role in transmission.

Under the assumption that persons with an asymptomatic infection do not change their social contact behaviour, we can use our data to estimate the fraction of cases infected by asymptomatic persons, by comparing the age distribution of incidence from general practitioners' consultation data (points in Figure 3.4 (both panels)) with that predicted using the ill and healthy contact data (lines in Figure 3.4 (right panel)). The left panel of Figure 3.4 shows this estimation method for different ϕ values (assuming q_R=1): If ϕ=0, then only symptomatic persons transmit infection, and the age distribution of cases during the exponential phase of the epidemic would be given by the solid line. If ϕ=1, then only asymptomatic persons transmit infection, and the age distribution of cases would be given by the dashed line. We find (see Figure 3.4 (right panel) and Table 3.3) that the observed distribution of ILI cases during the early stage of the epidemic is consistent with the majority of transmission's being due to persons with overt symptoms of ILI; we found that approximately two-thirds of the infections (Table 3.3, "all contacts") were caused by symptomatic persons. The best fit to the data was found when using patterns of mixing derived from all reported contacts; restricting encounters to those involving some physical contact or those lasting longer than an hour resulted in a poorer fit to the incidence data (Table 3.3).

Using the contact matrices describing all contacts, the best fitting value for $\tilde{\phi}$ was 5.86. It has been suggested that approximately two-thirds of infections involved any symptoms of ILI (Carrat *et al.*, 2008), which leads to $q_R \approx 2.93$, implying that persons with symptomatic infections are about 3 times as infectious per social contact as persons with asymptomatic infections. If it were the case that one-third of infections were symptomatic (e.g., considering only infection that causes fever (Carrat *et al.*, 2008), then q_R would be estimated to be 11.72. If half of the infections were symptomatic, then q_R would be estimated to be 5.86.

3.1.4 Discussion & Conclusion

We have quantified the marked effect that being infected with ILI has on people's social mixing behaviour. People report significantly fewer social encounters when they have influenza symptoms. Not only does the number of encounters fall, but the patterns of contact change; people with ILI have fewer contacts at work/school and during leisure activities, which results in fewer contacts in their own age group. This can be understood as a result of people taking time off from work/school and avoiding social activities when ill.

We have found that patterns of incidence during the early stages of the influenza A/H1N1pdm pandemic in England and Wales suggest that symptomatic persons generated the majority of cases. Therefore, there is scope to focus intervention resources on targeting symptomatic individuals, encouraging efficient treatment, and recommending time off from work and social distancing for persons with symptoms. It appears that during the A/H1N1pdm epidemic, symptomatic individuals were, in general, sufficiently unwell that they moderated their social behaviour; had a substantial fraction of them continued to engage in normal social activities, there would have been a large increase in transmission.

R_0 is a common measure of viral fitness (Alizon *et al.*, 2009). Using this measure, we have demonstrated that there is a considerable fitness cost associated with illness through the disruption of social networks. Changes in behaviour, in our sample, were correlated with severity of illness (measured by the number of symptoms reported) (Eames *et al.*, 2010). However, the presence of symptoms is also correlated with higher transmissibility: If, as seems plausible, between one-third and two-thirds of A/H1N1pdm infections were symptomatic (Carrat *et al.*, 2008), then our calculations suggest that symptomatic infections were 3-12 times as infectious as asymptomatic infections. Therefore, there appears to be a trade-off between infectiousness and opportunities for transmission. Symptomatic persons are more infectious per contact but make fewer contacts than asymptomatic persons.

Accessing symptomatic persons is challenging and, as with any such study, it is possible that the participants in this study were not representative of all people with ILI. However, the fact that patterns of incidence during the early stages of the epidemic were so well predicted by the mixing patterns reported by participants suggests that we succeeded in capturing true epidemiologically relevant behaviour. Furthermore, when healthy, the participants had mixing patterns very similar to those found in other social contact studies, suggesting that the sample was reasonably representative regarding normal social mixing behaviour. We note that the social contact data used here were collected only in England, while data from other sources have covered all of Great Britain or England and Wales; however, in all cases, the majority of the information came from England, so we are confident that the differences in sampling locations did not affect our conclusions.

The potential limitations of the diary-based approach are well known and touched upon in Chapter 1 (Read *et al.*, 2012). Collecting data from young children is difficult since we have to use proxies and people (of any age) may not record all of their contacts. Electronic methods (Cattuto *et al.*, 2010) can reduce these problems, but such methods require very high rates of participation in order to record most contacts (Read *et al.*, 2012) and are

Table 3.2: Basic reproduction numbers for different types of social contacts made during the 2009 A/H1N1pdm influenza epidemic, England, 2009-2010[a,b]. Table from Van Kerckhove *et al.* (2013).[a]

Comparison	**R_0 Ratio**	**95% CI**[b]
Results based on all social encounters		
Paired ill vs. healthy (n= 140)	0.29	0.15, 0.56
Healthy (n= 159) vs. POLYMOD GB (n= 1,012)	1.07	0.73, 1.88
Ill (n= 283) vs. POLYMOD GB (n= 1,012)	0.34	0.28, 0.43
Results based on skin-to-skin encounters		
Paired ill vs. healthy (n= 140)	0.45	0.22, 0.60
Healthy (n= 159) vs. POLYMOD GB (n= 1,012)	0.83	0.57, 1.19
Ill (n= 283) vs. POLYMOD GB (n= 1,012)	0.36	0.31, 0.49

[a] Comparisons of ill participants with healthy participants and ill participants with POLYMOD GB participants showed a significant difference in the R_0 ratio. Healthy participants and POLYMOD GB participants did not differ significantly.

[b] Bootstrap-based 95% confidence interval.

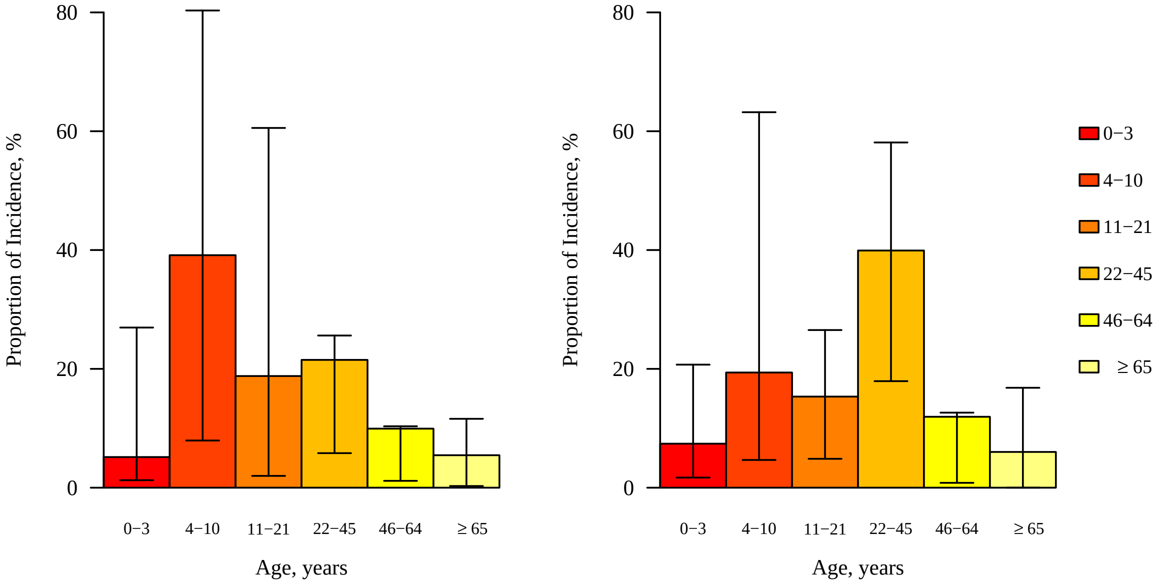


Figure 3.3: Age distribution of cases in the early stages of the 2009 A/H1N1pdm influenza epidemic as predicted by asymptomatic contact patterns (left, n=140) and symptomatic contact patterns (right, n=140) in the paired data set, assuming full susceptibility in the population, England, 2009-2010. The indicated 95% confidence intervals (T-shaped bars) were obtained using a nonparametric bootstrap. Because of the skewness of the underlying distributions, the mean values are not in the middle of the intervals. Figure from Van Kerckhove *et al.* (2013).

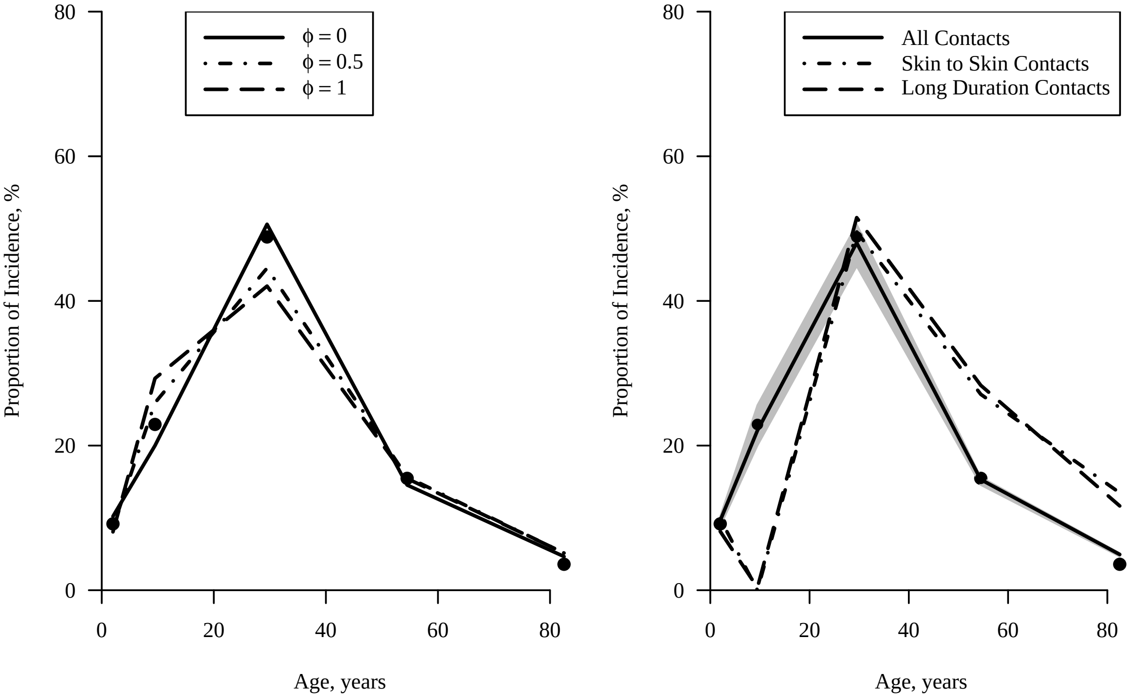


Figure 3.4: Theory (left) and fit (right) of a model to the observed age distribution from general practitioners' consultation data from England and Wales for the early phase of the 2009 A/H1N1pdm influenza pandemic (black dots), England, 2009-2010. The left-hand panel shows different ϕ values (assuming q_R=1, for illustration) leading to different age distributions. The right-hand panel shows the fit obtained using the best-fitting value of $\tilde{\phi}$ from all capped contacts (the gray area surrounding it indicates the 95% confidence interval, bootstrapping for $\tilde{\phi}$), from skin-to-skin contacts, and from long-duration contacts ($>$ 1 hour). Figure from Van Kerckhove *et al.* (2013).

Table 3.3: Estimated proportions of influenza infections caused by asymptomatic people during the 2009 A/H1N1pdm influenza epidemic, obtained using different contact matrices, England, 2009-2010. Table from Van Kerckhove *et al.* (2013).

Contact Matrix	Fraction of Cases Caused by Asymptomatic Infections	95% CI	Sum of Squared Residuals[a]
All contacts	0.34	3.61×10^{-4}, 0.77	0.0003
Skin-to-skin contacts	2.35×10^{-4}	6.00×10^{-6}, 2.91×10^{-1}	0.0187
Long-duration contacts ($>$ 1 hour)	0.18	1.06×10^{-3}, 0.35	0.0773

[a] A lower value of the sum of squared residuals indicates a better fit to the data.

therefore not suitable for a general population survey. Different biases could appear when participants are unwell, which could have affected our results. For example, diaries could be filled in better when people are ill since fewer contacts are easier to remember and record or they could be filled in less well, since illness may make participation more burdensome. We were unable to quantify the magnitude of such effects.

In this analysis, the duration of infectiousness was assumed to be equal for symptomatic and asymptomatic persons. The influences of this assumption have not yet been investigated, since they were beyond the scope of this initial paper, but will be the focus of further work. We assumed that infected persons are behaviourally "asymptomatic" or "symptomatic" throughout their infectious period. Alternatively, it is possible that people with symptoms change their social mixing behaviour some time after symptom onset or that people are infectious for some time before symptoms appear. Exploring the impact of these possibilities is done in Follow-up study I (Section 3.2); in the models presented here, in the absence of adequate data with which to link infectiousness, symptom onset, and behaviour change, we have taken the simpler approach.

To our knowledge, this study was the first to assess the individual-level impact of infection on social contact patterns and the consequences of altered social encounters on epidemic transmission. Even though the 2009 pandemic influenza strain was relatively mild, the change in contact patterns induced was large and epidemiologically significant. Epidemic models that fail to take these changes into account are likely to be inaccurate - even more so if disease severity is high. Symptomatic persons appear to have been responsible for most of the infection transmission, and efforts to reduce influenza transmission should target them. Indeed, should this result hold for other strains of influenza, identifying, treating, and isolating symptomatic individuals should be the focus of public health efforts in order to prevent transmission to others in the community. In Follow-up study II (Section 3.3) we present a pilot study with one of the aims to quantify the effect of other strains as well.

3.2 Follow-up study I: Structural differences in mixing behaviour informing the role of asymptomatic infection and testing symptom heritability

This section is based on the published work: "Santermans, E., Van Kerckhove, K., Azmon, A., Edmunds, W.J., Beutels, P., Faes, C., Hens, N. (2017). Structural differences in mixing behaviour informing the role of asymptomatic infection and testing symptom heritability. Mathematical Biosciences, **285***, 43-54.".*

In the previous sections, a simpler approach was considered to infer the proportion of (a)symptomatic infections from social contact data and incidence data. In this section, we present a follow-up study, described by Santermans *et al.* (2017), which employs and tests an extension. In this extension, we incorporate the aforementioned social contact data to inform mixing patterns in a compartmental model which we describe by a system of ordinary differential equations. Using the same incidence data as in Section 3.1.2.4, we infer on similar and additional parameters. We also investigate the possibility that the probability of developing ILI symptoms is related to the source of the infection, e.g. a symptomatic or asymptomatic individual.

Two different compartmental models are introduced and used to infer model parameters. The first one is called the non-preferential model (Figure 3.5) in which individuals either develop symptoms or not after a pre-symptomatic stage, but we consider no dependency on the status of the infector. In this case, we assume that susceptible individuals are infected at rate $\lambda(t)$. When infected, people enter the exposed compartment (E) in which they are infected but not yet infectious. Individuals become asymptomatic after a mean latent period $1/\gamma$ and thus enter the compartment I_1^a, with the superscript indicating the state of the person. A proportion of cases ϕ $(0 \leq \phi \leq 1)$ develops symptoms and $1-\phi$ of the cases remains asymptomatic. Thus infectious individuals move from the asymptomatic compartment I_1^a to the symptomatic I^s or asymptomatic I_2^a compartments with a rate $\phi\theta$ and $(1-\phi)\theta$, respectively. Individuals move to the recovered compartment (R) at rates σ^a and σ^s, respectively. The second model is the preferential model (Figure 3.6) which is an extension of the first model as it keeps track of whether a susceptible individual obtained the infection from an asymptomatic or a symptomatic case. Infection by an asymptomatic case happens at a rate $\lambda_a(t)$ and by a symptomatic case at a rate $\lambda_s(t)$, respectively. When an asymptomatic case infected the infected individual, it moves from S to E_a. If, on the other hand, the infector is symptomatic, the infected individual moves to E_s. Similarly, as in the non-preferential model, people become asymptomatic infectious at rate γ and thus move to I_a^a or I_s^a. Next, they can either remain asymptomatic or develop symptoms. Let ϕ_a be the probability that an individual infected by an asymptomatic case remains asymptomatic and ϕ_s be the probability that an individual infected by a symptomatic case develops symptoms.

As there was no information in the literature, at that time, on possible differences in the incubation and the latent period between individuals infected by symptomatic and asymptomatic cases, the length of the incubation period was assumed to be independent of the infector-type. Under this assumption, the preferential model simplifies to the non-preferential model if $\phi_s = 1 - \phi_a$.

Adding an age structure to the model can be done through the social contact data and by considering K age classes. Using the definitions and notations from Section 1.2.3.3, we denote the force of infection for the non-preferential transmission model as follows:

$$\boldsymbol{\lambda}(t) = \boldsymbol{\beta}^a \times (\mathbf{I}_1^a(t) + \mathbf{I}_2^a(t)) + \boldsymbol{\beta}^s \times \mathbf{I}^s(t), \tag{3.3}$$

where × denotes matrix multiplication. For the preferential transmission model the rate at which a susceptible individual acquires infection from an asymptomatic or symptomatic individual at time t, respectively, are given by:

$$\boldsymbol{\lambda}_a(t) = \boldsymbol{\beta}^a \times (\mathbf{I}_a^a(t) + \mathbf{I}_s^a(t) + \mathbf{I}^a(t)), \tag{3.4}$$

$$\boldsymbol{\lambda}_s(t) = \boldsymbol{\beta}^s \times \mathbf{I}^s(t). \tag{3.5}$$

The total force of infection is then $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_a(t) + \boldsymbol{\lambda}_s(t)$.

Taking into account that not all ILI cases are reported via general practitioners we introduce an under-reporting rate which differs by age. We use a likelihood-based approach by assuming

$$y_{k,j} \sim \mathrm{Po}(\rho_k(I^s_{new,k}(j) - I^s_{new,k}(j-1))), \tag{3.6}$$

where $y_{k,j}$ is the observed number of new cases in age group k in week j. $I^s_{new,k}(t)$ is the expected cumulative number of new symptomatic cases in age group k at time t obtained by solving $d\mathbf{I}^s_{new}(t)/dt = \phi\theta\mathbf{I}^{\mathbf{a}}_{\mathbf{1}}(t)$ for the non-preferential model and $d\mathbf{I}^s_{new}(t)/dt = (1-\phi_a)\theta\mathbf{I}^{\mathbf{a}}_{\mathbf{a}}(t) + \phi_s\theta\mathbf{I}^{\mathbf{a}}_{\mathbf{s}}(t)$ in the preferential model. ρ_k $(k = 1, \ldots, 5)$ denote the age-specific reporting rate of ILI cases. The system of differential equations is initiated by taking into

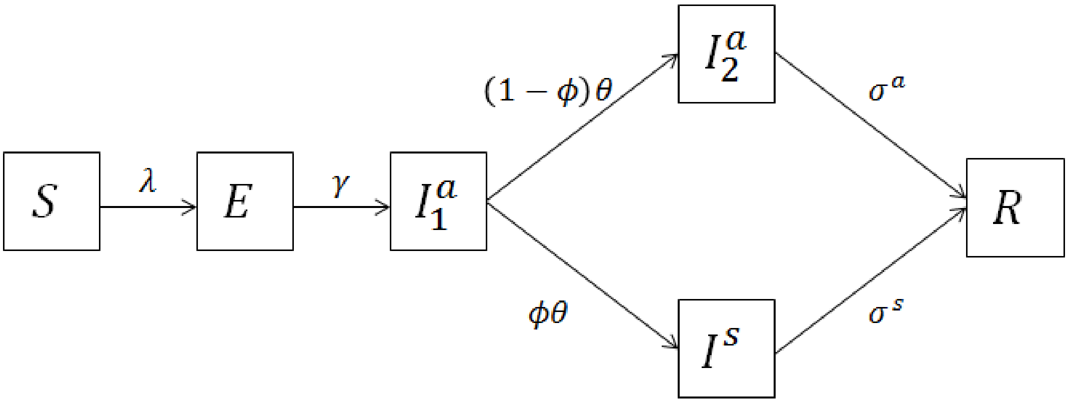


Figure 3.5: Schematic diagram of the non-preferential transmission model. Superscripts indicate presence (s) or absence (a) of symptoms. Figure from Santermans *et al.* (2017).

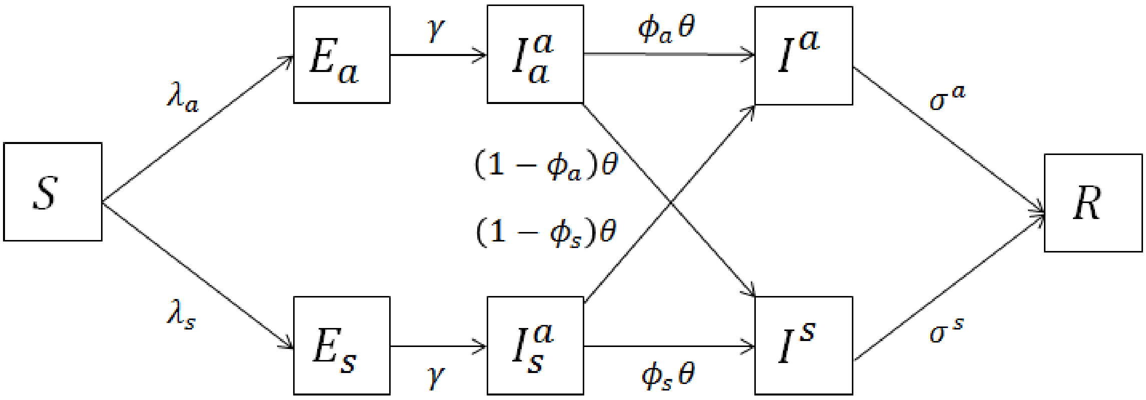


Figure 3.6: Schematic diagram of the preferential transmission model. Superscripts indicate clinical status of the infected individual: symptomatic (s) or asymptomatic (a). Subscripts indicate whether the infector was symptomatic (s) or asymptomatic (a). Figure from Santermans *et al.* (2017).

account the pre-existing immunity to the pandemic strain $\mathbf{S}(0)$. Furthermore, since we observed a large impact of the initial number of symptomatic cases $\mathbf{I}^s(0)$, these five parameters (one for each age category) are included in the estimation procedure. The number of asymptomatic cases at time 0 is assumed to be 0. The initial number of recovered individuals is $\mathbf{R}(0) = \mathbf{N} - \mathbf{S}(0) - \mathbf{I^s}(0)$, with $\mathbf{N}$ the population size. Our aim is to estimate ϕ, ϕ_a, ϕ_s, q^a, q^r, ρ_k, and $I^s_k(0)$ $(k = 1, \ldots, 5)$. Other parameters are assumed to be known based on a literature review by Dorjee *et al.* (2013). The model was not able to capture all age-specific reporting rates, thus the reporting rate in one age class is fixed and all others are interpreted as estimated relative differences. Parameters are estimated via a Markov Chain Monte Carlo (MCMC) approach. Uninformative priors are used for all parameters except for the initial number of cases, for which the priors are based on the ILI incidence of week 22 from the data used in Section 3.1.2.4. Whenever necessary, transformations are applied, as they ensure the estimates to lay within their proper parameter space. For ϕ, ϕ_a, ϕ_s and ρ_k $(k = 1, \ldots, 5)$ representing the age-specific reporting rates, a logit transformation is used whereas for q^a and q^r a log transform is taken. q^r is defined as before, i.e. q^s/q^a and assumed to be larger than 1 as symptomatic cases are considered to be more infectious than asymptomatic cases.

We furthermore extend the findings from the previous sections stating that social contact behaviour at home was significantly different from social contact behaviour when symptomatic. We do this by assuming that only a proportion p of symptomatic individuals stay at home immediately after symptom onset. The contact matrix for symptomatic individuals $\mathbf{C}^s$ is replaced by $p\mathbf{C}^s_h + (1-p)\mathbf{C}^s$, with $\mathbf{C}^s_h$ the contact matrix based on contacts made at home by symptomatic individuals. As such we assume that these contact rates do not increase when individuals stay at home. For fixed values of p, we solved the system of ODEs from this isolation model by using the obtained posterior parameter samples from the (non-)preferential model. This way we can assess the impact of p on the difference in the number of (a)symptomatic cases.

For both transmission models the posterior credible intervals for ϕ (ϕ_s for the preferential model) and q^r are wide, indicating the difficulty to estimate these parameters from the data (Table 3.4). The parameters show a strong link, and hence we can either estimate the proportion of symptomatic cases or the relative infectiousness from the data at hand. Using the non-preferential model we see that when only 15% of the cases develop symptoms, symptomatic cases are estimated to be about 2.76 times more infectious than asymptomatic cases. Furthermore, the reporting rate in the 15-44 years age group is estimated to be about 1.15 times higher compared to the reporting rate in the 45-65 years age group. For the 65+ age group, this is 0.75 times lower compared to the reporting rate in the 45-65 years age group. Using the preferential model we see that when 22% of cases infected by a symptomatic case develop symptoms, symptomatic cases are estimated to be about 2.62 times more infectious than asymptomatic cases. The reduced reporting rate for the 65+ age groups is also confirmed. The preferential transmission model has a better fit compared to the non-preferential model, given the lower DIC value. Furthermore, the preferential model simplifies to the non-preferential model when $\phi_s = (1 - \phi_a)$. When calculating the difference between these two parts for each posterior sample, a 95% credible interval of $[0.05, 0.55]$ is obtained. This interval indicates that we cannot simplify the preferential model to the non-preferential model.

Concerning the effect of home isolation, we find an increase in reduction of cases when the proportion p of symptomatics staying at home increased. From Figure 3.7 we see no visible difference between the symptomatic and asymptomatic cases for the non-preferential model, for the preferential model, we observe a larger reduction

Table 3.4: Posterior median (95% posterior credible intervals) and DIC value for the non-preferential and preferential model. Table from Santermans *et al.* (2017).

Parameter	Non-preferential	Preferential
ϕ	0.15(0.04, 0.39)	
ϕ_a		0.98(0.89, 1.0 0)
ϕ_s		0.22(0.062, 0.58)
q^a	0.082(0.069, 0.093)	0.10(0.092, 0.12)
q^r	2.76(1.04, 9.21)	2.62(1.04, 9.20)
ρ_1	0.21(0.18, 0.25)	0.20(0.17, 0.24)
ρ_2	0.20(0.17, 0.22)	0.20(0.18, 0.23)
ρ_3	0.23(0.20, 0.25)	0.21(0.19, 0.23)
ρ_4	0.2	0.2
ρ_5	0.15(0.12, 0.19)	0.15(0.12, 0.18)
R	1.36(1.33, 1.40)	1.41(1.23, 1.63)
DIC	298.75	288.49

in symptomatic cases compared to asymptomatic cases. The reduction of cases is greater according to the non-preferential model in comparison to the preferential model. In conclusion, this follow-up study shows the possibility to estimate parameters related to asymptomatic infection using data on symptomatic cases only, mainly due to the present differences in mixing behaviour as shown in the previous sections of this chapter. Furthermore, using a Bayesian approach, it is possible to estimate the proportion of symptomatic infections or the relative infectiousness of symptomatic cases compared to asymptomatic cases in the non-preferential model. Next, the data support the preferential transmission hypothesis, i.e. the development of ILI symptoms depends on whether the infector was a symptomatic or an asymptomatic case. Lastly, we found a reduction in the total number of cases of 39% [0.30, 0.45] when 50% of individuals would stay home immediately after symptom onset. On the other side, if everyone with symptoms would stay home, the observed reduction was 63% [0.53, 0.70].

Lin *et al.* (2016) use a model similar to the non-preferential model to explore the trade-off between contact rates and infectiousness, i.e. decreasing contact rates and increasing infectiousness with increasing symptom severity. They found a non-monotonically varying R_0 with symptom severity, implying that certain interventions such as antivirals for influenza, can increase R_0. This research highlights the importance of using empirical data describing the relation between contact rates and symptom severity in epidemiological models. Ball and Britton (2007) use a similar model to the preferential model but assuming homogeneous mixing in the population and not estimating model parameters. They showed that in certain situations the proportion of mildly (asymptomatic in our case) infected individuals could increase with increasing vaccination coverage. This emphasises the practical importance of our model for a broad range of pathogens with different levels of symptom severity.

One of the biggest limitations to the presented approach is that the reporting rates are not estimable from the data. Another assumption made is that the contact behaviour of asymptomatic cases is similar to the contacts of recovered individuals. Given the presented comparison with the historical control of the POLYMOD Study, this seems to be a reasonable assumption. Nevertheless, it is possible that asymptomatic cases do not feel entirely in

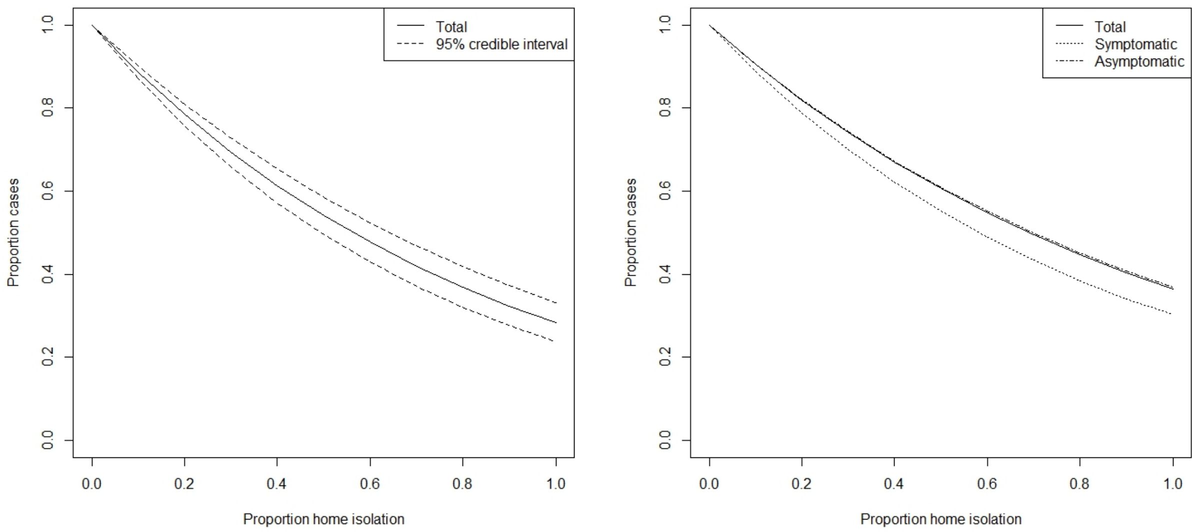


Figure 3.7: Proportion of cases plotted against the proportion of symptomatic individuals staying home immediately after symptom onset. Left panel: reduction in total number of cases for the non-preferential model with 95% confidence intervals. Right panel: reduction in the number of total, symptomatic and asymptomatic cases for the preferential model. Figure from Santermans *et al.* (2017).

best shape and thus might have a different contact behaviour than healthy individuals. The conclusions reported are based on data from A/H1N1pdm in 2009, and it is unclear how this applies to other influenza strains. If the aim is to extend this model to other diseases, more empirical data on how contact rates change with symptom severity are needed, hence the reason for Follow-up Study II.

3.3 Follow-up study II: The impact of influenza-like-illness and chronic diseases on social contact behaviour - a pilot study.

This section is based on work in preparation: "Van Kerckhove, K. et al. The impact of influenza-like illness and chronic diseases on social contact patterns relevant for infectious disease transmission - a pilot study. In preparation.".

3.3.1 Introduction

In Section 3.1.4, we showed the importance of the study conducted during the influenza A/H1N1pdm 2009 epidmic. Next, we also indicated the need to confirm the results for other influenza strains to properly focus the public health efforts to reduce or prevent transmission to others. Furthermore, in Section 3.2 we mentioned the need for additional surveys when one aims to extend the proposed models to other infectious diseases. Therefore, we conducted a pilot study, described in Section 2.5, to test the feasibility of a large-scale survey to quantify the effect of seasonal influenza on social contact behaviour. Driven by the fact that people with a chronic condition suffer more from influenza infection (World Health Organization, 2017a), we additionally aimed to investigate the social mixing behaviour of chronically ill people. In Section 3.2, we showed the ability to quantify the proportion of (a)symptomatic infections based upon data from symptomatic cases only, extending this method to people with chronic conditions and implementing intervention strategies might help in gaining understanding about the possible effects of those intervention strategies. Here, we present the preliminary results of the ICSOC Study and discuss the important aspects for a large-scale survey.

3.3.2 Results

Based upon the returned - not yet distributed - questionnaires, we quantify the response rate to vary between 15 and 100% with a median rate of 42.59% and a mean rate of 50.75%. The response rate of the follow-up questionnaire is 67.57%, based upon those participants whom entered their postal address. Booklets are predominantly filled in during the week with no booklets filled in on a Sunday, possibly induced by a reduction in consultation rates during the weekends (Figure 3.8).

Due to the small number of participants and the skewed distribution of the total number of contacts we base comparisons of the number of contacts upon the Wilcoxon rank sum test with exact p-values. We perform the analyses in R and JMP.

In each of the groups, except the healthy group, more female than male participants are present with larger differences when chronic conditions are involved. As such in the *Chronic only* group there are 12 females compared to 6 males and in the *Chronic and ILI* group 9 out of 10 participants is a woman. In the paired sample the genders are almost equally distributed. The median and mean ages in the groups with chronic conditions are higher compared to the group with ILI cases only. The mean (and median) age in the *Chronic only* group is 59.44 years (57 years), in the *Chronic and ILI* group this is 48.90 years (49.5 years) compared to 32.71 years (30 years) in the *ILI only* group. In the *Healthy* group, the mean age is 44.77 years, and the median age is 39 years. In the *Paired sample* the mean age is 34.57 years, and the median age is 27.5 years.

Given the broad categories of chronic conditions proposed, we see a large variety of chronic conditions in our sample as depicted in Table 3.5. Besides the offered choices, participants could note other chronic diseases, a selection being tonsil inflammation, angioneurotic oedema, bladder stoma, Crohn disease, diabetics, and problems with the thyroid gland.

Based on the different groups, various research questions can be stated and possibly answered. Figure 3.9 gives an overview of the number of contacts across the various groups. As indicated an overlap between the paired and other groups exist and thus in case the overlap is present, the figure shows the groups without the overlapping individuals, referred to as single. In the paired case, the four boxes on the right in each panel, we discriminate between all paired cases (most right) or paired cases without those having a chronic condition, which is comparable to the comparison in Section 3.1. The left panel shows the comparison for all contacts; clearly, the healthy individuals report a higher number of contacts compared to the other groups. The number of contacts in the *Chronic only* group shows a larger spread compared to the *ILI only* and *Chronic and ILI* group. In the single comparison, the *Chronic only* group shows a higher number of contacts compared to the *Chronic and ILI* group. This tendency is also visible in the paired comparisons. Similar, yet less pronounced effects are present when considering all contacts except the work contacts. In general, participants report few contacts at work which is

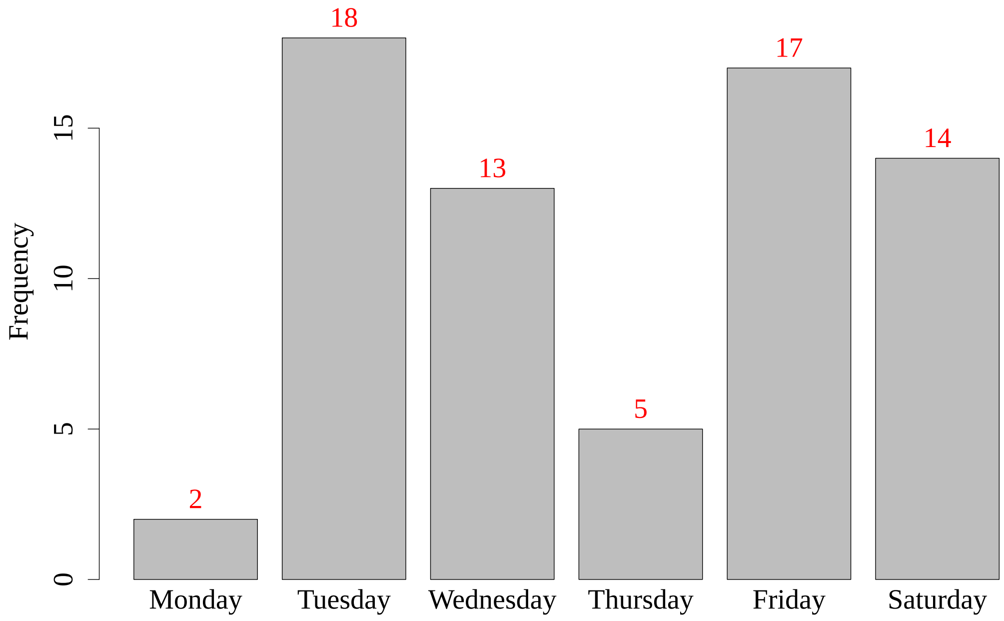


Figure 3.8: Distribution of booklets over the days of the week. The numbers represent the number of booklets filled in on a respective day, no booklets are filled in on a Sunday.

Table 3.5: Chronic conditions and their presence in the sample.

Condition	**Nb. participants**
Arthrosis	7
Heart and vascular disease	4
Epilepsy	2
Obesity	3
Hypertension	5
Rheumatoid arthritis	2
HIV / AIDS	1
Asthma	1
COPD	1
Rheumatism	5
Lyme disease	0
Chronic depression	1
Hepatitis C	0
Cardiomyomathy	0
Other	14

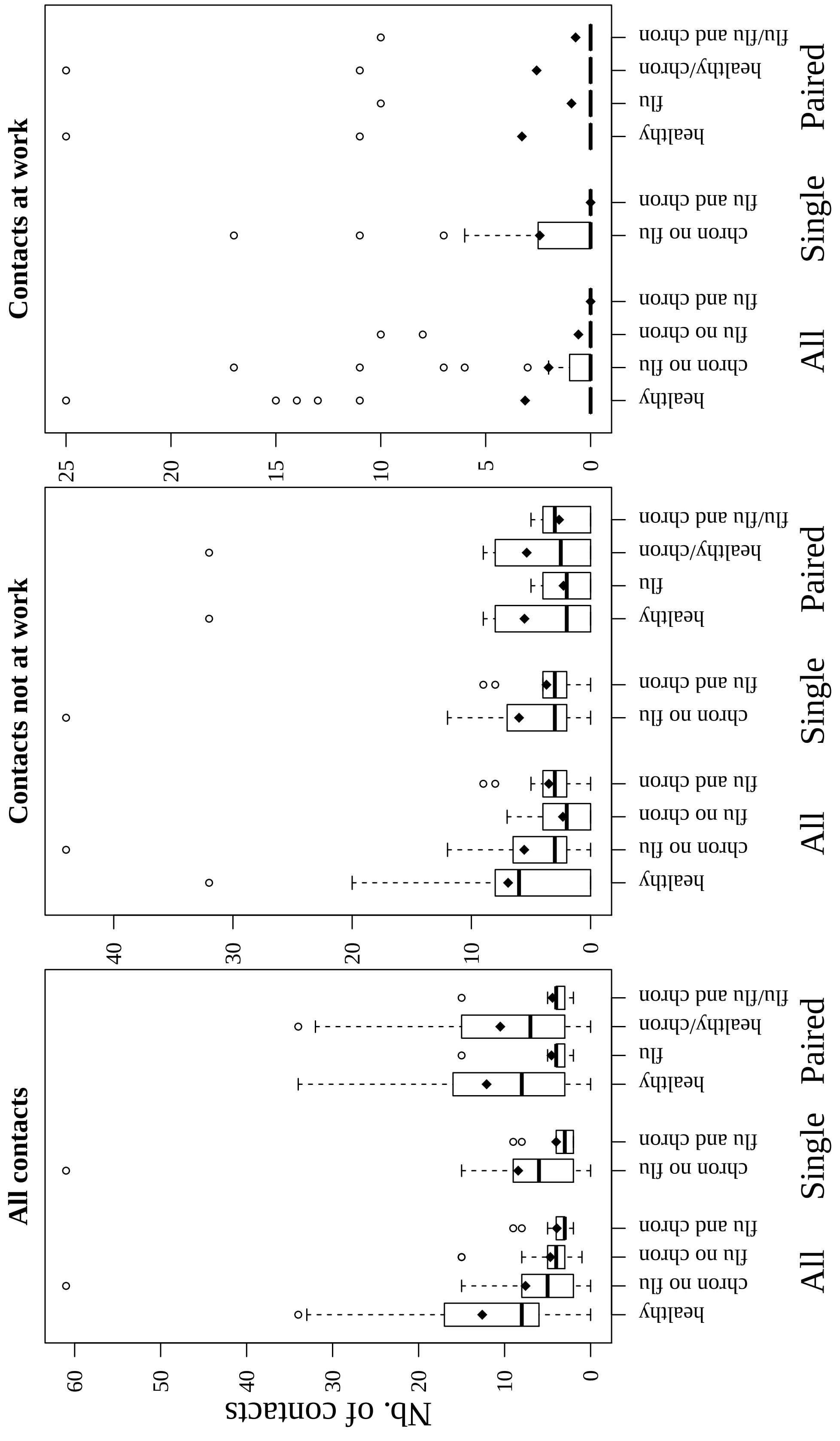


Figure 3.9: Visual comparison of the number of contacts. The panels show from left to right the box plots of all contacts, including additional work contacts if applicable, the number of contacts without work contacts and the work contacts only. In each panel, the different groups are displayed where three main differentiations can be observed based on all data (first four boxes), based on the single data hence ignoring those booklets available in both groups (fifth and sixth box) and based on the paired data (last four boxes). The paired data either involves only those who had an ILI only and were recovered the second time (first two boxes) or includes those people who had a chronic condition as well (last two boxes).

evident from the right panel. Significant differences in the total number of contacts (including work contacts) are obtained for the comparison between *Chronic only* (n=23) and *Healthy* (n=25) (p=0.0065) and between *Chronic and ILI* (n=14) and *Healthy* (n=25) (p=0.0006). The comparison of the number of contacts within the paired sample, zooming in on the participants with an ILI only (n=11), we obtain a borderline result (p=0.0527) for the one-sided alternative, e.g. more contacts in a healthy condition compared to an ill condition. No other significant differences are observed, not even for the work contacts. However, the sample is only limited, which might be an issue for this case. Another issue is a bias as other determinants are not taken into account for this pilot study.

From the small survey (6 GP's participated) concerning the experience of the general practitioners, we observe that general practitioners mostly remembered to distribute booklets (4 out of 6) and that patients were mostly open to participate, however, 2 out of 6 indicated a lack of time to distribute the questionnaires. The time they spent when distributing varied between 30 seconds and 5 minutes and most of them reported to have explained a little, half of them even stated to have given a thorough explanation. Four of the general practitioners indicated the quota for age to be unreachable. Common reasons for refusal for participation are the difficulty of the booklet or feeling too ill. Besides the proposed quota, the general practitioners indicate to have selected patients based on language (Dutch), intelligence, sufficient amount of spare time or other reasons. Albeit these aspects, three general practitioners would perhaps join a large-scale survey; one indicated to enter when some changes would be made, and two indicated to be interested in participating in such a large-scale survey. Some general practitioners offer additional comments and suggestions indicating the booklets to be rather complex, hence requesting extra time from the general practitioners during an already busy period. Furthermore, the discrepancy of chronic only and chronic with an ILI was initially not clear to one of them.

3.3.3 Discussion & Conclusion

Notwithstanding the size of this pilot study, we observe a significant difference in the number of social contacts between people with a chronic condition and healthy individuals. This result implies that a difference is present and interesting to investigate more deeply. Additionally, we observe a significant difference in the number of contacts whenever a person has a chronic condition with an ILI compared to being healthy. Whether this is attributed to the chronic condition, the ILI or both at the same time is not distinguishable from the current data. Hence, a large-scale study could investigate this aspect. Although borderline, the results point towards similar conclusions when comparing one's behaviour during an ILI or when being healthy as was the case in Section 3.1 (Eames *et al.*, 2010; Van Kerckhove *et al.*, 2013).

In the current study, we considered a broad range of chronic conditions, as an exploration. However, different chronic diseases can lead to various types of behaviours. Some chronic diseases may add a substantial burden to patients and hence restrict their social contacts, whereas other chronic diseases allow the participant to be socially active. Another aspect that can influence the social contact behaviour is the need for health care assistance, suggesting a possible increase in the number of contacts. In a future study, one can limit the chronic conditions or make sure in the design that a sufficient amount of people are recruited for the various chronic conditions.

The feedback from the general practitioners is valuable, as they highlight the complexity of the booklets, additional selection biases and the time-consuming part to them. Concerning the first aspect, this complexity was also apparent from the data as participants did not fill in some questions as intended. Two such issues are the description of an infectious disease and the use of antiviral medications; participants had difficulties with both these issues. Hence another translation or more explanation is required there. The lengthiness of the questionnaire induces an additional complexity. This aspect is probably the driving factor for general practitioners to select their patients and causes the time-consuming aspect as well.

This survey, being the first to investigate changes in social contact behaviour as a result of seasonal influenza and, as important, of chronic conditions, offers valuable insights and opens avenues for future research. Aspects to be considered when setting up a large-scale survey include those already mentioned. Given the expertise of the general practitioners and the need for collaborating with them, the discussion of the design can be done with their input through a focus-group concept. This concept allows the general practitioners and the researchers to exchange experiences and find possible solutions to certain issues. Furthermore, to conduct the survey at the country-level or with other countries, collaborations with the sentinel networks of general practitioners and with public health authorities are necessary.

3.A Additional details for sensitivity analysis to type of contact

Table 3.A.1: A comparison of the basic reproductive number for different types of contacts, with the 95% bootstrap-based confidence intervals without capping the number of contacts at 33. Table from Van Kerckhove *et al.* (2013).

Based on	**Comparison (Number of participants)**	R_0 **ratio**	**95% CI**
All encounters	Ill vs. healthy (140)	0.22	0.09, 0.43
	Healthy (159) vs. POLYMOD GB (1,012)	1.33	0.88, 2.45
	Ill (283) vs. POLYMOD GB (1,012)	0.34	0.28, 0.43
Skin to skin encounters	Ill vs. healthy (140)	0.45	0.22, 0.59
	Healthy (159) vs. POLYMOD GB (1,012)	0.84	0.58, 1.16
	Ill (283) vs. POLYMOD GB (1,012)	0.36	0.30, 0.49

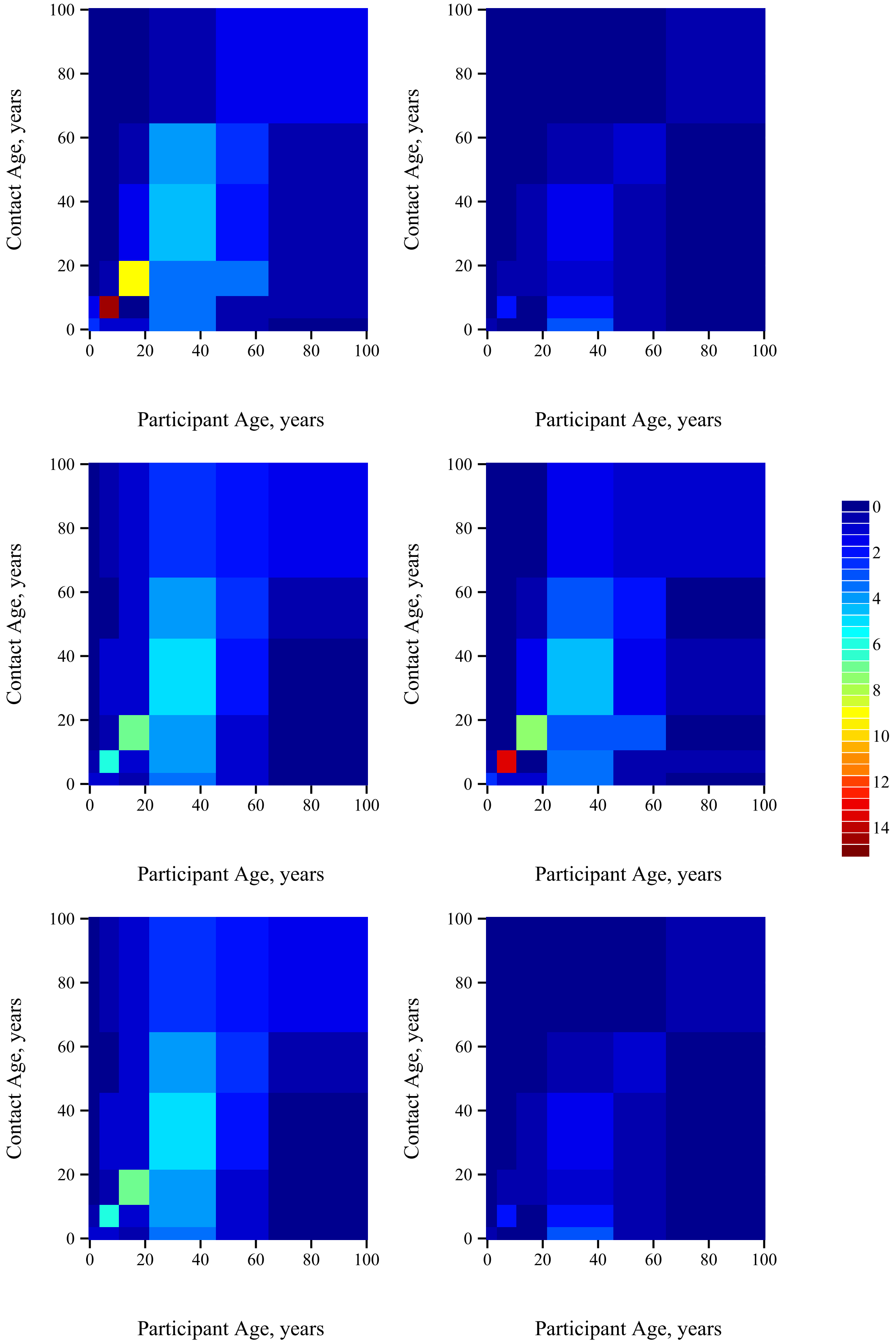


Figure 3.A.1: The daily number of contacts between age groups. The contact matrices are based on all contacts, with the number of contacts capped. The upper panel shows the contact matrices for the paired comparison between the healthy (left, n=140) and ill (right, n=140) samples, the middle panel shows the comparison between POLYMOD participants from GB (left, n=1,012) and the healthy participants (right, n=159) and the bottom panel displays the comparison between the POLYMOD GB (left, n=1,012) and ill (right, n=283) participants.

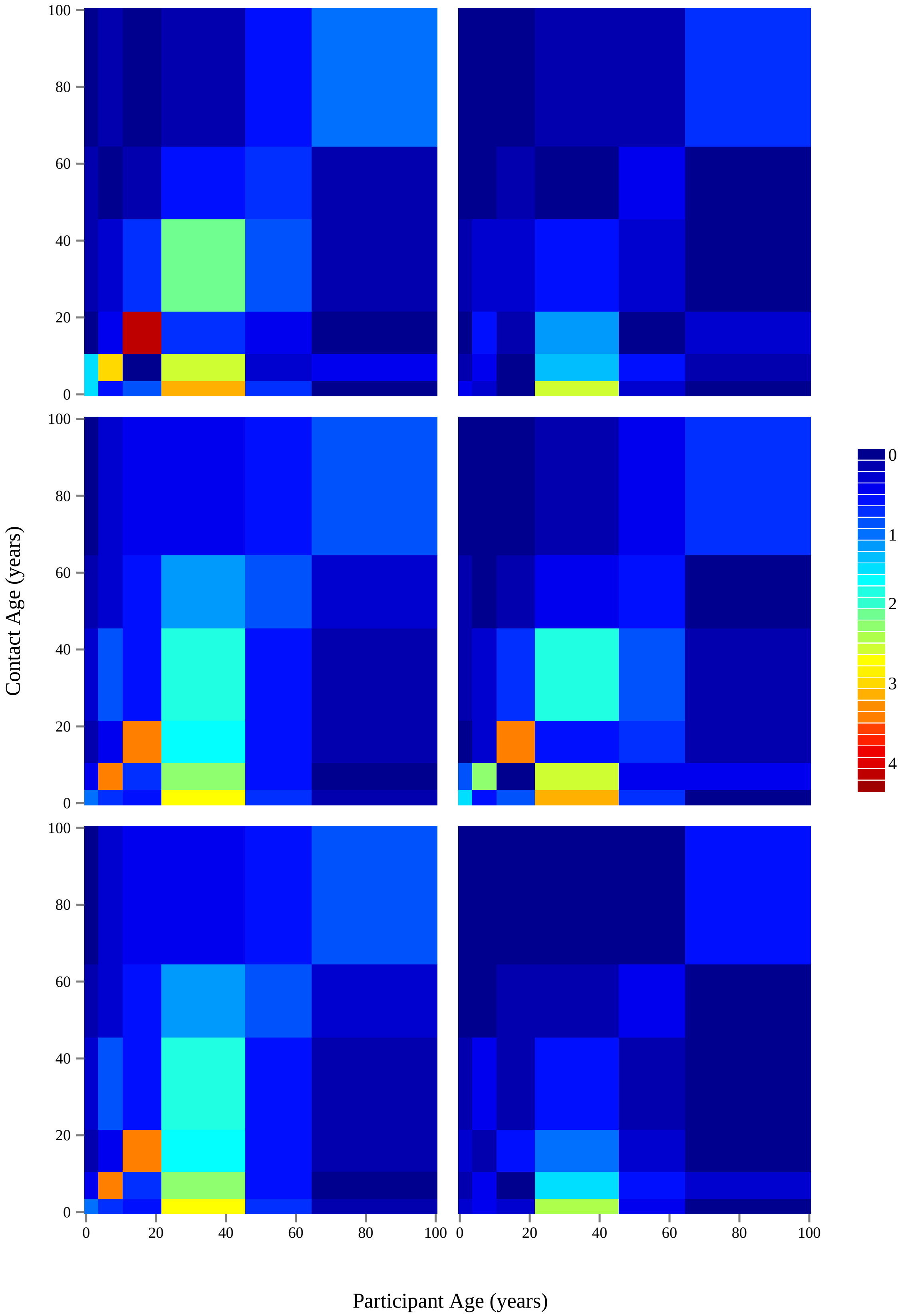


Figure 3.A.2: The daily number of contacts between age groups. The contacts are based on the physical contacts, with the number of contacts capped. The upper panel shows the contact matrices for the paired comparison between the healthy (left, n=140) and ill (right, n=140) samples, the middle panel shows the comparison between POLYMOD GB (left, n=1,012) participants and the healthy (right, n=159) participants and the bottom panel displays the comparison between the POLYMOD GB (left, n=1,012) and ill (right, n=283) participants.

Chapter 4

The Impact of Weather and Animal Ownership on Social Contact behaviour and the Structure of Within-Household Networks

4.1 Introduction

In comparison to the previous large-scale social contact survey, the Flemish study anno 2010-2011 considered several renewing aspects, including questions related to animal ownership and touching and the household survey. In this chapter, we use these two facets and combine information on weather with the social contact data to investigate three aspects: (1) the interplay between weather and social contact behaviour, (2) the interplay between animal ownership or touching and social contact behaviour and (3) the structure of within-household networks.

In Section 4.2, we describe the first aspect being the interplay of weather conditions on social contact behaviour. Previous research has shown that temperature and humidity exhibit a distinct seasonality in temperate climates. Absolute humidity has been hypothesized to drive influenza seasonality through modulating airborne survival and transmission (Shaman and Kohn, 2009; Shaman *et al.*, 2010; van Noort *et al.*, 2012). Furthermore, airborne transmission is believed to be the dominant mode of transmission in temperate regions (Lowen *et al.*, 2007). Social contact behaviour is predictive for the dynamics of infectious disease transmission (Wallinga *et al.*, 2006; Mikolajczyk *et al.*, 2008; Mossong *et al.*, 2008; Cauchemez *et al.*, 2008; Hens *et al.*, 2009a; Cauchemez *et al.*, 2009a; Eames *et al.*, 2012) and is influenced by different aspects as described in Chapters 2 & 3.

While most literature on influenza transmission aims to reveal the biological and physical mechanisms associated with different weather conditions, the influence of weather on social contact patterns remains unclear. Mikolajczyk *et al.* (2008) reported that school children reduced their contacts on rainy days by 16%. Nonetheless, seasonality of many respiratory diseases might be driven in part by the tendency of people to congregate indoors when the weather is bad (Lofgren *et al.*, 2007). Crowding may occur the year round at public gatherings like shopping malls, festivals, sporting events and conferences. Therefore, seasonal fluctuations in social contact patterns may not be large, however, they contribute to the spread of influenza during winter. (Lofgren *et al.*, 2007).

In Willem *et al.* (2012) we explore modifications of social contact patterns with respect to climatological changes and their effect on influenza seasonality. In the absence of a previous in-depth weather impact analysis of social contact patterns, we identify opportunities and limitations of existing datasets. Using the social contact data as described in Section 2.4, we create a partition according to the weather on the day of the survey and estimate for each weather condition the relative changes in contact patterns by the ratio of the mean number of contacts and of the R_0 values (Hens *et al.*, 2009b). In Section 4.2 we elaborate on the methods and results of Willem *et al.* (2012).

Second, we study the interplay between animal ownership and human-animal interactions and social contact behaviour. As many emerging and re-emerging infectious diseases are transmitted to humans via animals (Cleaveland *et al.*, 2001; Taylor *et al.*, 2001; Woolhouse and Gowtage-Sequeria, 2005; Jones *et al.*, 2008), there is a public health interest in human-animal interactions. This research topic is important especially in low and middle-income countries where social, cultural and environmental circumstances may increase interspecies transmission, with the potential to start a new pandemic in humans. Nonetheless, in high-income countries (HICs), infectious diseases in humans such as hantavirus, rabies, tick-borne encephalitis and Lyme disease are almost exclusively transmitted to humans by animals, though these have generally no potential to spread further between humans. Moreover, in HICs, about half of the households own pets (Voith, 2009) and these companion animals may carry zoonotic infections (Greene and Levy, 2006). In Kifle *et al.* (2016), we study both animal ownership and human-animal interactions through touching, as they both represent important risk factors for zoonotic infectious disease transmission, based on the data described in Section 2.4. We relate the human-animal interaction with social contact frequencies between humans, and propose a method to estimate the associated risks of causing a major zoonotic outbreak, conditional on a flu-like pathogen being present in each of the animal groups. We discuss this work further in Section 4.3.

The third aspect studies the structure of within-household networks based on the household survey in which households with young children are surveyed and each household member participates. As households are crucial units in the epidemiology of airborne diseases, e.g. influenza, smallpox and SARS, relations between household members are typically incorporated in models. These relations are typically characterized by frequent and intimate contacts, that allow for rapid disease spread within the household after introduction of an infectious case. Ferguson *et al.* (2006) state that the largest single risk factor for being infected oneself is being a member of a household containing an influenza case (Longini *et al.*, 1982; Cauchemez *et al.*, 2004). On top of that, households allow an infection to spread from schools to workplaces and vice versa, hence they hold a bridging function. Inference from household final size data revealed that children play a key role in bringing influenza infection into the household and in transmitting the infection to other household members (Cauchemez *et al.*, 2004).

The assumption of homogeneous mixing within households is often made in epidemic models. In early work, Reed-Frost type of models were used to estimate household and community parameters from household final size data, assuming a constant probability of infection from the community (Longini Jr and Koopman, 1982; Becker, 1989; Addy *et al.*, 1991). The so-called 'households model' is a generalisation, by Ball *et al.* (1997), of the previous models by including two levels of mixing whereby assuming random mixing within households (local) and in the entire population (global). The latter typically occurs at a much lower rate. The analytical tractability of this model allowed the theoretical study of epidemic phenomena. This has led to the definition of threshold parameters such as the reproduction number R^*, representing the average number of households infected by a typical infected household in a totally susceptible population (Ball *et al.*, 1997; Ball and Neal, 2002). Meyers *et al.* (2005) used a contact network model in an urban setting incorporating households as complete networks (cliques) to explain the early epidemiology of SARS. Individual-based simulation models of infectious disease transmission incorporate detailed individual-level information in order to account for heterogeneities relevant to the application (e.g., demography, socio-economics, genetics; Chao *et al.*, 2010b; Mniszewski *et al.*, 2008; Grefenstette *et al.*, 2013). Sometimes these models also incorporate details related to specific settings such as schools and workplaces, whereas assuming random mixing in households. Studies highlighting within-household transmission and control policies targeting households can be found in Halloran *et al.* (2002) and Ferguson *et al.* (2006).

Until recently there was no direct empirical evidence to support the assumption of homogeneous mixing within households. Egocentric contact surveys entailed partially observed within-household contact networks and only allowed for indirect inference of the unobserved network links (Potter *et al.*, 2011; Potter and Hens, 2013). It has been argued that by considering different household compositions and contact heterogeneity within households (Danon *et al.*, 2011) a more realistic modelling could be achieved.

Therefore, in Goeyvaerts *et al.* (In preparation) we use the household contact survey (Section 2.4) to empirically assess the assumption of homogeneous mixing, e.g. by studying the effect of age and gender on social distance within households. Furthermore, the data obtained provides an answer to one of the key questions from inference on household models: how does the density of the contact network scales with the household size (Danon *et al.*, 2011)? When ignoring contact heterogeneity between household members, the contact network density equals the contact rate between two individuals in a household and is a determinant for the within-household transmission rate of airborne infectious diseases (Wallinga *et al.*, 2006; Mossong *et al.*, 2008). Lastly, we assess reporting quality for diary-reported contact surveys by looking at reciprocity, i.e. symmetry in contact reporting. We use Exponential family Random Graph Models (ERGMs; Robins *et al.*, 2007) to gain insight in the factors driving contacts amongst household members and to develop a plausible model for within-household physical contact networks. In the mise en scene of the two-level mixing model, we compare the empirically grounded ERGMs to the random mixing assumption by using stochastic simulations of an epidemic in the model. We elaborate on these analysis in Section 4.4.

4.2 A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission

This section is based on the published work: "Willem, L., Van Kerckhove, K., Chao, D.L., Hens, N., Beutels, P. (2012). A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. PLoS One, **7***, e48695.".*

As mentioned in the introduction, the data from the Flemish Study (Section 2.4) enables to link social contact behaviour to weather conditions. We investigate this link by combining the described data with public data on daily precipitation and temperature from the National Oceanic and Atmospheric Administration (NOAA, http://www.noaa.gov/). To be consistent with previous findings related to influenza transmission (Shaman and Kohn, 2009), we calculate the vapor pressure, as a measure for the absolute humidity, from the dew point depression according to the work of Wallace and Hobbs (2006).

We divide the contact data in subsamples based on the type of day and the weather conditions. The subsamples for the weather conditions are created using median values of daily weather condition measures. As described in Section 1.2.3.1, dairy weights are applied and given the subsamples, they are recalculated for each specific situation. In this section, the weights are based on age, household size (Census data 2001) and day of the week, and they are capped at 3 to prevent disproportionate contributions of certain participants.

To quantify the differences between the various conditions, the mean number of contacts and the R_0 ratio are calculated. We include additional comparisons depending on specifics of the contacts, such as the location and duration. For the calculation of the R_0 ratio, and hence, the social contact matrix, we used the methods from Section 1.2.3.3 applying Eq. (1.9) to obtain R_0. We assume reciprocity when using all contacts, contacts involving skin-to-skin touching or contacts of a specific duration. Whereas for the contact matrices at specific locations, the reciprocity is not enforced. When assuming q to be constant the R_0 ratios are estimated using only social contact rates. Notwithstanding that this q might vary by weather condition, assuming a constant q enables us to focus on the effect of modified social contact behaviour. Uncertainty is expressed by 95% confidence intervals based on a nonparametric bootstrap (Section 1.2.4.2), with an age-stratification to retain the original age distribution.

The results are based on 1,752 participants of which 413 were sampled during a school holiday period or during a weekend adjacent to a holiday period. Overall, 1,046 participants reported on a regular weekday, 293 on regular weekend days, 286 on weekdays during public holidays and 127 on weekend days during or adjacent to public holidays. Apparently, the first holiday period in November coincided with higher temperature (Spearman correlation of 0.53), which is taken into account. This pattern is depicted in Figure 4.1 where the squares at the bottom indicate holiday periods. Between the precipitation and holiday variable the correlation was only 0.005. Figure 4.1 also shows, in the middle panel, that 70% of the participants reported on 20% of the survey period (October-November).

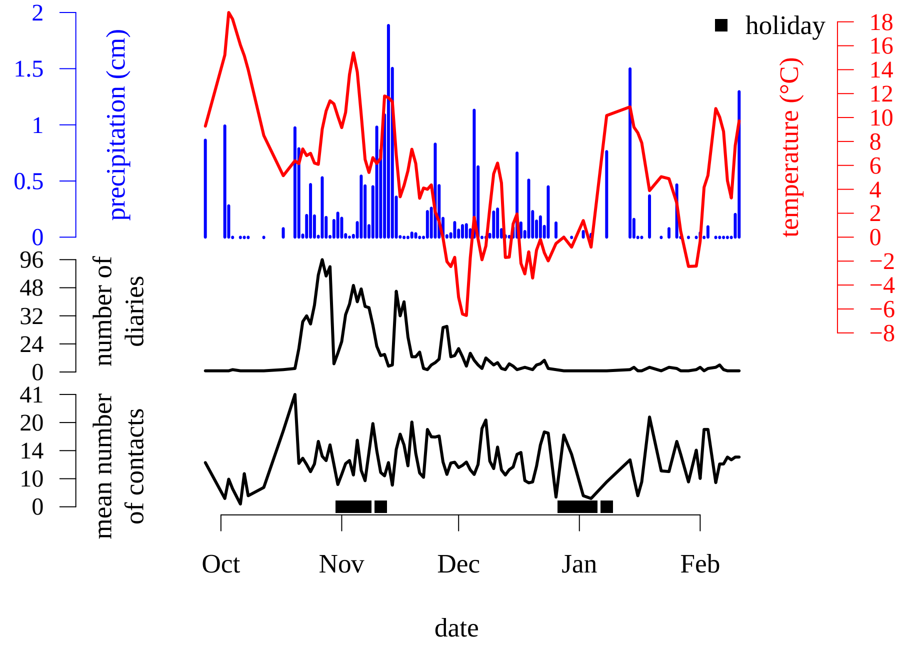


Figure 4.1: Overview of the weather and the social contact data from October 2010 until February 2011. Top: the mean daily temperature (line) and the total daily precipitation (bars). Middle: the daily number of diaries. Bottom: the mean number of social contacts per day. Holiday periods are marked at the both with squares. Figure adapted from Willem *et al.* (2012).

When comparing the data by day-types, we observe a significant increase in the number of contacts during weekdays compared to weekend days both for all contacts (1.553 [1.397,1.739]) and contacts involving skin-to-skin touching (1.184 [1.049,1.342]). Similar (significant) results are obtained when comparing regular to holiday periods.

During the period the data was collected the median temperature was 6.83°C, the median precipitation level was 0.05 cm/day and the median absolute humidity was 8.41 mbar. Remark that high and low temperature are low temperatures overall as participants mostly participated during winter. This information is considered in the partition of the contact data and in the estimation of the mean number of contacts as well as the R_0 ratios. Comparing the number of contacts on days with low temperatures to days with high temperatures, we observe a non-significant increase of 13% (95% CI: [0.993,1.283]). On the other hand, both school contacts and contacts of at least 15 minutes increase significantly (2.004 [1.363,3.027] and 1.190 [1.039,1.341], respectively) with cold temperatures, i.e. days with temperatures below the median value. We do not find any significant change in the contact pattern in relation to precipitation.

In addition, we study the regular and holiday periods separately to eliminate possible confounding by weather condition. We restrict our analysis to regular weekdays as a consequence to the limited sample size when considering weekend. Data sparseness is present as well for the precipitation levels, and consequently, a combined analysis for temperature and precipitation is underpowered even for regular weekdays.

Restricting ourselves to regular weekdays, we find a significant increase in the mean number of contacts of long duration ($>$ 1 hour) (1.188 [1.025; 1.361]), this tendency is also observed for all contacts taking longer than 15 minutes (1.143 [0.995; 1.307]). Also, we find a significant decrease in the number of non-specific contacts, which are contacts during leisure, transport, family visits and others (0.461 [0.276; 0.737]). In addition, we observe a significant increase in the number of contacts taking longer than 15 minutes (1.183 [1.032; 1.342]) and school contacts (1.502 [1.085, 2.041]) on days with low precipitation. For home and work contacts, no significant change is detected in relation to precipitation stratification. We observe a significant increase in the number of contacts lasting longer than 1 hour during regular weekdays (1.162 [1.004; 1.323]) with dry air and a significant decrease in the number of non-specific contacts (0.445 [0.279; 0.732]). Comparisons based on R_0 showed similar trends for the day-type, in line with previous results, e.g. transmission decreases during weekends and holiday periods. For the weather conditions, also similar conclusions based on the R_0 ratio are seen as based on the number of contacts, however non-significant.

4.3 Animal ownership and touching enrich the context of social contacts relevant to the spread of human infectious diseases

This section is based on the published work: "Kifle, Y.W., Goeyvaerts, N., Van Kerckhove, K., Willem, L., Kucharski, A., Faes, C., Leirs, H., Hens, N., Beutels, P. (2016). Correction: Animal ownership and touching enrich the context of social contacts relevant to the spread of human infectious diseases. PLoS One, ***11****, e0148718.".*

The Flemish survey (Section 2.4) included a questionnaire related to animal ownership and touching, allowing to study human-animal interactions and the effect on spillover infections. We use this data to model the animal ownership and animal touching by employing logistic regression models. We consider the total number of contacts, animal ownership and demographic as well as temporal factors as possible predictors for animal touching whereas for animal ownership we only consider socio-demographic factors. The demographic factors considered are age, gender, educational attainment (which is included as three separate covariates depending on age: educational attainment of the mother (0-12 years), educational attainment of the participant ($\geq$ 13 years) who was studying during the survey, educational attainment of the participant ($\geq$ 13 years) who was not studying during the survey), province (the five Flemish provinces) and household size (household sizes over 4 are comprised in one single group). The temporal factors reflect the type of day, i.e. a week- or weekend day and a holiday or regular period. We perform a stepwise backward selection procedure based on Akaike's Information Criterium (AIC; Akaike, 1974) which includes all possible interactions, thus starting from the most complex model, the model was reduced to the model with the lowest AIC. We apply further model building using a likelihood ratio test (LRT; at 5% level of significance) to assess the overall significance of the obtained covariates.

We also employ a weighted negative binomial regression model for the total number of contacts (including professional contacts, physical contacts and non-physical contacts). The weights are dairy weights similar to the ones discussed in Section 1.2.3.1 with age and household size (Census 2001 data) as factors and weights limited to 3. The predictors for this model are the demographic and temporal factors combined with animal ownership and all significant two-way interaction terms. The same selection criteria as for animal ownership and touching are used.

Lastly, we calculate the probability of a major outbreak of zoonoses with flu-like transmissibility, conditional on the presence of zoonoses in the animal groups. To do so, we assume that animal pathogens can infect humans through owning or touching infected pets, livestock and poultry. Due to the age and socio-demographic factors of some people, they are at higher risk of infection compared to others when in contact with these infected animals. We categorise age based on the Belgian Census data of 2000 into 15 categories: 0-4, 5-9, 10-14, 15-19,20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, +70 years. We estimate the probabilities for each age group, by transferring a continuous-time branching process from the context of demography to the context of epidemiology (Diekmann *et al.*, 1990), conditional on the pathogen being present in the animals and being able to transmit to humans with a frequency proportional to the frequency of touching or owning these animals. By way of example and without loss of generalisability we assume the basic reproduction number to equal 1.5 (Flahault *et al.*, 1988). This value implies a level of transmissibility of (seasonal) influenza in this case. We calculate the outbreak probability as the product of three quantities: (1) the observed proportion of participants within a certain age group, owning or touching pets, livestock or/and poultry; (2) the probability that an infection does not die out, given that it starts in a particular age class; and (3) the size of the age group proportional to the total population. We either base the calculation on age-heterogeneous mixing patterns or age-homogeneous mixing patterns.

Of the 1,768 participants, 59.4% and 49.4% owned and touched at least one animal, respectively (with 0.7% and 2.5% of missingness, respectively). The most reported animals are pets, 905 (51.2%) and 812 (45.9%) owns and touches them, respectively (Table 4.1). While many participants have chickens at home (14.9%), only few touch poultry and less than 5% own or touch livestock (Table 4.1). We found that animal ownership

is significantly associated with age and household sizes (Table 4.2), with a good fit to the observed data (Hosmer-Lemeshow test (Agresti, 2002): $p = 0.926$). We show that participants aged between 6 and 64 years are more likely to own animals compared to young participants aged 0-5 years (reference). The elderly (age above 65 years) and the reference group did not differ significantly. Additionally, for larger households, we find a higher probability of owning animals compared to single-person households (reference). We furthermore find a significant effect of age, animal ownership and their interaction with animal touching (Table 4.A.1) with a good fit to the observed data (Hosmer-Lemeshow: p=0.999). For those who own animals, participants aged 18-64 years are more likely to touch animals compared to children aged 0-5 years (reference), whereas there is no significant difference between the other age groups and the reference. As expected, the odds of touching animals is significantly higher for animal owners as compared to participants without animals in their household (“non-animal owners”).

Table 4.1: Number and proportion of participants who reported owning and touching pets (cats, dogs, fish), livestock (horses, sheep, cows or pigs), poultry (chicken, turkey, pigeons) and other animals* in Flanders, Belgium, 2010-2011 (n=1,768). Table from Kifle *et al.* (2016).

	Owning	**Touching**
Pets	905 (51.19%)	812 (45.93%)
Cats	503 (28.45%)	450 (25.45%)
Dogs	437 (24.72%)	491 (27.77%)
Fish	267 (15.05%)	6 (0.34%)
Livestock	82 (4.64%)	45 (2.55%)
Horses	53 (3.00%)	32 (1.81%)
Sheep	21 (1.19%)	6 (0.34%)
Cows	18 (1.02%)	10 (0.56%)
Pigs	9 (0.51%)	1 (0.06%)
Poultry	272 (15.38%)	38 (2.15%)
Chickens	264 (14.93%)	29 (1.64%)
Pigeons	23 (1.30%)	10 (0.56%)
Turkeys	5 (0.28%)	0 (0.00%)

* Out of total participants, 326 (18.44%) and 117 (6.62%) participants owned/touched other animals.

Table 4.2: Multiple-logistic regression model for animal ownership in Flanders, Belgium, 2010-2011 (n=1,756). Table from Kifle *et al.* (2016).

Category	**Sample size**	**Parameter Estimates (SE)**	**OR & 95% CI**	**P-value**
Age				<0.001
0-5 years*	174		1.00	
6-11 years	127	0.88 (0.27)	2.41 [1.43, 4.06]	
12-17 years	79	0.72 (0.30)	2.05 [1.13, 3.72]	
18-44 years	621	0.59 (0.18)	1.81 [1.27, 2.60]	
45-64 years	468	0.67 (0.20)	1.95 [1.31, 2.88]	
65+ years	287	-0.31 (0.32)	0.73 [0.39, 1.36]	
Household size				<0.001
1*	98		1.00	
2	312	0.78 (0.24)	2.19 [1.38, 3.48]	
3	328	1.23 (0.24)	3.44 [2.14, 5.52]	
4	439	1.44 (0.24)	4.21 [2.62, 6.74]	
5	218	1.48 (0.27)	4.41 [2.60, 7.48]	
Missing	361	0.58 (0.31)	1.78 [0.97, 3.27]	

* Reference Category.

We observe the mean number of contacts to be 13.5 with a variance of 11.8, suggesting overdispersion to be present. Our final model for the number of contacts includes interaction effects of age and weekday, province and gender, animal ownership and weekday and main effects only of household size and holiday period (Table 4.A.2). The overdispersion parameter is estimated at 2.43 (95% CI [2.24, 2.62]), indicating significant overdispersion. Participants older than five years report significantly more contacts during the weekend compared to children aged 0-5 years, while on a weekday participants older than 17 years report fewer of them compared to children aged 0-5 years. People living alone report significantly fewer contacts compared to people living with others. Overall, participants reported more social contacts on a weekday compared to a weekend day, and during a regular period compared to a holiday period. Animal owners have more social contacts than non-animal owners during the weekend, but this difference disappears on weekdays.

Figure 4.2 and 4.3 (Kifle *et al.*, 2016) show the probabilities of a human-transmissible pathogen to cause a large zoonotic outbreak in each of the age groups based on the age-heterogeneous social contact patterns and by assuming an R_0 of 1.5 to reflect an influenza-like pathogen. In the context of Flanders, the overall probability of a major outbreak is small, and pathogens originating from pets would pose a greater risk compared to poultry or livestock pathogens. Furthermore, if direct contact with live animals would be required, the relative risk posed by poultry or livestock pathogens is negligible, simply because far fewer people have direct interactions with such animals than with pets in Belgium (Figure 4.3). If ownership would be the primary driver (Figure 4.2), then poultry owners are more likely to be instrumental in causing a major outbreak than livestock owners. In general, elderly ($>$55 years) have the lowest probability to cause a large zoonotic outbreak, whereas younger adults (25-54 years) and children (0-9 years) are most likely to cause a zoonotic epidemic. Kifle *et al.* (2016) also report the analysis assuming homogeneous mixing patterns, which leads to similar results.

Nonetheless the probability for children (0-4 years) and adults older than 55 years to cause a major outbreak increases with homogeneous mixing as compared to heterogeneous mixing. The epidemic potential of children, teenagers and young adults (5-24 years) decreases under these conditions (Supplementary Material of Kifle *et al.*, 2016).

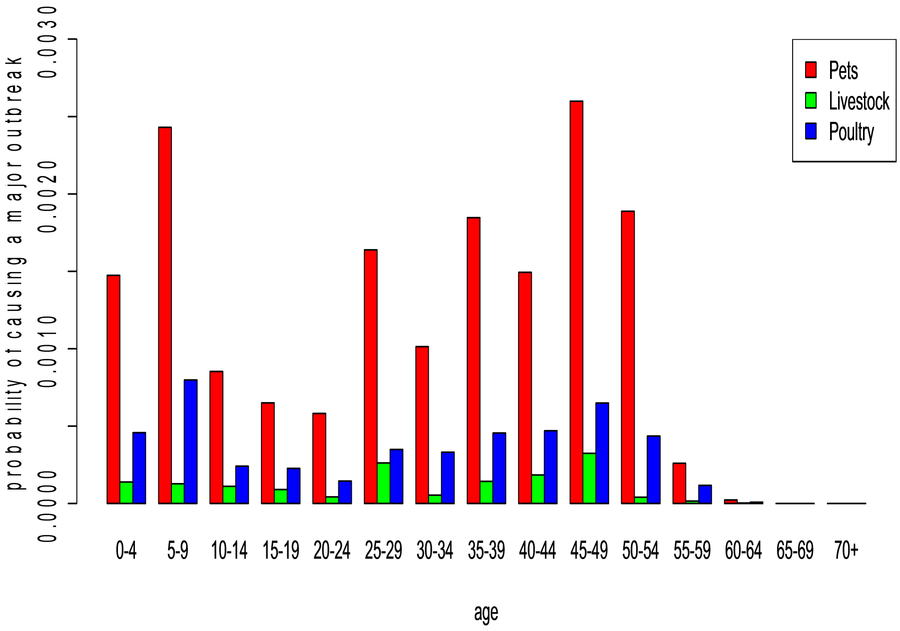


Figure 4.2: The probability of a human-transmissible pathogen causing a major outbreak in each age group, with owning of pets, livestock and poultry as exposure of pathogen spillover (assuming realistic - heterogeneously distributed - contacts by age). Figure from Kifle *et al.* (2016).

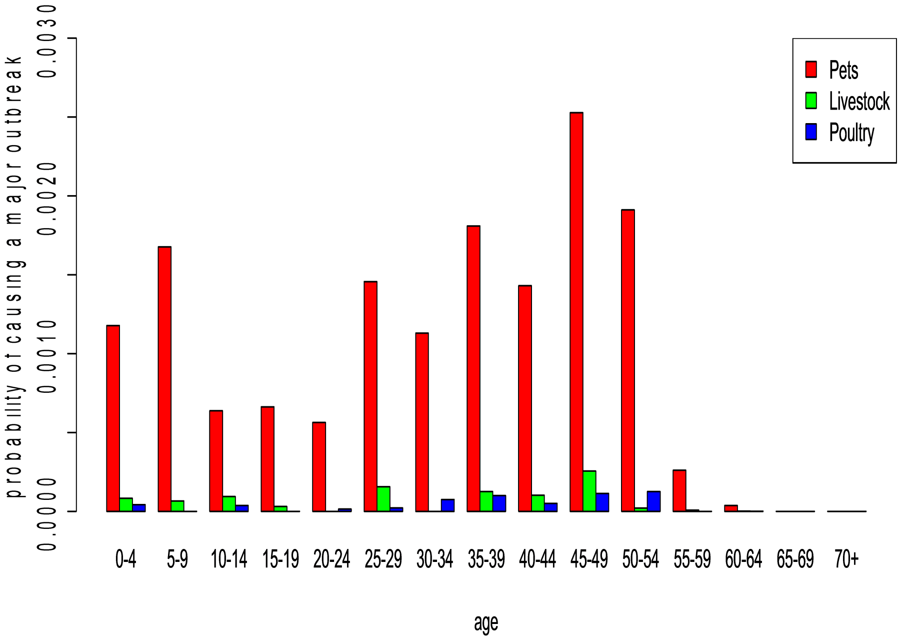


Figure 4.3: The probability of a human-transmissible pathogen causing a major outbreak in each age group, with touching of pets, livestock and poultry as exposure of pathogen spillover (assuming realistic - heterogeneously distributed - contacts by age). Figure from Kifle *et al.* (2016).

4.4 Household members do not contact each other at random: implications for infectious disease modelling

This section is based on work in preparation: "Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., Van Kerckhove, K., Willem, L., Aerts, M., Beutels, P., Hens, N. Household members do not contact each other at random: implications for infectious disease modelling. In preparation.".

The Flemish survey consisted, as mentioned before, of two parallel surveys of which one is the household one, querying every member of a household about his/her contact behaviour on one randomly assigned day. This day was the same for all members. Participants were requested to specify whether or not the contacted person belonged to their household. We link household contacts to other household members following these criteria: matching household identification number, gender and age (allowing the recorded age to deviate from the actual age by one year). As such, we link all contacts reported as household contacts to a single household member. Amongst the remaining contacts, i.e. with missing or negative household member indicator, which occurred at home, an additional small subset of household contacts is identified using the same criteria as before but requiring an exact age match. As such we determine the within-household contacts. Furthermore, we assume all contacts to be reciprocal. Hence each household can be depicted as an undirected network where nodes represent household members and the edges the contacts between household members. We assigned child, mother and father roles by using age, gender and household size. We use three measures to quantify within-household clustering: the clustering coefficient (Kolaczyk, 2009), the mean correlation coefficient (Morris *et al.*, 2008) and the proportion of observed versus potential triangles (see Supplementary Material of Goeyvaerts *et al.*, In preparation).

To model the within-household physical contact networks, we use Exponential-family Random Graph Models (ERGMs; Robins *et al.* (2007); Section 4.B.1). We infer on the process driving physical contacts between household members by incorporating network statistics based on nodal covariate information. The considered nodal covariates are mixing between siblings, among children and their parents and between partners, gender-preferential mixing and age effects in children, and the effect of household size, distinguishing small ($\leq$3 members), medium (4 members) and large ($\geq$5 members) households. We also explore higher-order dependency effects between household members by including in the model the number of isolated individuals, 2-stars, triangles and triangles in households of size ≥ 6. A 2-star is a person connected to two other household members, and a triangle is a set of three household members which are all linked to each other. Although we focus our analysis on within-household contact networks, we fit a single ERGM including all households. We include thus in our model a household effect (within-household edges) which captures the tendency to contact others in one's household. Given no between-household contact reports are present in the survey, the probability of contacts between households is essentially zero; thus the coefficients for the household preference effect should be estimated extremely large. We obtain approximate maximum likelihood estimates using a stochastic MCMC algorithm (Geyer and Thompson, 1992). We describe more details in Additional Details 4.B.1. We model the within-household physical networks separately for weekdays and weekends. We assess the goodness-of-fit of the models by simulating new sets of physical contact networks from the fitted ERGM and by comparing specific contact network characteristics, not included in the model, to the observed ones. These characteristics are the proportion of complete networks, the mean network density and the proportion of observed versus potential triangles, stratified by household size.

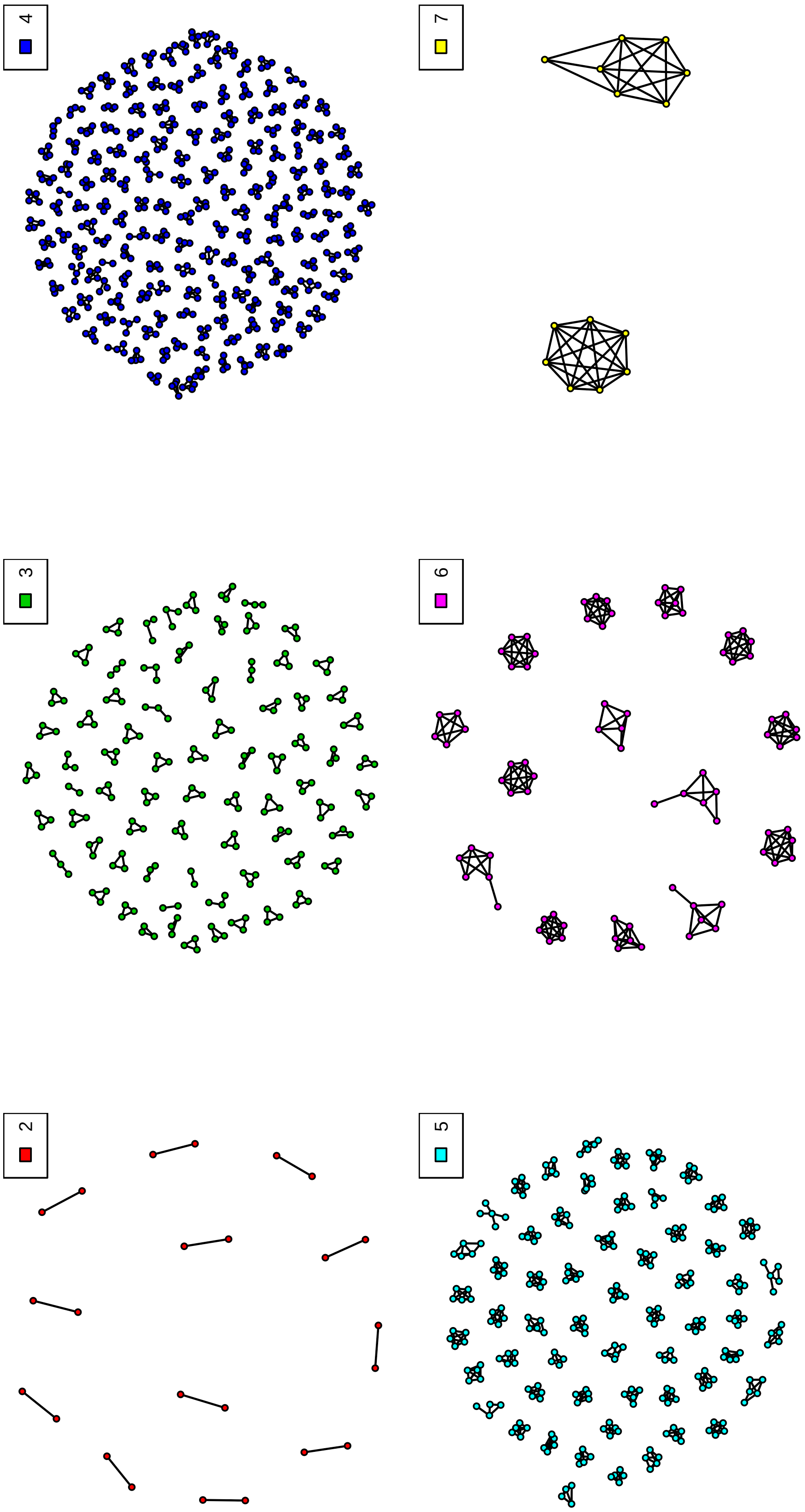


Figure 4.4: Observed within-household physical contact networks by household size (see color legend). Nodes represent household members and edges represent physical contacts. Figure adapted from Goeyvaerts *et al.* (In preparation).

We also simulate the spread of a newly emerging infection in a closed fully susceptible population of households by a discrete-time chain binomial SIR model (Bailey (1957); Section 4.B.2). Similar as in the households model of Ball *et al.* (1997), we assume two levels of mixing, high-intensity mixing within households and low-intensity mixing in the community, i.e. between households. We use the households that were utilised in the ERGM analysis to construct the community of households and compare two different figurations of within-household mixing: random mixing and empirical-based mixing. The last one employs the physical contact networks simulated from the fitted ERGMs; one from the weekday model and one from the weekend day model. These networks were kept fixed during the entire simulation.

As we aim to study the effect of contact heterogeneity, we make the following assumptions: susceptibility and infectiousness are age-invariant, no latent period is present, infected individuals recover with a constant probability of 0.22, hence a mean duration of infectiousness of 3.5 days. We employ transmission parameters based on literature estimates for influenza using household final size data and symptom onset data. We determine the first day of the epidemic randomly to be a week- or weekend day and start it by infecting three individuals at random. We track the spread until all infected individuals recover and no new infections occur.

We base our results on 318 households; we excluded 24 due to missing data from the set of 342 households, which encompass 1,266 participants who recorded 19,685 contacts in total with household sizes varying from 2 to 7. We identify and match the household members, which results in 3,821 within-household contacts which show 98% reciprocity, thus indicating good quality of reporting as expected in this household setting (Smieszek *et al.*, 2014). By assuming the contacts to be reciprocal, we find 1946 distinct within-household contacts of which 1,861 (96%) involved physical contact. We observe that contacts between household members are of long duration, consistent with findings from previous social contact surveys (Mossong *et al.*, 2008) and from individual-based simulation models creating so-called synthetic populations (Del Valle *et al.*, 2007). Furthermore, we observe that within-household contacts occurred (almost) daily and 66% of household members only met each other at home on the assigned day, while 33% met at multiple locations of which 98% included home. For the further analysis, we focus on physical contacts (with and without conversation) given that these are shown to better explain the observed age-specific seroprevalence of airborne infections compared to non-physical contacts (Ogunjimi *et al.*, 2009; Goeyvaerts *et al.*, 2010; Melegaro *et al.*, 2011). Figure 4.4 allows to appreciate at a glance the diversity in household size and network configurations we study through the survey.

We excluded two additional households from the data, given that we cannot assign the role of mother, father and child to the household with a grandparent and a same-sex parent household. Hence, we base the results on 316 households with 1,259 participants.

Table 4.3 summarises the proportion of complete (i.e. fully connected) networks and the mean network density for the within-household physical contact networks by household size, disentangled by week and weekend days and holiday and regular periods. The ratio of the number of observed edges to the number of potential edges defines the network density. Overall the type of day does not have a large impact on the within-household contacts, although the data suggest some decreasing connectedness with increasing household size, mainly on weekdays and during regular periods. The observed proportion of fully-connected networks is 0.77 on weekdays and 0.85 on weekend days for households of size 4. Goeyvaerts *et al.* (In preparation) further report a high degree of clustering of physical contacts within households.

The final ERGMs for weekdays and weekend days resulted in a negative estimate for the edge effect, which is there to counterbalance the large within-household edge effect as the data does not contain contacts between households. Table 4.4 describes the estimates as log odds ratios and hence, they need to be exponentiated to obtain odds ratios. We find a non-significant effect of the gender-preferential mixing and the number of isolates (LRT p=0.5766 for weekdays) for both types of days. For weekend days, we find no significant effect of household size, thus leading to an 8-parameter model (LRT p=0.5134). The network density for physical contacts decreases with increasing household size, as the odds of a physical contact occurring in a household of size ≤ 3 and ≥ 5 is estimated to be 2.10 and 0.67 times the odds of a physical contact in a household of size four on weekdays. Next, on both types of days, the probability for siblings to make physical contact decreases with increasing age. Furthermore, the odds of a physical contact between father and child is smaller than for any other pair except older siblings. For households of size ≤ 5, the odds of a physical contact that will complete a triangle is estimated to be 7.85 and 35.87 times the odds of a physical contact that will not complete a triangle on week and weekend days, respectively. These last results demonstrate the overall high degree of contact clustering within households. On weekdays, the degree of clustering is slightly lower in households of size ≥ 6 (conditional odds of 5.93). Overall, the final ERGMs fit the data well since the estimated network characteristics, which were not included in the model, are similar to the observed characteristics as can be seen from the figures in the Supplementary Materials of Goeyvaerts *et al.* (In preparation).

In a first scenario for the epidemic spread in a community of households, we find the proportion of small outbreaks to be significantly smaller in the random mixing setting compared to the empirical-based mixing, 0.43 and 0.50, respectively (Fisher's exact test, p=0.0027). The mean proportion of individuals ultimately infected and the mean proportion of households infected is slightly larger under random mixing compared to realistic mixing: 0.39 [0.12, 0.56] vs. 0.36 [0.12, 0.53], and 0.70 [0.28, 0.88] vs. 0.67 [0.29, 0.86], respectively. Furthermore, the household attack rate, defined as the mean proportion of individuals infected per household (Longini Jr and Koopman, 1982) increases with household size under both settings.

In an additional scenario, a scaling factor is included to account for different densities which could have lead to the small differences observed in the first scenario. We observe a similar relation between the attack rate and household size for both settings (week vs. weekend). Furthermore, there are barely any differences in mean final fraction of individuals (0.37 [0.13, 0.52] vs. 0.36 [0.12, 0.53]) and mean final fraction of households (0.68 [0.31, 0.86] vs. 0.67 [0.29, 0.86]). The proportion of small outbreaks is similar in both settings, 0.48 and 0.50, respectively (Fisher's exact test, p-value: 0.3954).

The preliminary results of other scenarios with different transmission rates, age-dependent household transmission rates, a larger community of households, or a partially vaccinated population are similar (not shown). Yet, when considering a more 'extreme' setting which focuses on physical contacts with a duration of more than 4 hours and which assumes higher within-household transmission rates, we obtain lower incidence for empirical-based mixing regardless of correcting for the within-household density (Supplementary Material of Goeyvaerts *et al.*, In

Table 4.3: Proportion of complete networks and mean network density, stratified by household size, for the observed within-household physical contact networks, comparing week and weekend days (top) and regular and holiday periods (bottom). Table from Goeyvaerts *et al.* (In preparation).

	Week			Weekend		
HH size	Nr. HHs	Proportion complete	Mean density	Nr. HHs	Proportion complete	Mean density
2	9	1.00	1.00	3	1.00	1.00
3	53	0.91	0.96	19	0.74	0.88
4	111	0.77	0.93	48	0.85	0.96
5	39	0.64	0.90	18	0.78	0.95
≥6	13	0.46	0.85	3	1.00	1.00
Total	225	0.77	0.93	91	0.82	0.94

	Regular period			Holiday period		
HH size	Nr. HHs	Proportion complete	Mean density	Nr. HHs	Proportion complete	Mean density
2	9	1.00	1.00	3	1.00	1.00
3	42	0.86	0.94	30	0.87	0.93
4	105	0.82	0.94	54	0.76	0.93
5	38	0.66	0.91	19	0.74	0.92
≥ 6	12	0.50	0.84	4	0.75	0.98
Total	206	0.79	0.93	110	0.79	0.93

Table 4.4: ERGM for within-household physical contact networks on week- and weekend days: parameter estimates and Wald test p-values, log-likelihood and AIC. Table from Goeyvaerts *et al.* (In preparation).

	Week		**Weekend**	
Network statistic	Estimate	p-value	Estimate	p-value
Edges	-28.16	<0.01	-20.63	<0.01
Within-household edges	28.97	<0.01	22.78	<0.01
Child-father mixing	-0.60	0.23	-1.15	0.45
Child-mother mixing	0.16	0.76	0.14	0.93
Father-mother mixing	0.27	0.66	-0.76	0.63
Age effect children	-0.07	<0.01	-0.18	<0.01
Small households (≤ 3)	0.74	<0.01		
Large households (≥5)	-0.40	<0.01		
2-stars	-0.26	0.25	-0.87	0.01
Triangles	2.06	<0.01	3.58	<0.01
Triangles in households ≥6	-0.28	0.02		
Log-likelihood	-306.80		-65.98	
AIC	635.59		147.95	

preparation).

4.5 Discussion & Conclusion

We obtained the results in this chapter using the Flemish survey (Section 2.4). The findings related to the number of social contacts are in agreement with published results as we have found more contacts on regular weekdays compared to weekend and holiday periods in Section 4.2 & 4.3 (Hens *et al.*, 2009a). Furthermore, the effect of age and household size on the total number of contacts are also in line with previous results (Hens *et al.*, 2009a), however, in this chapter we have found a significant effect of animal ownership, gender and provinces (Section 4.3). Hens *et al.* (2009a) found no significant difference between males and females, whereas here we have shown a significant gender effect, varying by the different provinces of Flanders.

In Section 4.2 we have observed an increased number of school contacts with high temperatures and low precipitation, and the latter is in line with a previous study in school children (Mikolajczyk *et al.*, 2008). Furthermore, the number of contacts longer than an hour increases on regular weekdays with low temperature and almost no precipitation. Next, we observed an increase in the long duration contacts when absolute humidity of the air was below the median value. We have also calculated R_0 ratios for different weather conditions to look at transmission dynamics and observed similar trends for long duration contacts on regular weekdays with low temperature, precipitation and humidity.

The relation between weather conditions and biological mechanisms of influenza transmission has been studied before, but the link with social contact patterns remained unclear (Lofgren *et al.*, 2007; Lowen *et al.*, 2007; Flasche *et al.*, 2011; Fuhrmann, 2010; Tamerius *et al.*, 2011; Steel *et al.*, 2011; van Noort *et al.*, 2012). We have presented evidence that weather may influence social mixing patterns. Although one may expect people to leave their home more often on a fine day, this does not necessarily imply that they will make more contacts. We did not find a relationship between the number of contacts and the weather, based on the present data, but the contact duration seemed to depend on the weather. One plausible explanation might be that people aggregate longer inside when the weather is bad, which time use studies (Mccurdy and Graham, 2003; Graham and Mccurdy, 2004) observed as well. Contacts of longer duration are of particular importance for infectious disease transmission since they tend to be more intensive and more often involve closer interactions (Smieszek, 2009). Fluctuations in social contact durations together with other seasonal adjustments might give influenza a greater opportunity to spread during wintertime (Fuhrmann, 2010).

In this chapter we have also shown that people interact more with pets (dogs, cats, fish) than livestock or poultry, as expected given Belgium is a HIC with little private agricultural activity. The most popular pets are dogs and cats. Ownership was slightly more (28.5% compared to 24.7%), and touching a little less frequent (25.5% in comparison to 27.8%) for cats compared to dogs. The important factors for owning animals are the household size and age. Participants aged 0-5 years are less likely to own animals compared to those older than six years, as are participants living alone compared to those living in larger households.

The results of Section 4.3 furthermore, suggest that the probability of touching an animal during a randomly assigned day depends on the main and interaction effects of age and animal ownership. People who did not own animals had a low chance to touch an animal on a random day. We furthermore observed that animal owners have more social contacts than non-animal owners during the weekend.

Previous research related to mathematical models of zoonotic outbreaks was lacking at the time we published our results (Kifle *et al.*, 2016), however, a simple stochastic model of directly transmitted zoonoses was used by Singh *et al.* (2013). We have demonstrated a method to characterise the relative probability of a major zoonotic outbreak from different animals, by relating human-animal interactions with social contact frequencies between humans in different age groups. We divided the study population by age and calculated the probability that a major outbreak would occur through animal contacts by age group. As a result, we have found that children (0-9 years) and adults (25-54 years) were more likely to cause a major zoonotic outbreak.

In this chapter, we furthermore presented results based on the first social contact study focusing specifically on contact networks within households. Previous studies based their inference on within-household contact networks on egocentric contact data from the POLYMOD Study (Mossong *et al.*, 2008; Section 2.2; Potter *et al.*, 2011; Potter and Hens, 2013) or on limited data in a very specific setting (rural Peru; Grijalva *et al.*, 2015). Estimates of the proportion of fully-connected networks inferred by Potter *et al.* (2011) and Potter and Hens (2013) ranged from 0.34 to 0.65 for households of size four, and are thus smaller than the proportions that we have observed (0.77 on weekdays and 0.85 on weekend days). Potter *et al.* (2011) and Potter and Hens (2013) obtained their estimates from partially observed within-household contact networks and, therefore, likely underestimated the actual proportion of fully-connected networks. For the purpose of studying household contacts, the current household-based survey design is considered an improvement upon the individual-based survey design (POLYMOD Study (Mossong *et al.*, 2008; Section 2.2)). We have analysed the household network data using ERGMs to assess the effect of factors such as role in the household, gender, children's age and household size on within-household physical contacts. We have found that contacts between father and children are less likely than contacts between father and mother, mother and children and between children (except older siblings). This discrepancy is in line with conclusions reported by de Greeff *et al.* (2012), who analysed data for pertussis in households with young infants and found that fathers were less likely to contract pertussis than other household members. Targeted vaccination of mothers and siblings was found to be most effective, as siblings are more likely to introduce an infection in the household. We have found that the contact density decreases and that the mean number of contacts increases with increasing household size (see Supplementary Material of Goeyvaerts *et al.*, In preparation). This implies that the mean contact degree is proportional to z^w where z is the household size and $0 < w < 1$. This result supplements findings from studies on household epidemic data of close-contact infections (Melegaro *et al.*, 2004; Cauchemez *et al.*, 2004; de Greeff *et al.*, 2012) from a social contact data perspective. To assess the standard assumption of random mixing within households, we simulated epidemics in a two-level SIR setting based on either the empirically grounded networks or random mixing using literature-based influenza parameters. We found that solely incorporating contact heterogeneity has no impact on epidemic spread. This indicates that the assumption of random mixing between household members may be an adequate approximation of social contact behaviour for infections transmitted via close contacts. However, the results do suggest that accounting for the within-household contact density is of importance. Furthermore, we found that in a more extreme setting with intenser contacts and a higher within-household transmission rate a density correction is not sufficient to bridge the differences between both mixing assumptions. This suggests that informing mixing between household members with social contact data could impact modelling efforts in certain settings

Albeit the innovative analysis and results presented in this chapter, we made different assumptions throughout the various analysis. As such in Section 4.2 we assumed constant transmission parameters for different weather conditions to be able to focus on the effect of modified social contact behaviour. Notwithstanding, temperature and humidity are likely to have some impact on virus transmission and host immunity (Lowen *et al.*, 2007; Steel *et al.*, 2011; Shaman and Kohn, 2009). Especially low absolute humidity is believed to be beneficial for influenza virus survival and transmission (Shaman and Kohn, 2009). Two key assumptions made in Section 4.3 to estimate the probability of zoonotic outbreaks are that the probability is conditional on the pathogen's presence in the animal group and that pathogen spillover from animals to humans was driven by animal ownership and touching. When inferring the within-household networks, we assumed that a contact occurred if at least one household member reports it. As a result, contacts forgotten by both members could lead to an underestimation of the network density. Potter *et al.* (2015) developed a model to deal with the issue of reporting errors on network edges. However, given that the high reciprocity rates (98%) indicate an excellent reporting quality of the survey, we believe such an adjustment will not have an impact on our conclusions.

Additionally, as with every study, some limitations are present. Proper estimation of the impact of weather conditions on social contact patterns requires considering the effect of weekend days and holiday periods. However, we experienced difficulties related to data sparseness when partitioning the data by day-type. The dataset was designed to estimate the effect of day-types but not for an extra partitioning based on weather types. Each subpopulation has to be representative of the general population, which was not the case for weekend and holiday periods after partitioning for a weather condition. We used weights to account for different age distributions, but this was insufficient if the exposure to a particular type of weather on a particular day-type occurred for only five people of a single age group. The sample was only substantial on regular weekdays so that our results are mainly limited to this day type. The number of unique weather conditions in our dataset was limited since the majority of the participants were recruited at the end of October and during November. A limitation when studying the human-animal interactions was the focus on ownership and direct touching of animals as potential risk factors for infectious disease transmission. In this survey, there was no attempt to study potential exposure to animal excrements (e.g. due to gardening or vicinity of forests) or raw meat consumption. Nevertheless, the work in Section 4.3 is the first that analysis human-animal interactions in the context of inter-human contacts. Although the results apply to a high-income country setting, one can also use the survey and analytical methods in low- and middle-income countries. The amount of missing data was limited, and therefore complete case analysis was used. In the household networks context, the results depend on the contact definition used to determine the within-household network links and cannot be generalised to the spread of any infectious disease. Based on the work of Goeyvaerts *et al.* (2010), we opted to use physical contacts in our study as a surrogate for potential transmission events for close-contact infections. Though, even for two airborne diseases, different networks may be appropriate because differing levels of interaction are required to constitute an effective contact (Keeling and Eames, 2005).

The results presented in this chapter open many future avenues. Given the data sparseness present for the weather analysis, careful considerations related to this aspect in a future study design is necessary. Since weather conditions cannot be predetermined, we recommend sampling on many different days to get a broad spectrum of weather conditions. Also, a sufficiently large number of respondents should participate on weekends and holiday periods to gain predictive power for these periods. Additional data on the location of the contacts - indoors or outdoors - can also lead to new insights in this challenging research field. Although we found associations between the weather and contact duration, this is no guarantee for a clear-cut causal relation with influenza seasonality. More studies are needed assessing the role of climatic conditions and contact patterns on observed epidemiological transmission patterns to provide adequate information to plan and evaluate mitigating strategies (Shaman *et al.*, 2010; Chao *et al.*, 2010a). One needs to interpret the results on human-animal contacts, and inter-human contacts in the context of a densely populated HIC in Europe. Similar studies in both similar and different settings are needed to confirm these findings and test the presented methods. The methods presented concerning within-household contacts, on the other hand, could be extended in various ways. We observed a relationship-specific heterogeneity in the duration of contacts, which might be relevant for some diseases. The ERGM framework can be adapted to model 'valued' within-household contact networks (Krivitsky, 2012), with the value of a link determined by its total duration, and by weighting the transmission rates in the epidemic simulation model accordingly. It is also of potential interest to capture temporal dynamics of within-household contacts and to simulate the impact of contact formation and dissolution on the spread of infections (Hanneke *et al.*, 2010; Krivitsky and Handcock, 2014). Combining time use like data with social contact data would allow to infer the possible timing of (physical) contacts with household members, and to estimate dynamic within-household contact networks. This time aspect would also be valuable to inform large-scale individual-based simulation models of infectious disease spread. Finally, combining the model for within-household contact networks developed in the work by Goeyvaerts *et al.* (In preparation) with epidemic data from a similar community of households would allow improving estimates of age-specific heterogeneity in susceptibility and infectiousness for infections such as influenza (Addy *et al.*, 1991).

In conclusion, this chapter discussed different research topics, each innovative on its own. The analysis of weather conditions and social contact behaviour has not been investigated previously. Furthermore, it allows for possible elements for future studies. When considering animal-human interactions, we proposed and applied a new approach to estimate the probability of a zoonotic outbreak. Lastly, the household study provides unique insights into within-household contacts, which are considered to be important drivers of many close-contact infections. It is the first empirical evidence resulting from a large household contact survey supporting the use of the random mixing assumption in epidemic models incorporating household structure.

4.A Additional details for animal ownership and touching

Table 4.A.1: Multiple-logistic regression model for animal touching in Flanders, Belgium, 2010-2011 (n=1,722). Table from Kifle *et al.* (2016).

Category	Sample size	Parameter Estimates (SE)	OR & 95% CI	P-value
Age				<0.001
0-5 years*	170		1.00	
6-11 years	125	-0.21 (0.30)	0.81 [0.45, 1.45]	
12-17 years	79	0.18 (0.35)	1.19 [0.60, 2.38]	
18-44 years	615	0.76 (0.25)	2.15 [1.32, 3.47]	
45-64 years	456	0.90 (0.26)	2.45 [1.47, 4.10]	
65+ years	277	0.03 (0.31)	1.03 [0.56, 1.88]	
Animal ownership				<0.001
Owner*	1037		1.00	
Not owner	685	-2.71 (0.43)	0.07 [0.03, 0.15]	
Age: Animal Ownership				0.038
6-11 years: Not owner	28	0.75 (0.69)	2.11 [0.55, 8.11]	
12-17 years: Not owner	20	0.79 (0.73)	2.20 [0.53, 9.21]	
18-44 years: Not owner	212	-0.39 (0.49)	0.68 [0.26, 1.76]	
45-64 years: Not owner	166	-0.94 (0.52)	0.39 [0.14, 1.09]	
65+ years: Not owner	188	-0.34 (0.55)	0.71 [0.24, 2.10]	

* Reference Category.

Table 4.A.2: Weighted-negative binomial regression for the total number of contacts in Flanders, Belgium, 2010-2011 (n=1,742). Table from Kifle *et al.* (2016).

Covariate	Nb. Part.‡	Median (IQR‡)	Param. Est.‡ (SE)	RNC‡ [95% CI]	P-value
Age					0.002
0-5 years*	174	10.00 (11.00)		1.00	
6-11 years	127	18.00 (20.00)	0.56 (0.18)	1.76 [1.24, 2.51]	
12-17 years	79	15.00 (19.50)	0.47 (0.18)	1.61 [1.13, 2.30]	
18-44 years	621	11.00 (11.00)	0.58 (0.14)	1.79 [1.37, 2.33]	
45-64 years	466	11.00 (10.00)	0.59 (0.14)	1.80 [1.35, 2.37]	
65+ years	275	7.00 (9.50)	0.68 (0.18)	1.98 [1.38, 2.84]	
Household size					<0.001
1*	98	8.00 (10.00)		1.00	
2	312	10.00 (10.00)	0.15 (0.08)	1.17 [1.00, 1.35]	
3	328	11.00 (11.00)	0.21 (0.08)	1.24 [1.06, 1.44]	
4	439	13.00 (12.00)	0.29 (0.08)	1.34 [1.16, 1.56]	
5	218	14.00 (15.00)	0.39 (0.08)	1.48 [1.26, 1.74]	
Missing	347	7.00 (10.00)	0.03 (0.10)	1.04 [0.84, 1.27]	
Animal ownership					0.002
Owner*	1,042	11.00 (12.00)		1.00	
Not owner	700	10.00 (11.00)	-0.24 (0.07)	0.79 [0.68, 0.91]	
Gender					0.483
Female*	930	11.00 (12.00)		1.00	
Male	812	11.00 (11.00)	0.05 (0.06)	1.05 [0.92, 1.19]	
Province†					0.024
Antwerp*	487	11.00 (11.00)		1.00	
Limburg	264	11.00 (13.25)	0.18 (0.07)	1.20 [1.04, 1.39]	
East Flanders	407	10.00 (11.00)	0.05 (0.06)	1.05 [0.93, 1.19]	
Flemish Brabant	257	10.00 (13.00)	0.20 (0.07)	1.22 [1.05, 1.41]	
West Flanders	327	11.00 (10.00)	0.14 (0.07)	1.15 [1.00, 1.32]	
Weekday indicator					<0.001
Weekend*	419	10.00 (10.00)		1.00	
Weekday	1,323	11.00 (12.00)	0.62 (0.15)	1.85 [1.37, 2.48]	
Holiday indicator					0.007
Regular period*	1,632	11.00 (12.00)		1.00	
Holiday period	110	8.00 (9.00)	-0.20 (0.07)	0.82 [0.71, 0.94]	
Age: Weekday indicator					<0.001
6-11 years: Weekday	96	20.00 (23.00)	-0.28 (0.21)	0.75 [0.50, 1.13]	
12-17 years: Weekday	58	18.50 (20.00)	-0.20 (0.21)	0.82 [0.54, 1.24]	
18-44 years: Weekday	466	12.00 (11.00)	-0.69 (0.16)	0.50 [0.37, 0.69]	

45-64 years: Weekday	358	11.00 (11.00)	-0.76 (0.16)	0.47 [0.34, 0.65]	
65+ years: Weekday	221	7.00 (9.00)	-1.13 (0.18)	0.32 [0.22, 0.46]	
Province: Gender					0.020
Limburg: Male	124	10.00 (10.75)	0.04 (0.11)	1.04 [0.84, 1.29]	
East Flanders: Male	176	11.00 (11.50)	-0.02 (0.10)	0.98 [0.81, 1.18]	
Flemish-Brabant: Male	121	14.00 (13.00)	-0.32 (0.11)	0.72 [0.58, 0.90]	
West Flanders: Male	171	15.00 (14.00)	-0.09 (0.10)	0.91 [0.75, 1.11]	
Animal ownership: Weekday indicator					0.003
Not an owner: Weekday	530	10.50 (11.00)	0.26 (0.08)	1.30 [1.09, 1.53]	

[*] Reference Category.

[†] We exclude 14 observations with missing provinces because of not being enough for estimation.

[‡] Nb. Part. = Number of Participants (Sample size), Param. Est. = Parameter Estimates, IQR = Inter-quartile range, RNC = Relative Number of Contact.

4.B Additional details for within-household networks

4.B.1 Exponential-family Random Graph Models (ERGMs)

Let $\mathbf{Y}$ be the random adjacency matrix of an undirected network, with $Y_{ij} = Y_{ji} = 1$ if person i and j make physical contact and zero otherwise, and let Ω be the support of $\mathbf{Y}$, i.e. the set of all obtainable networks. In an ERGM the probability of observing a set of network edges is defined as follows:

$$P_{\boldsymbol{\theta},\Omega}(\mathbf{Y} = \mathbf{y}) = \frac{\exp\{\boldsymbol{\theta}'\mathbf{g}(\mathbf{y},\mathbf{X})\}}{\kappa(\boldsymbol{\theta},\Omega)}, \ \mathbf{y} \in \Omega, \tag{4.B.1}$$

where $\mathbf{g}(\mathbf{y},\mathbf{X})$ is a vector of network statistics that may depend on additional covariate information $\mathbf{X}$, with $\boldsymbol{\theta}$ the corresponding vector of coefficients, and $\kappa(\boldsymbol{\theta},\Omega)$ a normalising factor. $\boldsymbol{\theta}$ can be interpreted as the increase in the conditional log-odds of the network, per unit increase in the corresponding component of $\mathbf{g}(\mathbf{y},\mathbf{X})$, resulting from switching a particular dyad Y_{ij} from 0 to 1 while leaving the rest of the network fixed at $\mathbf{Y}^c_{ij}$. This interpretation is clear from the alternative specification on Hunter *et al.* (2008), which define the vector of change statistics for a specific pair of nodes (i, j) as:

$$\boldsymbol{\delta}_{\boldsymbol{g}}(\mathbf{y},\mathbf{X})_{ij} = \mathbf{g}(\mathbf{y}^+_{ij},\mathbf{X}) - \mathbf{g}(\mathbf{y}^-_{ij},\mathbf{X}), \tag{4.B.2}$$

where $\mathbf{y}^+_{ij}$ and $\mathbf{y}^-_{ij}$ are the networks realised by fitting $y_{ij} = 1$ and $y_{ij} = 0$ respectively, while leaving all the rest of $\mathbf{y}$ fixed. Hence, we can use a logistic interpretation of the coefficients in $\boldsymbol{\theta}$:

$$\text{logit}\{P_{\boldsymbol{\theta},\Omega}(Y_{ij} = 1|\mathbf{Y}^c_{ij} = \mathbf{y}^c_{ij})\} = \boldsymbol{\theta}'\boldsymbol{\delta}_{\boldsymbol{g}}(\mathbf{y},\mathbf{X})_{ij}, \tag{4.B.3}$$

where $\mathbf{Y}^c_{ij}$ represents the rest of the network other than Y_{ij}.

Approximate maximum likelihood estimates are obtained using a stochastic MCMC algorithm (Geyer and Thompson, 1992). In short, a distribution of random networks is simulated from a starting set of parameter values using MCMC and the parameter values are refined by comparing this distribution of networks against the observed network in a Newton-Raphson type algorithm, repeating this process until the parameter estimates stabilise (Robins *et al.*, 2007). We use a burn-in of length 10^6, intervals between sampled networks of length 10^3 and a total sample

Table 4.B.4: Network statistics considered in the ERGMs, where an edge is defined as a physical contact between two individuals. Reference categories are child-child mixing, boy-girl mixing, and mixing within households of size 4. Table from Goeyvaerts *et al.* (In preparation).

Network statistic	**Legend**
Edges	Total number of edges
Within-household edges	Total number of edges within households
Child-father mixing	Total number of edges between children and fathers
Child-mother mixing	Total number of edges between children and mothers
Father-mother mixing	Total number of edges between partners
Boy-boy mixing	Total number of edges between male children
Girl-girl mixing	Total number of edges between female children
Age effect children	The sum of age(i) and age(j) for all edges (i, j) between siblings
Small ($\leq$3) households	Total number of edges within households of size ≤ 3
Large ($\geq$5) households	Total number of edges within households of size ≥ 5
Isolates	Total number of isolates
2-stars	Total number of 2-stars
Triangles	Total number of triangles
Triangles in households of size ≥ 6	Total number of triangles in households of size ≥ 6

size equal to 5×10^5. The initial value of $\boldsymbol{\theta}$ is obtained by maximum pseudolikelihood estimation, considering Eq. (4.B.3) as a logistic regression model assuming all Y_{ij} mutually independent (Strauss and Ikeda, 1990). The MCMC estimation is performed with the "ergm" package in R (Hunter *et al.*, 2008; Handcock *et al.*, 2003) that is part of the "statnet" suite of packages for statistical network analysis (Handcock *et al.*, 2008, 2014; Goodreau *et al.*, 2008).

4.B.2 Discrete-time chain binomial SIR model

At time step t (in days), assuming infection is spread through physical contacts, each susceptible i acquires infection with probability:

$$p_{i,1}(t) = 1 - (1-\beta_h)^{\sum_{j\neq i\in h_i} y_{ij} I_j(t)} (1-\beta_{c,11})^{\sum_{j\notin h_i} I_{j,1}(t)} (1-\beta_{c,12})^{\sum_{j\notin h_i} I_{j,2}(t)}, \tag{4.B.4}$$

$$p_{i,2}(t) = 1 - (1-\beta_h)^{\sum_{j\neq i\in h_i} y_{ij} I_j(t)} (1-\beta_{c,21})^{\sum_{j\notin h_i} I_{j,1}(t)} (1-\beta_{c,22})^{\sum_{j\notin h_i} I_{j,2}(t)}, \tag{4.B.5}$$

where index 1 corresponds to children ≤ 18 years and index 2 to adults > 18 years. β_h denotes the within-household transmission probability for physical contact, per time step. The 2×2 community transmission matrix $\boldsymbol{\beta}_c$ (with $\beta_{c,12} = \beta_{c,21}$ is taken directly proportional to the per capita physical contact rates estimated from the Belgian POLYMOD contact survey, with a proportionality constant q_c (Mossong *et al.*, 2008; Goeyvaerts *et al.*, 2010).

$$\boldsymbol{\beta}_c = q_c \begin{bmatrix} 17.35 & 6.26 \\ 6.26 & 7.88 \end{bmatrix} 10^{-7}. \tag{4.B.6}$$

Further, y_{ij} denotes the observed adjacency matrix and under the random mixing scenario, y_{ij} equals 1 for all household members i and j. Finally, h_i denotes the household of node i and $I_j(t)$ indicates whether node j is infected (1) or not (0) at time t with subscripts referring to children and adults.

Literature estimates are used to choose β_h and q_c: $\beta_h = 0.05$ and $\bar{\beta}_c = 0.00026$. These parameter values result in estimates of the community probability of infection for children and adults ($\overline{\text{CPI}}_{\text{child}}$ and $\overline{\text{CPI}}_{\text{adult}}$) between 0.18 and 0.20, and 0.11 and 0.12, respectively. Furthermore, $q_{HH} = 1 - \beta_h = 0.95$ is the probability to escape infection from an infected household member per day.

A second scenario is considered, which includes a scaling factor to account for the differences in density between empirical-based and random mixing:

$$p_{i,1}(t) = 1 - (1 - \beta_h \delta_h)^{\sum_{j \neq i \in h_i} y_{ij} I_j(t)} (1 - \beta_{c,11})^{\sum_{j \notin h_i} I_{j,1}(t)} (1 - \beta_{c,12})^{\sum_{j \notin h_i} I_{j,2}(t)}, \tag{4.B.7}$$

$$p_{i,2}(t) = 1 - (1 - \beta_h \delta_h)^{\sum_{j \neq i \in h_i} y_{ij} I_j(t)} (1 - \beta_{c,21})^{\sum_{j \notin h_i} I_{j,1}(t)} (1 - \beta_{c,22})^{\sum_{j \notin h_i} I_{j,2}(t)}. \tag{4.B.8}$$

Hence, δ_h is chosen 1 for empirical-based mixing, while for random mixing it equals the network density of the simulated contact network in the realistic mixing scenario.

Chapter 5

The Interplay Between Time Use, Mobility and Social Contact Behaviour

In this chapter we begin with a study describing the dispersion of social contacts by making use of time use data combined with social contact data. Next, we combine mobility and social contact data in a metapopulation model which aims to describe the transmission dynamics of seasonal influenza in the 2008/2009 season.

5.1 The dispersion of social contacts and their relevance for the construction of epidemic models

This section is based on work in preparation: "Van Kerckhove, K., Willem, L., Faes, C., Beutels, P., Hens, N. The dispersion of social contacts and their relevance for the construction of epidemic models. In preparation.".

5.1.1 Introduction

In the last decade, Held *et al.* (2005, 2006); Paul *et al.* (2008); Paul and Held (2011) and Held and Paul (2012) built and extended a regression-oriented, endemic-epidemic time-series model for spatial-temporal count data of infectious diseases. More recently, Geilhufe *et al.* (2014) integrated movement data, through weights, into this model to capture the spatial spread of influenza A in North Norway. Weights based on a power law approximation of traffic data led to a better fit compared to the weights based on the raw traffic data. By running analysis separately for children and adults, travel data were shown to be only a rough proxy for children. Based on these findings, the authors suggest the use of social contact data as a possible improvement.

In the same year, Meyer and Held (2014) published their results on two different spatio-temporal models with various power law extensions for the spatial component. They applied the models to individual cases of invasive meningococcal disease in Germany and to count data on influenza in Southern Germany. In both cases, the models with the power law extensions outperformed the previously used formulations.

Read *et al.* (2014) observed, based on a social contact survey in China, that contacts occur mainly close to home, and only some occur at a distance further than 500 km from home. Driven by these findings, Meyer and Held (2017) adapted their model for stratified, area-level infectious disease counts to include information on social contact data. Meyer and Held (2017) applied this modified model to norovirus gastroenteritis in Berlin's city districts and found an improvement of the model when incorporating social contact data. However an age-dependent decay could not be identified, and thus the authors suggest to "replace the parametric form by a social contact matrix, stratified by spatial distance in addition to age groups". Based on the work of Kucharski *et al.* (2015), Meyer and Held (2017) furthermore suggest to use separate movement information for children and adults to quantify the strength of epidemiological coupling between regions.

A common finding of Geilhufe *et al.* (2014), Meyer and Held (2014) and Meyer and Held (2017) is a power law dispersal when modelling disease counts. Since transmission can be related to social contact behaviour through the social contact hypothesis (Wallinga *et al.*, 2006), one can question whether this power law behaviour can be observed in social contact dispersal as well or whether transmission and social contact dispersal differ.

A systematic review of social contacts surveys by Hoang *et al.* (In preparation) reveals that only a limited number of contact studies investigated the relationship between contacts and distance. Danon *et al.* (2012) collected information on the distance from home for each contact and examined the degree distribution in a network setting. Based on the same survey, Danon *et al.* (2013) showed results on the total time spent with contacts by the distance from home. They observed a decrease in contact duration with increasing distance from home. Furthermore, they showed a relation between clustering and distance from home, with high clustering within two miles, dominated by home contacts, but the highest value of clustering occurring at a distance 50 miles or more from home. The authors hypothesise that this might be due to differences in the purpose behind contacts made at various distances. Fu *et al.* (2012) performed a large-scale study in Taiwan, whereby the distance from home seems requested for each contact, but the authors did not report on it unambiguously. They report that 52.7% of contacts took place at a distance less than 1 km from home, 29.2% at a distance between 1 and 9 km from home, 14.6% at a distance 10-49 km from home and the rest (3.5%) at a distance beyond 50 km from home. Read *et al.* (2014) were able to quantify distances from home based on the recording of the location (latitude and longitude) of each reported contact. Based on this information they observe an increase in age-assortativity when contacts were made further from home. Stein *et al.* (2015) deduced the distance between recruiters and their recruited participants, van Hoek *et al.* (2013) requested for infants to report the maximum distance travelled from home, but, to date, did not report results thereof. Others, amongst whom Melegaro *et al.* (2017) asked for the distance between home and work but thus far did not report results related to this information.

Based on the Flemish social contact survey (Section 2.4), we present results on the relationship between contacts and distance. This study is, to the best of our knowledge, the first to estimate the social contact dispersal over distance. Although Kucharski *et al.* (2015) and Meyer and Held (2017) recently indicated the necessity for collecting such data; we ran this social contact survey earlier. As indicated previously, this survey recorded time use of the participants and their contact behaviour. In the time use, the distance from home, in categories, for visited locations is requested, allowing to link contacts to the most probable distance from home.

Our aim is twofold: (1) to test various underlying continuous degree densities over distance, amongst which power law approximations as these seem most predictive in transmission models. (2) To estimate social contact matrices over distance, given their need in models for infectious diseases as indicated by Meyer and Held (2017). We discuss the implications for the spread of infectious diseases when jointly considering spatial behaviour and social contacts, using the force of infection over distance.

In Section 5.1.2.1 we describe the data and explain the process to link the contacts to a distance. In Section 5.1.2.2, we introduce the methodology for the estimation of the continuous underlying degree distribution over distance; we present the notation for the social contact matrices over a distance in Section 5.1.2.3. Section 5.1.2.4 ends the methods section with the details about the infection pressure over distance.

5.1.2 Methods

5.1.2.1 Data

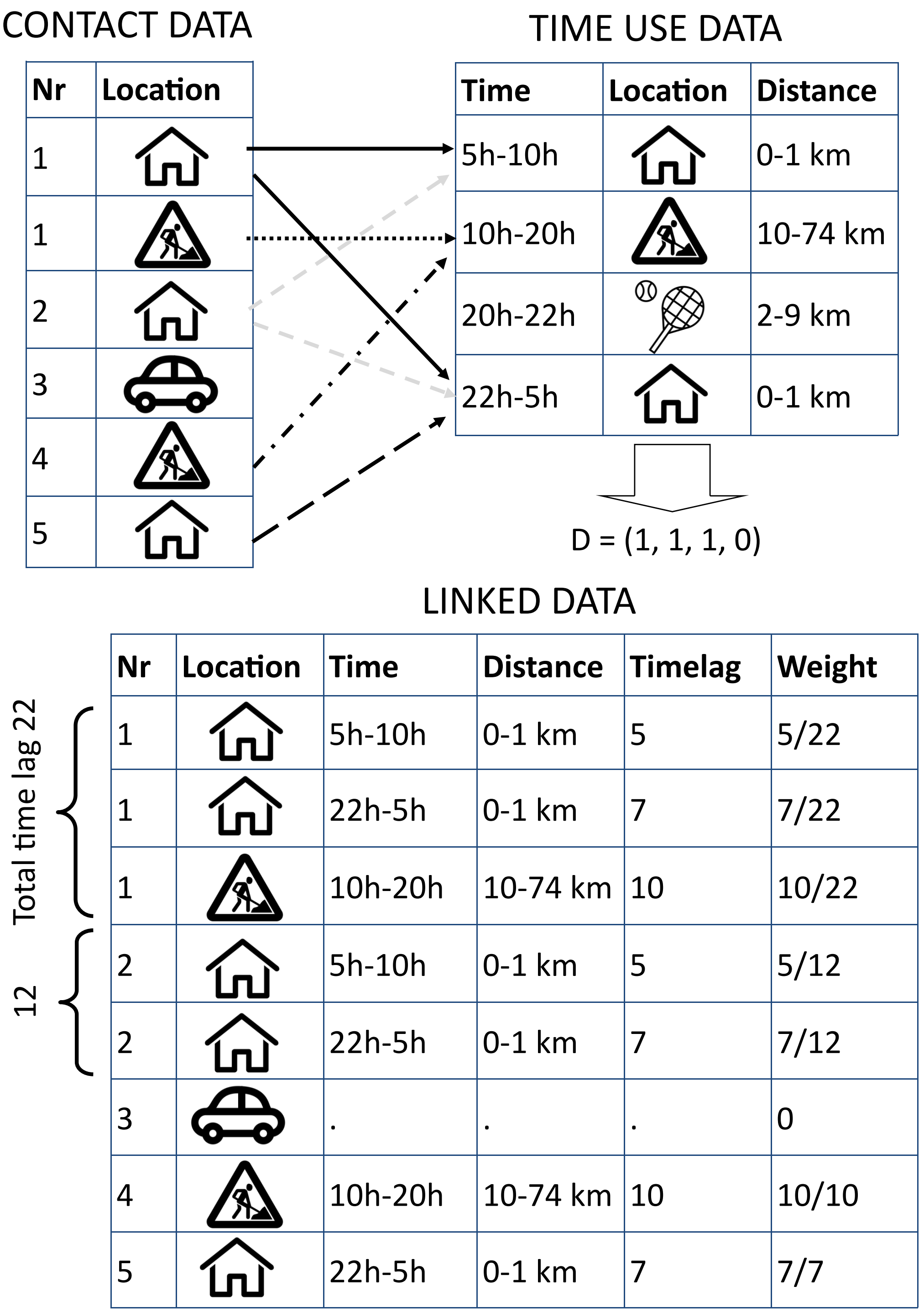
As discussed before, we use data from the Flemish social contact survey (Section 2.4). In brief: elements requested about contacts were the age or estimated age of the contact, the gender, whether touching was involved, the frequency of meeting this contact (e.g. daily - the first time) and whether the contacted person was part of the household.

In the time use part, for pre-specified time blocks participants indicated the one location at which they spent most of their time and the distance from home for this location (0-1 km, 2-9 km, 10-74 km or 75 km or more). The time blocks were of 1-hour length, except in the morning (2-5h, 5-8h) and in the evening (between 20h and 2h, 2 hours time blocks).

In the contact diary, repeated contacts per day were asked to be reported only once at first occurrence and in a chronological order allowing for specifying multiple locations. It was, therefore, possible to probabilistically link those contacts to the time use survey. Linking contacts to time use results in connecting contacts to potentially multiple and different distances. We call a subcontact a part of the contact with one particular location and one linked distance. To avoid an artificial increase in the total number of contacts, we give standardised weights (weights sum up to one per contacted individual) for subcontacts. In case a participant does not report any contacts, the weight equals 1, to ensure a 0 count contribution of the participant. Figure 5.1 shows a visual example of this process. We give a more detailed description of this process in Section 5.A.1.

The time use data allows for describing the travel of participants across distances, as can be seen from Figure 5.1. These patterns are summarised in a vector $\mathbf{D}$ for each participant, which has 1 for each distance visited and 0 otherwise (more rigorous description in Section 5.1.2.2). Table 5.1 shows the prominent (present for more than 2% of the participants) patterns for those participants for which the linking was feasible. Table 5.A.1 is a similar table for all possible participants.

Based on the linked data, and by retaining the contacts with a non-missing distance, we describe the average number of contacts over distance. Figure 5.2 shows the average for all participants (full line) combined with different situations of week/weekend or regular/holiday periods, by creating lines between the observed points. Discrepancies are apparent and can also be seen when stratifying by age (Figure 5.A.1).



Nr	Location
1	🏠
1	🚧
2	🏠
3	🚗
4	🚧
5	🏠

Time	Location	Distance
5h-10h	🏠	0-1 km
10h-20h	🚧	10-74 km
20h-22h	🎾	2-9 km
22h-5h	🏠	0-1 km

Nr	Location	Time	Distance	Timelag	Weight
1	🏠	5h-10h	0-1 km	5	5/22
1	🏠	22h-5h	0-1 km	7	7/22
1	🚧	10h-20h	10-74 km	10	10/22
2	🏠	5h-10h	0-1 km	5	5/12
2	🏠	22h-5h	0-1 km	7	7/12
3	🚗	.	.	.	0
4	🚧	10h-20h	10-74 km	10	10/10
5	🏠	22h-5h	0-1 km	7	7/7

Figure 5.1: A visual example of the process linking the contact data (upper left) to the time use data (top right) to achieve the linked contact data (bottom part). Based on the time use data, we deduce a vector **D**, which is 1 for each of the distance categories visited and 0 otherwise. The arrows between the contact data and the time use data indicate the possible time slots the contact could have occurred. Whenever no link was possible (as for travel), the time, distance and other information are left blank in the linked data, and we set the weight to 0. In case a location occurs multiple times in the time use data, they can contribute multiple times. However, when it is evident from the contact list that the contact took place at a precise location (when contacts are present at this single location), earlier time blocks of a location will not be considered for the linking. As an example, contact 4 occurs at work. Thus, contact 5 at home can not be linked to the first time block at home since a work contact took place in between. (Icons from the Noun Project: Home from il Capitano, Work from Gleb Khorunzhiy, Travel from Ralf Schmitzer and Leisure from Larisa Skosyrska)

Table 5.1: Patterns travelled over distance based on time use data of participants with proper linking (n=1560). The row frequencies and percentages (total and split by age) show the number and proportion of people exhibiting a particular travel behaviour over distance throughout the day. We merged travel patterns for which the relative frequency is smaller than 2% into 'Other'. The column frequencies show the unique number of people travelling to a particular distance; the last two rows give them divided over two age classes: 0-18 years of age and 19 years of age and older.

	0-2 km	2-10 km	10-75 km	over 75 km	**Frequency**			**Perc. (%)**
					Total	age $[0,18]$	age $(18,100]$	
	1	0	0	0	324	100	224	20.77
	1	0	1	0	269	41	228	17.24
	1	1	0	0	559	161	398	35.83
	1	1	1	0	320	42	278	20.51
Other	.	.	.	.	88	18	70	5.64
Freq.	1544	904	635	83	1560	100		
Perc. (%)	76.99	57.95	40.71	5.32	100			
Freq. age $[0,18]$	358	206	90	18	362			
Perc. (%) age $[0,18]$	83.98	56.91	24.86	4.97	100			
Freq. age $(18,100]$	1186	698	545	65	1198			
Perc. (%) age $(18,100]$	74.87	58.26	45.49	5.43	100			

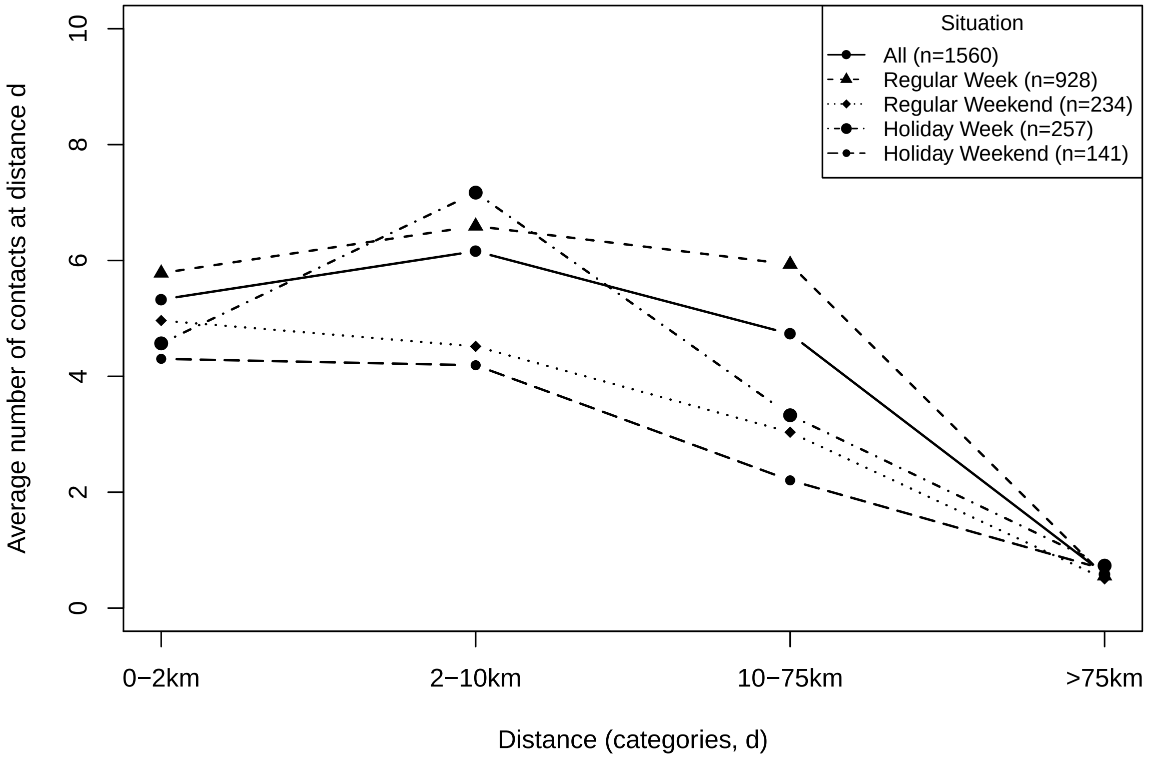


Figure 5.2: The average number of contacts at distance d ($\bar{Y}_{d\cdot}$) are connected by (full) lines for all participants. The dashed line connects the average number of contacts on a regular weekday, the dotted line those from a regular weekend day, the dot-dashed line those from a holiday weekday and the long dashed line those from a holiday weekend day. Discrepancies are present for the different type of days.

5.1.2.2 The degree distribution over distance

Let Y_{dt} denote the number of contacts made by participant t $(t = 1, \ldots, n)$ at the ordered distance category d $(d = 1, \ldots, 4)$ resulting in a 4-tuple $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})'$. Let D_{dt} denote the indicator variable taking value 1 if participant t has travelled distance d and 0 otherwise. Similarly, $\mathbf{D}_t$ denotes the corresponding 4-tuple. Note that we set $Y_{dt} = 0$ when $D_{dt} = 0$. Denote $Y_{\cdot t} = \sum_{d=1}^{4} Y_{dt}$ the total number of contacts made by participant t. Similarly, let $\mathbf{m} = (m_1, m_2, m_3, m_4)'$ and $m. = \sum_{d=1}^{4} m_d$ denote the corresponding expected values which are estimated by $\bar{\mathbf{Y}}. = (\bar{Y}_1., \bar{Y}_2., \bar{Y}_3., \bar{Y}_4.)'$ and $\bar{Y}.. = \sum_{d=1}^{4} \bar{Y}_{d\cdot}$ where $\bar{Y}_{d\cdot} = \frac{1}{n} \sum_{t=1}^{n} Y_{dt}$. During the linking process (Section 5.1.2.1), weights w_{klt} are given to each subcontact k of contact l made by participant t with $\sum_k w_{klt} = 1$. However, they do not imply Y_{dt} to be integer values thus leading to pseudo-counts and non-integer values y_{dt}'s.

Y_{dt} is modeled using a negative binomial regression model allowing for overdispersion

$$Y_{dt} \sim \text{NegBin}(m_{dt}, \phi_d),$$

where $\log(m_{dt}) = \log(\bar{y}..) + \log(f^*(d, \boldsymbol{\theta}_t))$ with $f^*(d, \boldsymbol{\theta}_t)$ a distance function for which $\sum_{d=1}^{4} f^*(d, \boldsymbol{\theta}_t) = 1$ and $\log(\phi_d) = \eta_d$. Denote $f(d^*, \boldsymbol{\theta}_t)$ for a continuous (parametric) distance density, with $\boldsymbol{\theta}_t$ representing the parameters of the distance function, which we regress on participant characteristics. We assume $f^*(d, \boldsymbol{\theta}_t)$ is obtained by integrating $f(d^*, \boldsymbol{\theta}_t)$ over the distance categories, e.g. $f^*(1, \boldsymbol{\theta}_t) = \int_0^2 f(d^*, \boldsymbol{\theta}_t) dd^*$. Denote with $\mathbf{X}_t$ the design matrix of covariates and $\boldsymbol{\eta}_t = \boldsymbol{\beta}' \mathbf{X}_t$ the linear predictor and $\boldsymbol{\theta}_t = \mathbf{g}(\boldsymbol{\eta}_t)$, where $\mathbf{g}(.)$ is a vector of link-functions. We consider the following model as the saturated model

$$f^*(d_i, \theta_t) = \begin{cases} \frac{\exp(\eta_i)}{1 + \sum_{k=1}^{3} \exp(\eta_k)} & \text{for } i = 1, 2, 3 \\ \frac{1}{1 + \sum_{k=1}^{3} \exp(\eta_k)} \end{cases} . \tag{5.1}$$

In order to achieve results representative to the general populations, a weighted likelihood approach is considered with diary weights w_t^d for each participant t. These diary weights are based upon the joint distribution of age and household size from the Flemish population and calculated as in Section 1.2.3.1. The weights are capped at 2 to prevent disproportionate contributions of participants.

Different continuous distance densities are considered to represent the underlying contact dispersal over distance:

(1) the exponential density with $f(d^*, \theta_t) = \frac{1}{\theta_t} e^{\frac{d^*}{\theta_t}}$ whereby $g(.) = \exp(.)$,

(2) the Weibull density with a hazard function parametrization,

$$f(d^*, \boldsymbol{\theta}_t) = \begin{cases} \lambda_t \rho_t (d^*)^{\rho_t - 1} \exp(-\lambda_t (d^*)^{\rho_t}) & d^* \geq 0 \\ 0 & d^* < 0 \end{cases},$$

whereby $\boldsymbol{\theta}_t = (\lambda_t, \rho_t)'$ and $\mathbf{g}(.) = (\exp(.), \exp(.))'$.

(3) To overcome the pole at $x = 0$ for the power law $f(x) = x^{-\alpha}$, a lagged powerlaw density is used and given by

$$f(d^*, \boldsymbol{\theta}_t) = \begin{cases} \frac{\alpha - 1}{\alpha\sigma} & \text{if } d^* \leq \sigma \\ \frac{\alpha - 1}{\alpha\sigma} \frac{d^*}{\sigma}^{-\alpha} & \text{otherwise} \end{cases},$$

whereby $\boldsymbol{\theta}_t = (\alpha_t, \sigma)'$ and $\mathbf{g}(.) = (\exp(.) + 1, \exp(.))'$.

(4) Based on Lomax (1954) the Lomax density is used:

$$f(d^*, \boldsymbol{\theta}_t) = \frac{\alpha}{\lambda}\left(1 + \frac{d^*}{\lambda}\right)^{-\alpha-1},$$

with $\boldsymbol{\theta}_t = (\alpha_t, \lambda)'$ and $\mathbf{g}(.) = (\exp(.), \exp(.))'$.

(5) Additionally, the lognormal density is considered:

$$f(d^*, \boldsymbol{\theta}_t) = \frac{1}{\sqrt{2\pi}\sigma d^*} e^{-\frac{(\ln(d^*)-\mu)^2}{2\sigma^2}},$$

for positive d^*-values and with $\boldsymbol{\theta}_t = (\mu_t, \sigma_t)'$ and $\mathbf{g}(.) = (I(.), \exp(.))'$, with $I(.)$ the identity-link function.

(6) Lastly, the double Pareto density (Reed, 2003) is considered:

$$f(d^*, \boldsymbol{\theta}_t) = \begin{cases} \frac{\alpha_t\beta_t}{\alpha_t+\beta_t}(d^*)^{\beta_t-1} & d^* \leq 1 \\ \frac{\alpha_t\beta_t}{\alpha_t+\beta_t}(d^*)^{-\alpha_t-1} & d^* > 1 \end{cases},$$

with $\boldsymbol{\theta}_t = (\alpha_t, \beta_t)'$ and $\mathbf{g}(.) = (\exp(.), \exp(.))'$.

We were inspired by the work of Meyer and Held (2014) to use the Lomax (4), and the lagged power law (3) density. All densities, except the exponential, are shown in Figure 5.A.2.

We base model comparison on three measures: Akaike's Information Criterium (AIC, Akaike, 1974), Bayesian Information Criterium (BIC, Schwarz, 1978) and the Mean Squared Prediction Error (MSPE, Lawson, 2009 (p. 58)). The latter compares the predictions based on the model to the empirical data and a lower value implies a better model. 4000 predictions are used to obtain a precise, yet stable value of the MSPE (not shown here). Correcting for the correlated nature of the data could be done by extending a GLMM or GEE model. Nevertheless, the imposed mathematical structure makes GEE difficult to use. Extending the GLMM to an NLMM turned out to be computationally too demanding as various correlation structures did not result in convergence. Therefore, point estimates are obtained using the independence model, and a nonparametric stratified bootstrap is employed to capture the variability appropriately (Efron and Tibshirani, 1993).

5.1.2.3 Contact matrices at a distance

Mathematical models taking into account mobility often rely on comparing and contrasting different spatial dispersion models using spatially explicit infectious disease data (Meyer and Held, 2014, 2017). Here we focus at disentangling mobility from contact behaviour and we look at mixing patterns at a distance from home. Remark that in the remainder we use distance from home and imply this to be from the participants' point of view.

Denote Y_{ijdt} for the number of contacts in age class j at distance d from home during one day as reported by participant t in age class i, whereby $t = 1, \ldots, T_i$ with T_i the number of participants in age class i and $i, j = 1, \ldots, K$ with K the number of age groups considered. Denote $m_{ijd} = E(Y_{ijdt}|D_{dt} = 1)$ the expected number of contacts in age class j during one day as reported by a participant in age class i, who travelled distance d, hence $D_{dt} = 1$. The matrix $\mathbf{M}_d$ consisting of elements m_{ijd} is called the social contact matrix at distance d

from home, conditional upon the displacement to distance d. Similarly, denote $m_{ijd^*} = E(Y_{ijdt})$ as the expected number of contacts in age class j during one day as reported by a participant from age class i at distance d. The matrix $\mathbf{M}_{d^*}$ consisting of elements m_{ijd^*} is called the social contact matrix at distance d from home. This last one represents a marginalized version of the first one. Denote $\mathbf{D}_{it}$ the vector indicating for each distance whether the distance is travelled or not, for participant t of age class i with elements D_{idt}. Whenever D_{idt}, for a certain distance d, equals 1, Y_{ijdt} will contribute to the estimation of the social contact matrix at distance d, $\mathbf{M}_d$. Furthermore, recall that whenever D_{idt} equals 0, Y_{ijdt} equals 0, which is used in the estimation of $\mathbf{M}_{d^*}$ (Eq. 5.3). Let w^d_{it} denote the diary weight as described in Section 1.2.3.1 but based on Flemish population data. We estimate the social contact matrix at a distance d, $\mathbf{M}_d$, conditional upon the displacement to distance d as:

$$\hat{m}_{ijd} = \frac{\sum_{t=1}^{T_{id}} w^d_{it} y_{ijdt} D_{idt}}{\sum_{t=1}^{T_{id}} w^d_{it} D_{idt}}. \tag{5.2}$$

Furthermore, the social contact matrix at distance d, $\mathbf{M}_{d*}$ can be estimated as:

$$\hat{m}_{ijd^*} = \frac{\sum_{t=1}^{T_{id}} w^d_{it} y_{ijdt}}{\sum_{t=1}^{T_{id}} w^d_{it}}. \tag{5.3}$$

We denote $\mathbf{M}$ for the social contact matrix based upon all individuals irrespective of distance from home.

We estimate the social contact matrices for stratified situations based on regular and holiday period and week or weekend days. In the general case, the diary weighs w^d_{it} include the joint distribution of age and household size together with the joint distribution of holiday/regular periods and week/weekend days, making the results representative to the school calendar of 2010-2011. Note that the diary weights for different type of days need recalculation because of stratification. The approach considered here reflects that people can meet people at various distances from their homes.

5.1.2.4 Infection pressure at a distance

Let λ_{id} denote the rate at which a susceptible person of age class i $(i = 1, \dots, K)$ acquires infection at distance d, conditional upon travelling distance d from home. We refer to this as the force of infection conditional upon displacement or the conditional force of infection (cFOI). Denote λ_{id^*} for the rate at which a susceptible person of age class i $(i = 1, \dots, K)$ acquires infection at distance d, whether or not the participant travelled distance d from home. We refer to this as the unconditional force of infection (uFOI). These two versions of the force of infection have a different perspective, whereby the first one represents an individual interpretation and the second a population perspective. As such, λ_{id} gives the probability that a person of age group i gets the infection at a distance d, whereas λ_{id^*} represents the probability that a population of age group i experiences infection at a distance d. This last one is closely related to the work of Meyer and Held (2014, 2017) and can be seen as an extension of the meta-population approach (see e.g. De Luca *et al.*, submitted). Assuming the social contact hypothesis (Wallinga *et al.*, 2006) holds, we calculate the cFOI as $\lambda_{id} = \sum_j q m_{ijd} \iota_{jd}$, and the uFOI as $\lambda_{id^*} = \sum_j q m_{ijd^*} \iota_{jd}$ where ι_{jd} is a vector of probabilities taken to be the assumed age-specific relative incidence at distance d, hence reflecting the relative frequency of infected individuals in age class j at distance d.

We study the FOIs by comparing different scenarios for ι_{jd}. More specifically we consider the two extreme cases of only children (ι_{jd}=(1,0)) or only adults (ι_{jd}=0,1)) infected and the case of a 50-50 distribution of infected individuals. To compare to current practices, we estimate the infection pressure using the conventional social contact matrix $\mathbf{M}$. To study the unconditional and conditional approach, we either use $\mathbf{M}$ at each of the distances or incorporate the fact that only a proportion of the people contribute to a distance and thus multiply the observed $\mathbf{M}$ with the age-specific proportions of travel to a particular distance d. We denote this last one as $\mathbf{M}_*$. Furthermore, we apply a rescaling factor to ensure comparability between the two approaches, i.e. the approach with $\mathbf{M}$ or $\mathbf{M}_*$ or the approach with $\mathbf{M}_d$ or $\mathbf{M}_{d*}$. This rescaling guarantees equality at the level of the total number of contacts regardless of distance. We consider different scenarios depending on the type of day reflected in the estimated social contact matrices. We limit ourselves to a regular weekday, a holiday weekday and a weekend day due to limited sample size in holiday and regular weekend, respectively.

Without loss of generalizability, we aim to reflect an influenza-like scenario. Hence the transmissibility (expressed by q) will be chosen to ensure a basic reproduction number of 2, based upon $\mathbf{M}$ for a regular weekday. We assume q to be constant by age as well as for each of the situations, as indicated in Section 1.2.3.3. We reflect on the uncertainty by use of confidence intervals based on a non-parametric bootstrap of the participants.

The analyses were done using the nlmixed procedure in SAS and R.

5.1.3 Results

5.1.3.1 General description

The study encompasses 1,773 individuals, 633 participants indicate no distance for home, another 83 participants fill in a different distance from home. Each of these entries for distance is corrected to 0-1 km as should have been by definition. We copy information about location and distance from a previous time block (when present) when certain time blocks have missing distance, and the location remains the same. Despite these corrections, 193 participants have blanks in the time use data, concerning distance and location, and we disregard those participants for the linking process and analysis.

Section 5.1.2.1 describes that the linking process results in weights for each subcontact, representing the probability that this subcontact occurs at a certain distance and time, and the sum of the weights should equal 1 per contact.

Contacts can happen during a short time frame; hence its location can be missing from the time use data, as the latter one looks at larger time spans of 1 hour or more. Hence, we cannot link those locations to a distance. We present tables and figures based on all (sub)contacts with a distance category, thus ignoring 2,525 unique contacts which had a missing distance for the contact, and hence we base all analysis on 1,560 participants for which the linking was possible.

5.1.3.2 The degree distribution over distance

In total 26,227.27 weighted contacts with distances are reported, leading to an average number of contacts of 16.81 per day and a median of 10 (IQR: 6-19) contacts. On average people reported 5.33 contacts (median 3, IQR: 2-5.35) at a distance 0-1 km from home, 6.16 (median 0.15, IQR: 0-6) at a distance 2-10 km from home, 4.73 (median 0, IQR: 0-3.42) at a distance 10-75 km from home and 0.59 (median 0, IQR: 0-0) at a distance of 75 km or more. Thus indicating more contacts closer to home.

The most prominent distance travel pattern is when remaining within the perimeter of 10 km from home, followed by staying as close as 2 km to home with almost equal frequency for the pattern in which people reach a distance up to 75 km. The other reported pattern ($\mathbf{D} = (1, 0, 1, 0)'$) implies that people travel to a further distance but do not stop or do activities (of sufficiently long duration) within the perimeter between 2 and 10 km from home. Table 5.1 shows that in the prominent distance travel patterns, people do not reach the furthest distance from home. The lack of people travelling a distance more than 75 km from home can be explained by the travel behaviour of Belgian people, as on average they travelled 42.12 km per day per person in 2010-2011 (Reumers *et al.*, 2016). Disentangling the patterns by age, a slightly different decrease can be observed with children showing sharper decline on the third distance category compared to adults and a higher percentage travelling within the 0-2 km radius for children compared to adults.

Table 5.2 shows the (median) number of contacts reported by children and adults at the various distances conditional and unconditional on visiting the respective distances. Close to home, children record more contacts compared to adults conditional upon travelling the distance. For distances beyond 10 km, the result is vice versa, with an increase in the number of contacts at a distance beyond 75 km for children ($n = 18$). The number of contacts in the radius of 2 to 10 km from home is higher for children compared to adults. When considering the unconditional perspective, we observe that children report fewer contacts for distances beyond 10 km, whereas this only occurs for adults at distance beyond 75 km.

Figure 5.2 shows the observed overall degree distribution over distance as a solid line and the degree distribution for regular weekdays (dashed line), regular weekend days (dotted line), holiday weekdays (dot-dashed line) and holiday weekend days (long dashed line). A clear discrepancy is present in these various situations, with the most visible one being between a regular and holiday weekday. As expected, a difference based on age is also present as can be observed in Figure 5.A.1. Tables with the observed (total) numbers of contacts can be found in Section 5.A.4.

Next, we present our model results. Table 5.3 shows the difference in AIC values (ΔAIC) compared to

Table 5.2: The median number of contacts (Q_1, Q_3) at specific distances (un)conditional upon visiting the distance shows a discrepancy between children and adults.

	Median number of contacts (Q_1, Q_3)			
Age category	**at 0-2 km**	**at 2-10 km**	**at 10-75 km**	**over 75 km**
Conditional upon travelling the distance				
$[0, 18]$	4 (3, 7.43)	8 (2.75, 18.06)	4.5 (2, 9.46)	8.82 (5, 13)
$(18, 100]$	3 (2, 5)	4.37 (2, 10.86)	6.05 (2, 14)	9.41 (3, 14.85)
Unconditional upon travelling the distance				
$[0, 18]$	4 (3, 7)	0.09 (0, 9)	0 (0, 0)	0 (0, 0)
$(18, 100]$	3 (2, 5)	0 (0, 5)	0 (0, 5)	0 (0, 0)

the best fitting model, e.g. with the lowest AIC value, thus ΔAIC=AIC-AIC$_{min}$ together with the difference in BIC values (ΔBIC=BIC-BIC$_{min}$) and the difference in MSPE values (ΔMSPE=MSPE-MSPE$_{min}$).

These results show that based on AIC and BIC the lognormal model with the log mean depending on age and week is the best model. Using BIC, the lognormal model with only age as a covariate for the log mean performs equally well. Based on the MSPE the saturated model with age, week and regular interaction leads to the best fit to the data. The MSPE of the Weibull model without covariates or with week for both parameters is close to one of the best fitting model. The AIC and BIC of the Lomax and Weibull model with one dependent parameter are close to the AIC and BIC of the best fit model and the models usually perform better compared to the saturated model.

As the nonparametric bootstrap of participants and fitting the best model was not final, Figure 5.3 shows observed results. The different data points (discrete) show the observed degree distribution over distance by the situations considered, whereas the various line types depict the fit from the best models for these situations. The colours indicate the fit by the saturated (black) or lognormal (red) model.

Both models show similar patterns, although the additional regular/holiday term in the weekend can induce the present deviations. With increasing age, the number of contacts at a distance increases and diminishes again for the elderly (above 65 years). The saturated model depicts even a larger probability of contacts at the 10 to 75 km distance for individuals aged between 25 and 45 years during regular weekdays. This age-specific increase can be explained by them being working adults who travel to work at a further distance. However, in this age class, the models do not fit as good as compared to other age classes, together with age class 18 to 25 years and 45 to 65 years.

The model includes a different overdispersion parameter for each distance category, the obtained estimates are displayed in Table 5.A.4. All estimated overdispersion values show a larger variation compared to the mean.

5.1.3.3 Contact matrices at a distance

Estimating a social contact matrix at a distance further than 75 km is only be based on a limited number of participants (83), which should be kept in mind. In Figure 5.5 we see the representation of the m_{ijd} values for each of the cells of the social contact matrices, whereby the line type and the shape of the dot differentiate between the the cells of $\mathbf{M}_d$. The cells are plotted over distance and connected with lines. As such, the evolution of the average number of contacts between groups of individuals over distance conditional upon the displacement is described. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap, more runs are necessary to obtain convergence. We observe an increase in the average number of contacts amongst adults with increasing distance; however, given the smaller number of participants this results in wider confidence bands. Concerning contacts between children, we observe a slight increase in the average number of contacts at a distance between 2 and 10 km, but further on a decline. Adults report a similar number of contacts with children over distance, except for distances beyond 75 km. Children, on the other hand, show a declining trend in average number of contacts with adults, except at distances beyond 75 km influenced by a limited number of participants (18).

Quite a similar pattern is observed on a regular weekday (Figure 5.6, top panel). During regular weekend

Table 5.3: Model comparison measures for the distance density functions with various covariate combinations. Covariates considered are age, week/weekend (*week*) and regular/holiday (*reg*) period, whereby + refers to the combination of covariates only as main effects and $*$ refers to using all interactions possible with each of the main effects. The columns show ΔAIC=AIC-AIC$_{min}$, ΔBIC=BIC-BIC$_{min}$ and ΔMSPE=MSPE-MSPE$_{min}$.

Model	ΔAIC	ΔBIC	ΔMSPE$_{4000}$	AIC	BIC	ΔMSPE$_{4000}$
	Saturated Model			**Lagged Powerlaw Model**		
no cov	150	102	142.07	184	129	771.58
age	20	114	195.82	49	42	1490.74
week	149	122	146.45	180	132	800.85
age + week	17	131	207.34	45	45	1,629.48
*age * week*	35	291	199.59	50	96	1,807.40
*age * week * reg*	47	625	**0**	455	610	12,027.84
	Exponential Model			**Lomax Model**		
no cov	1,369	1,308	414.85	166	111	311.18
age	1,240	1,226	409.81	22	14	792.24
week	1,371	1,316	415.60	161	114	371.30
age + week	1,242	1,234	405.76	17	17	864.98
*age * week*	1,198	1,238	406.89	21	67	800.00
*age * week * reg*	1,100	1,248	404.50	39	193	764.43
	Weibull Model (λ)			**Weibull Model (λ & ρ)**		
no cov	151	96	137.61	151	96	137.61
age	24	17	311.96	23	63	188.52
week	149	101	148.44	148	107	140.86
age + week	21	20	364.09	17	71	191.66
*age * week*	22	68	372.32	30	178	216.09
*age * week * reg*	41	195	296.64	11	374	175.00
	Lognormal (θ)			**Lognormal (θ & λ)**		
no cov	151	97	172.98	151	97	172.98
age	8	**0**	237.10	19	59	306.56
week	147	99	194.02	149	108	194.08
age + week	**0**	**0**	277.47	13	66	307.79
*age * week*	7	54	255.52	25	173	357.61
*age * week * reg*	26	180	272.15	25	388	297.20
	Double Pareto Model (α & β)					
no cov	184	129	792.18			
age	54	93	1,239.60			
week	182	140	772.59			
age + week	47	100	1,127.59			
*age * week*	61	209	1,446.18			
*age * week * reg*	82	445	1,170.99			

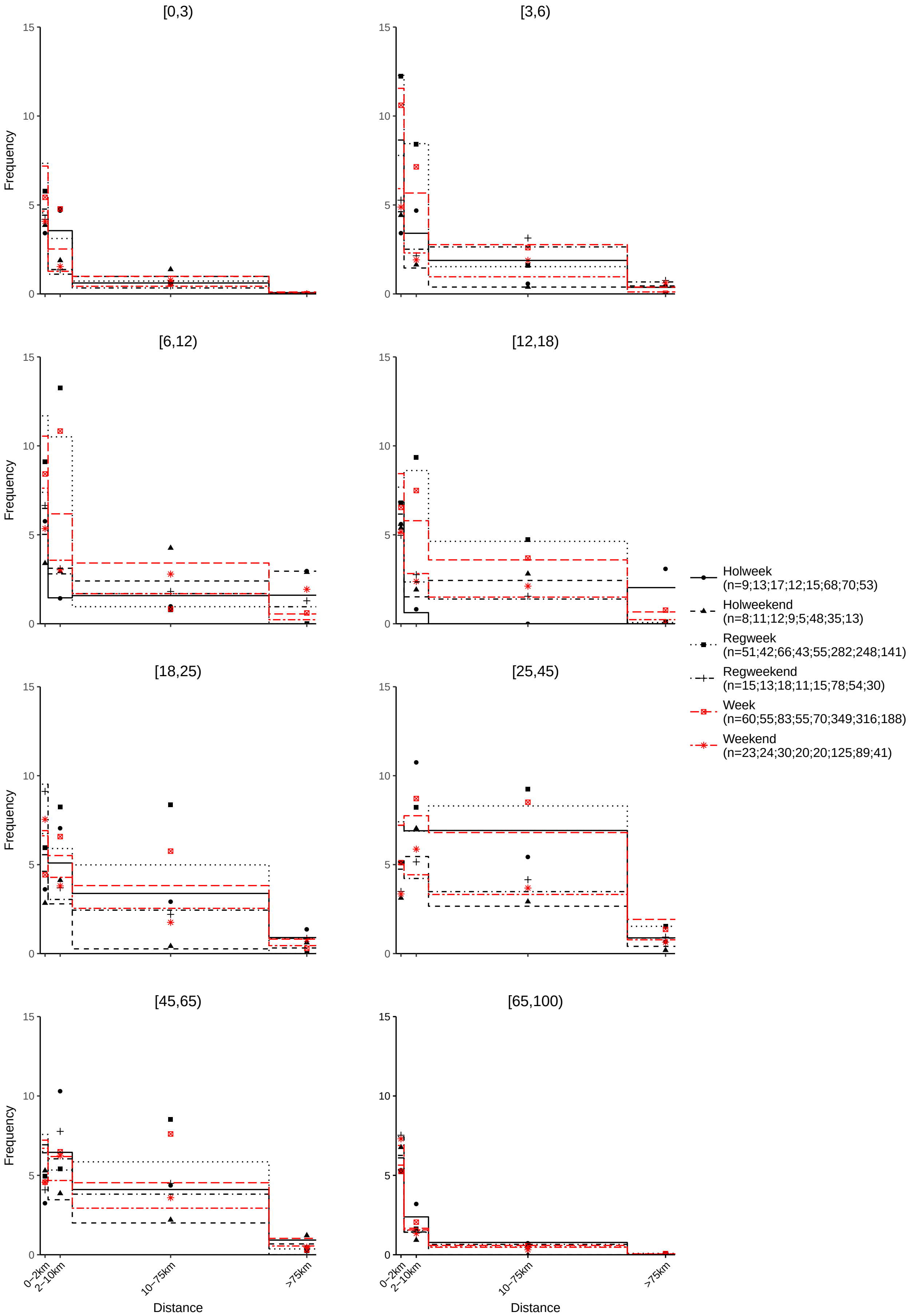


Figure 5.3: The observed (points) degree distribution over distance with the saturated model with age, week and regular periods included (black lines) and the lognormal model with age and week without interaction included (red lines). The type of dot, as well as the lines, deviate depending on the situation. The numbers in the brackets of the legends give the sample size for each of the situations. Each panel shows the fit for a particular age category.

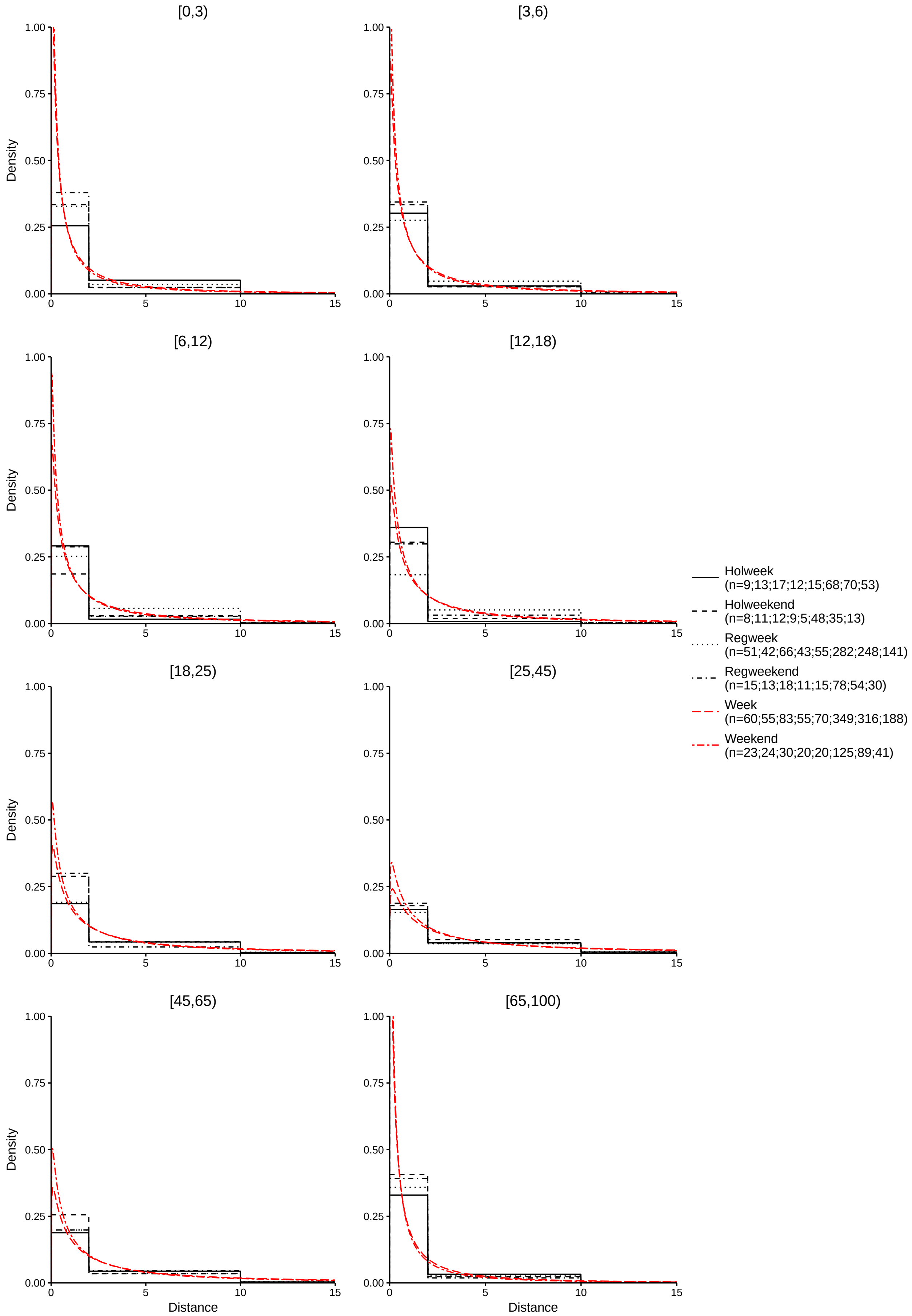


Figure 5.4: The estimated underlying degree densities over distance for the saturated model with age, week and regular periods included (black) and the lognormal model with age and week without interaction included (red). The type of lines deviates depending on the situation. The numbers in the brackets of the legends give the sample size for each of the situations. Each panel shows the densities for a particular age category.

days (Figure 5.6, second row) we find a slightly different pattern, with a slight increase in the number of contacts with adults over distance, reported by children. This tendency can also been seen for holiday weekdays (Figure 5.6, third row). For holiday weekend days (Figure 5.6, bottom panel), we see this pattern for contacts with children from the perspective of the adults.

When looking into the same aspects but based on the unconditional social contact matrices $\mathbf{M}_{d*}$, we find a declining trend over distance for each of the specific contacts (Figure 5.7). We observe this downward trend also for each of the types of days. However, we note the remarkable switch from predominantly child-child contacts at distance 2-10 km during regular week days to predominantly adult-adult contacts at this distance during holiday weekdays in Figure 5.12.

5.1.3.4 Infection pressure at a distance

We study the infection pressure, the cFOI and the uFOI, for different relative incidences as indicated in Section 5.1.2.4. For comparison a scenario with a constant m_{ij} over distance is considered as well. In Figures 5.9 & 5.10 the top rows are the results for the three different relative incidences based on the cFOI or the uFOI, respectively and the bottom row represents the scenario's when using constant m_{ij}'s, thus reflecting the current approach. The left panels show the result when only children are infectious, the middle when the infectious people are half-half children and adults, and the right panels show it when only adults are infectious. For the conditional case $\mathbf{M}$ is used at each distance thus based on 1,560 participants. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap, more runs are necessary to obtain convergence. Figures 5.11 & 5.12 show the same aspects but disentangling for the different type of days, i.e. a regular weekday, a holiday weekday or a weekend day.

From Figure 5.9, we observe an increase in the cFOI for children at a distance 2-10 km from home when only children are infectious at a distance d; adults show a constant force of infection over distance. In case the infectious are 50-50 adults and children, we see a rather constant force of infection over distance. When only adults are infectious at a distance d, we see an increasing trend in the cFOI over distance for adults and children. However, for the children a lot of uncertainty is present at a distance over 75 km, so we should not over-interpret this result. The bottom panel of Figure 5.9 shows the current practice the cFOI and displays a uniform distribution over distance with the heights depending on who was infectious at a distance d. When only children are infectious, the cFOI for children is highest and the other way around when only adults are infectious. For the 50-50 case, we observe similar cFOIs with a slightly higher value for children.

When zooming in on the different type of days (Figure 5.11), we see slightly decreasing cFOIs over distance on weekend days. This observation applies to each of the three situations. During a holiday week, situations where at least some adults are infectious show an increase of the cFOI at a distance 2-10 km from home. During regular weekdays children experience a higher cFOI at a distance 2-10 km from home when children are infectious at the distances. Remarkably, adults experience an increased risk over distance when adults are infectious at the distances. Hence, when adults are travelling, their risk of acquiring infection increases by the distance they travel.

From Figure 5.10, we observe a declining trend for children across distance for the three situations of rela-

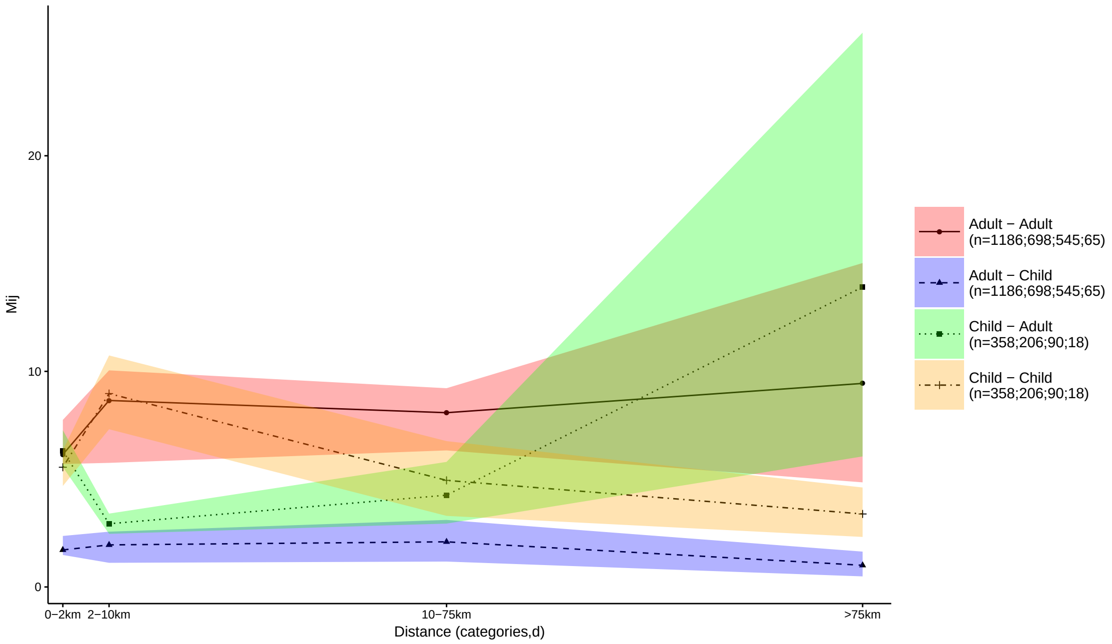


Figure 5.5: Lines connect the dots which represent the cells of the social contact matrix over distance $\mathbf{M}_d$ conditional upon travelling the distance. The first individual in the legend is the participant; the second individual is the contact. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap. The key indicates the number of participants contributing information by distance.

tive incidence based on the changing behaviour. For adults this trend can also be seen, except when only children are infectious at a distance d, adults experience a small uFOI overall. In the case of the constant social contact matrix situation (bottom panel of Figure 5.10), both adults and children experience a decreasing uFOI by distance. When only adults or children are infectious at a distance d, the adults or children suffer the highest risk of getting infectious, respectively. In the 50-50 case, their risk is similar.

Zooming in on the different type of days (Figure 5.12), more subtle changes are present. During weekends, the trends are still closer to a setting with a constant force of infection. During holiday weeks an increase in the uFOI is present at 2-10 km for adults in case adults are infectious at the distances and children experience the highest risk close to home (but still small) and almost no hazard at further distances. For regular weekdays children experience an increased risk at a distance 2-10 km from home when only children are infectious at each of the distances. Adults experience slightly more risk of infection at distance 10-75 km from home when only adults are infectious with an increasing trend over distance.

5.1.4 Discussion & Conclusion

In this work we used a social contact survey conducted in Flanders in 2010-2011, to link social contact behaviour to distances. This survey allowed us to study the dispersal of contacts over distance. We investigated three aspects related to the dispersal, e.g. the underlying distance dispersal kernel, the social contact matrices over distance and the infection pressure over distance.

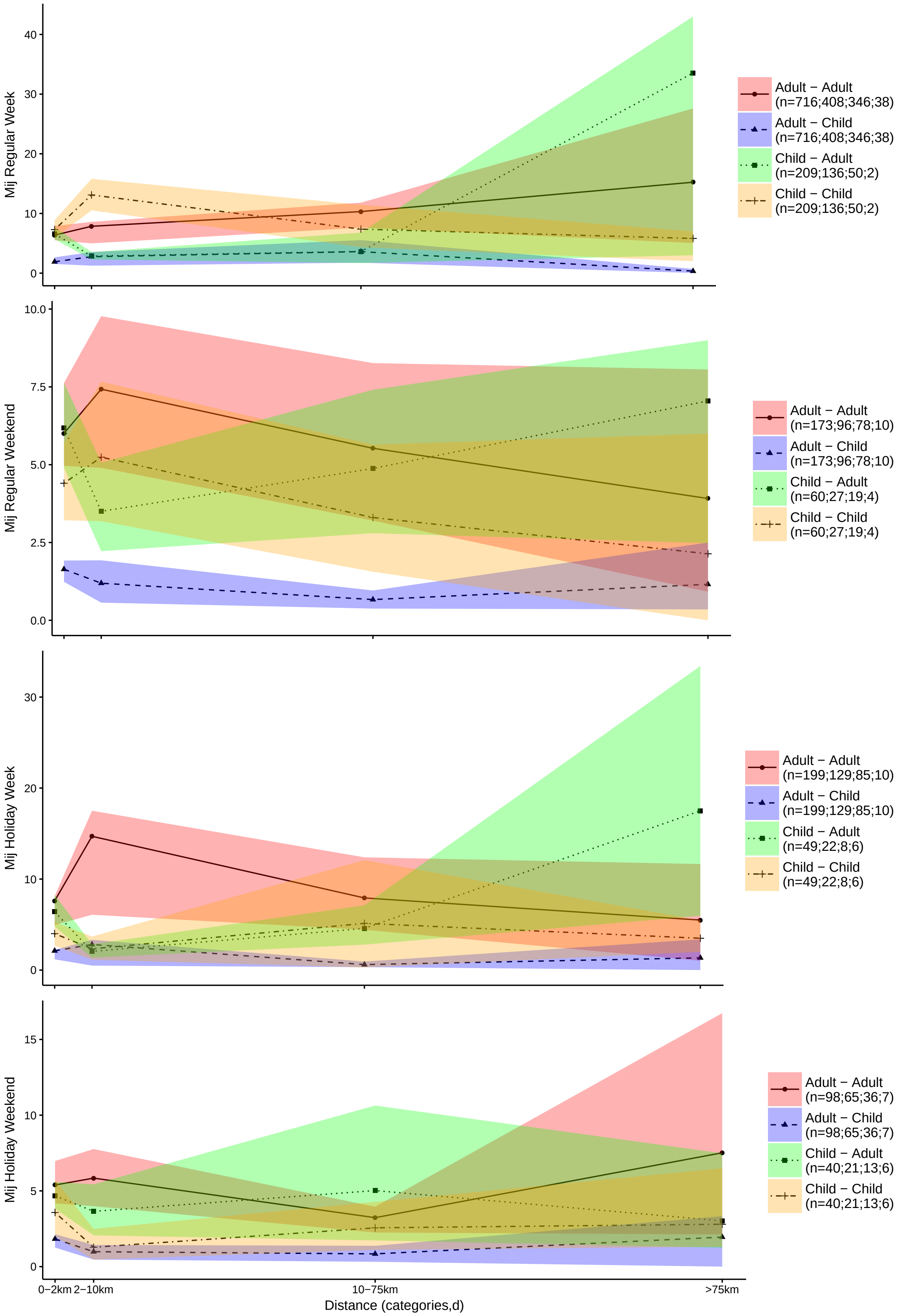


Figure 5.6: Lines connect the dots which represent the cells of the social contact matrix over distance $\mathbf{M}_d$ conditional upon travelling the distance, a panel for the different types of days. The first individual in the legend is the participant; the second individual is the contact. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap. The key indicates the number of participants contributing information by distance.

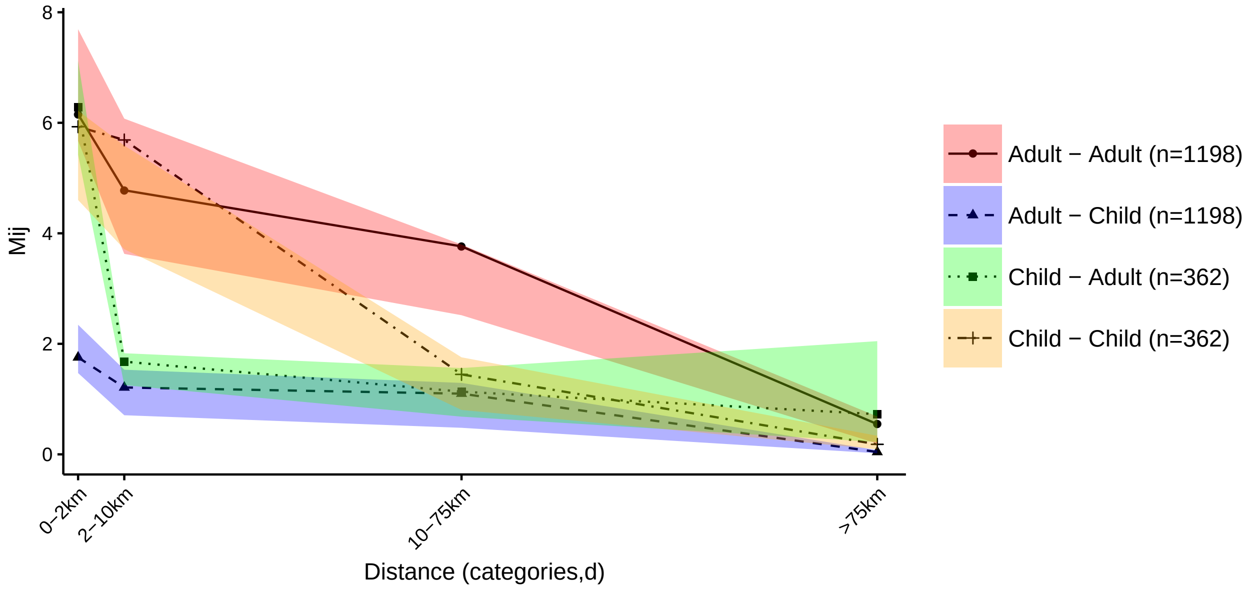


Figure 5.7: Lines connect the dots which represent the cells of the social contact matrix over distance $\mathbf{M}_{d*}$ unconditional upon travelling the distance. The first individual in the legend is the participant; the second individual is the contact. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap. The key indicates the number of participants contributing information by distance.

We observe the mean number of contacts to diminish by distance, with 5.33 contacts at a distance 0-2 km from home, 6.16 at a distance 2-10 km from home, 4.73 at a distance 10-75 km from home and 0.59 for contacts at a distance beyond 75 km. Fu *et al.* (2012) seem to have collected the distance from home for every contact, but apart from reporting percentages, they did not use the data or offered a more in-depth explanation. They observe almost half of the contacts to be reported at a distance less than 1 km from home, compared to our results, this is slightly more. Nevertheless, our results are in agreement with their observation of a declining trend across distance.

Our results favour a lognormal degree distribution over distance, and thus do not support a powerlaw model, in contrast to results of others based on data on the spread of infectious diseases (Meyer and Held, 2014, 2017). Furthermore, the age and the day of the week are important in the degree distribution over distance. However, the unconditional social contact matrices, seem to show a powerlaw alike decline in general. While the type of day induced subtle changes, a distinction not made by the previous authors. When studying the related infection pressure, we see discrepancies between the age groups, which are also affected by the age-specific distribution of infectious individuals. We show evidence that the approach considered by De Luca *et al.* (Submitted) and others, who use a constant social contact matrix across distance, insufficiently captures the truth of subtle differences of children having mainly an influence at a distance close to home, and adults having an overall effect when the infectious individuals are mostly adults. De Luca *et al.* (Submitted) studied changes across different types of days and observed the importance of changes in mixing behaviour on the spread of infectious diseases such as influenza. We show evidence of changes present in infection pressure according to the type of day. Additionally, our results

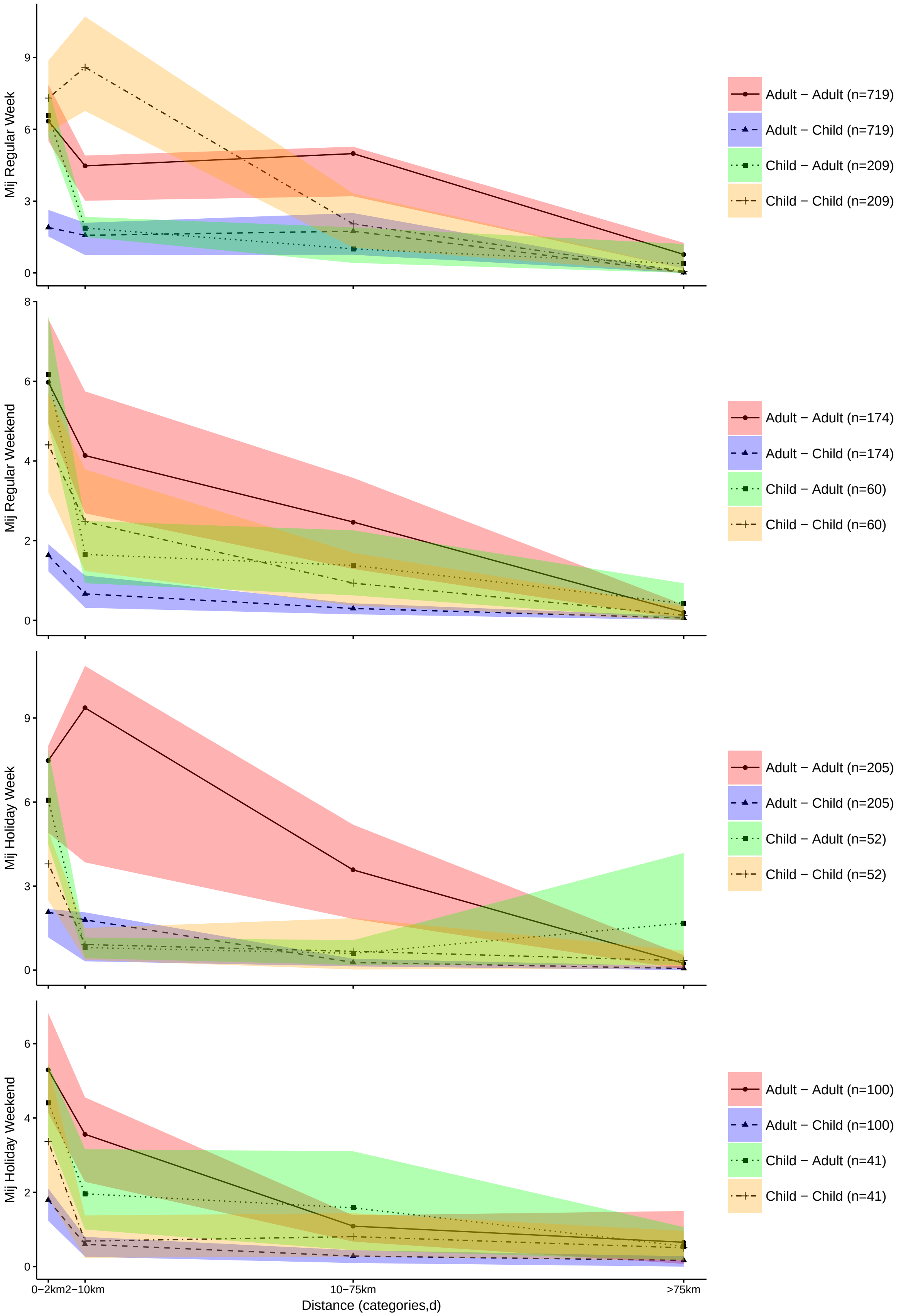


Figure 5.8: Lines connect the dots which represent the cells of the social contact matrix over distance $\mathbf{M}_{d*}$ unconditional upon travelling the distance, a panel for the different types of days. The first individual in the legend is the participant; the second individual is the contact. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap. The key indicates the number of participants contributing information by distance.

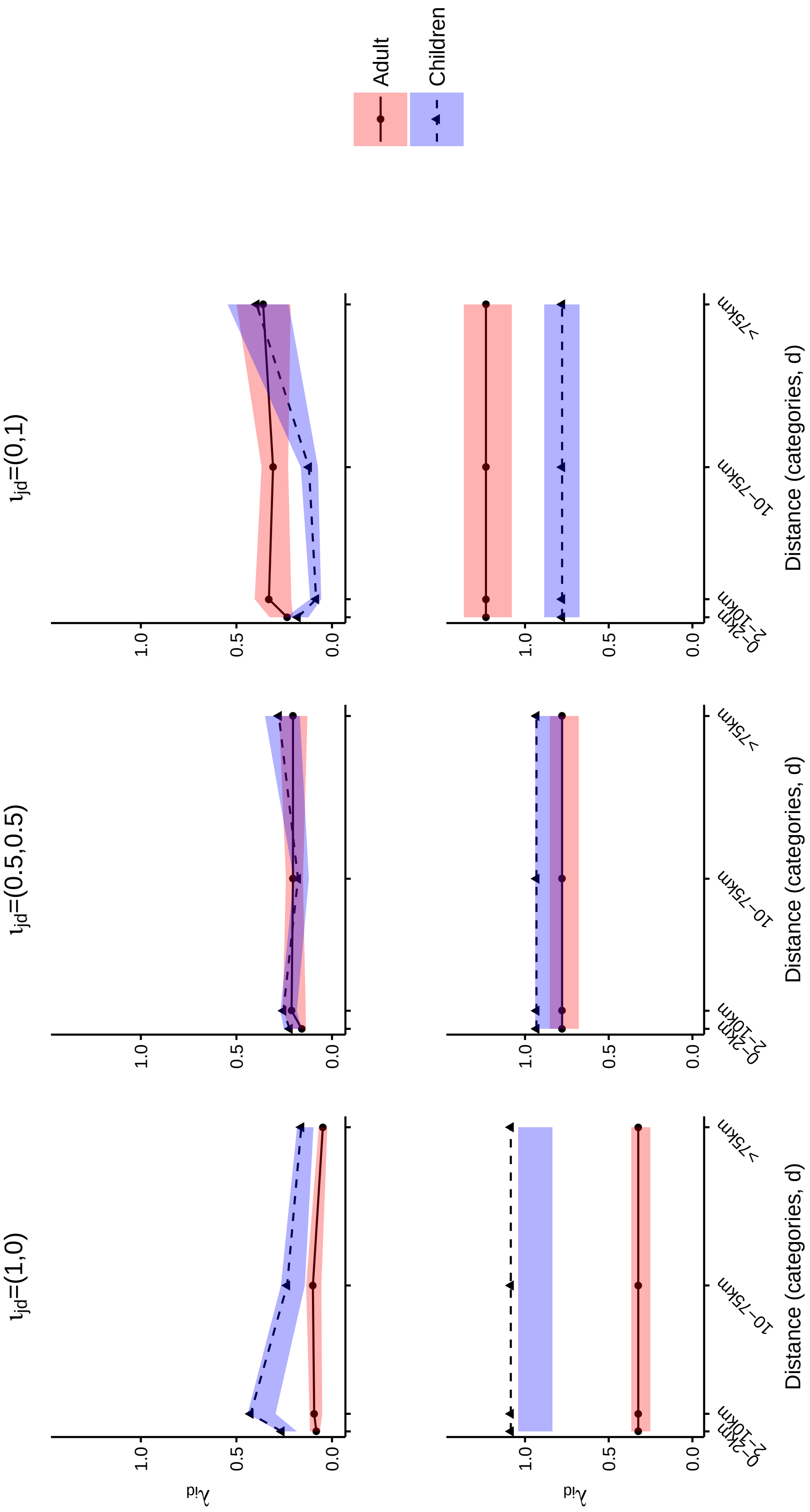


Figure 5.9: The conditional FOI (top row) and the constant matrix across distance approach (bottom row) for different relative incidences. Different line types and shapes of dots distinguish between children ($[0, 18]$) and adults ($(18, 100]$). The left panel shows the cFOI in case only children are infectious; the middle panel shows the cFOI given the relative incidence is equal across ages and the right panel shows the cFOI when only adults are infectious. We calculated a rescaling factor at the level of contact matrices (Section 5.1.2.4) and applied it for ease of comparison. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap.

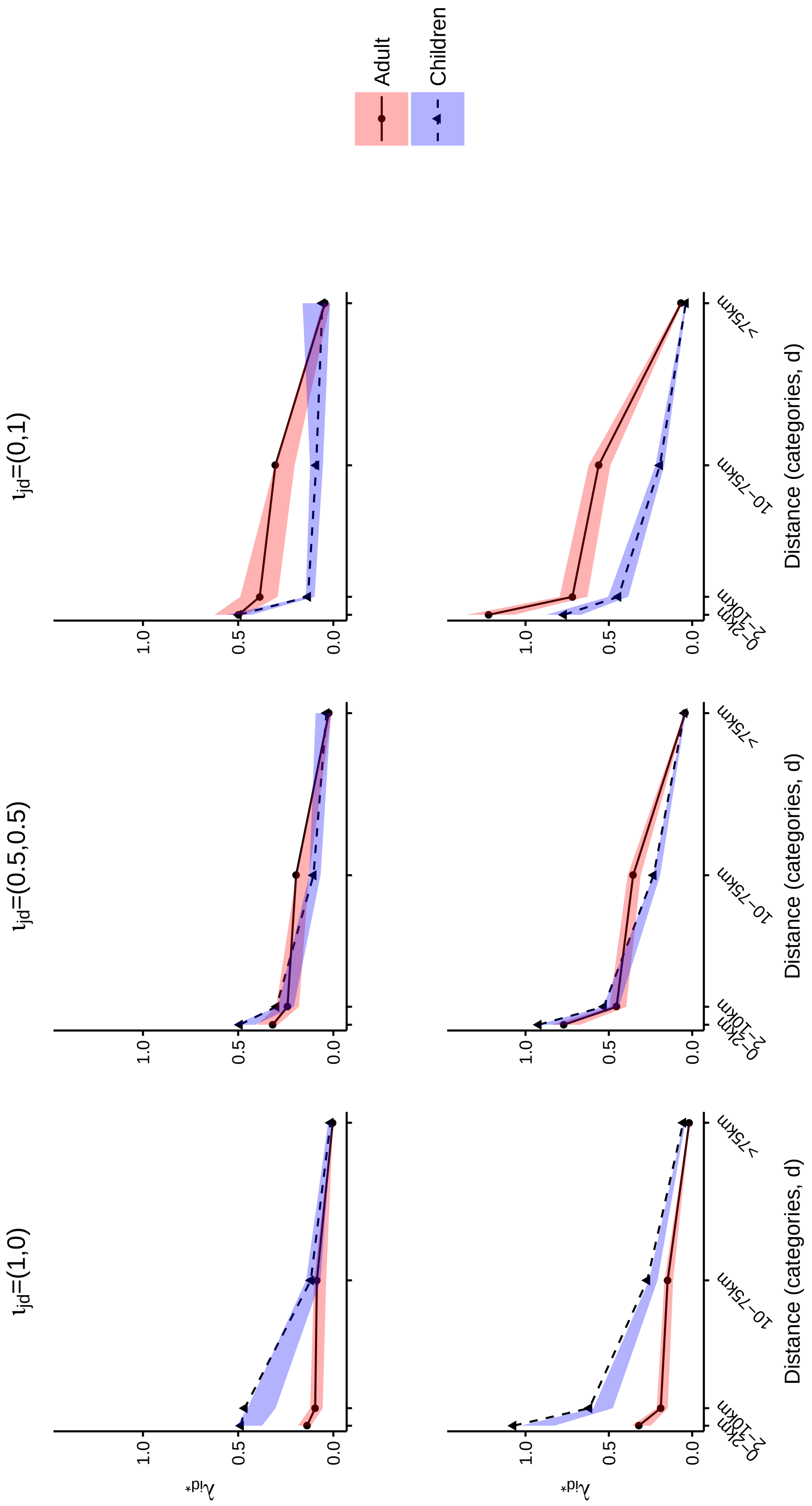


Figure 5.10: The unconditional FOI (top row) and the constant matrix across distance approach (bottom row) for different relative incidences. Different line types distinguish between children ($[0, 18]$) and adults ($(18, 100]$). The left panel shows the uFOI in case only children are infectious; the middle panel shows the uFOI given the relative incidence is equal across ages and the right panel shows the uFOI when only adults are infectious. We calculated a rescaling factor at the level of contact matrices (Section 5.1.2.4) and applied it for ease of comparison. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap.

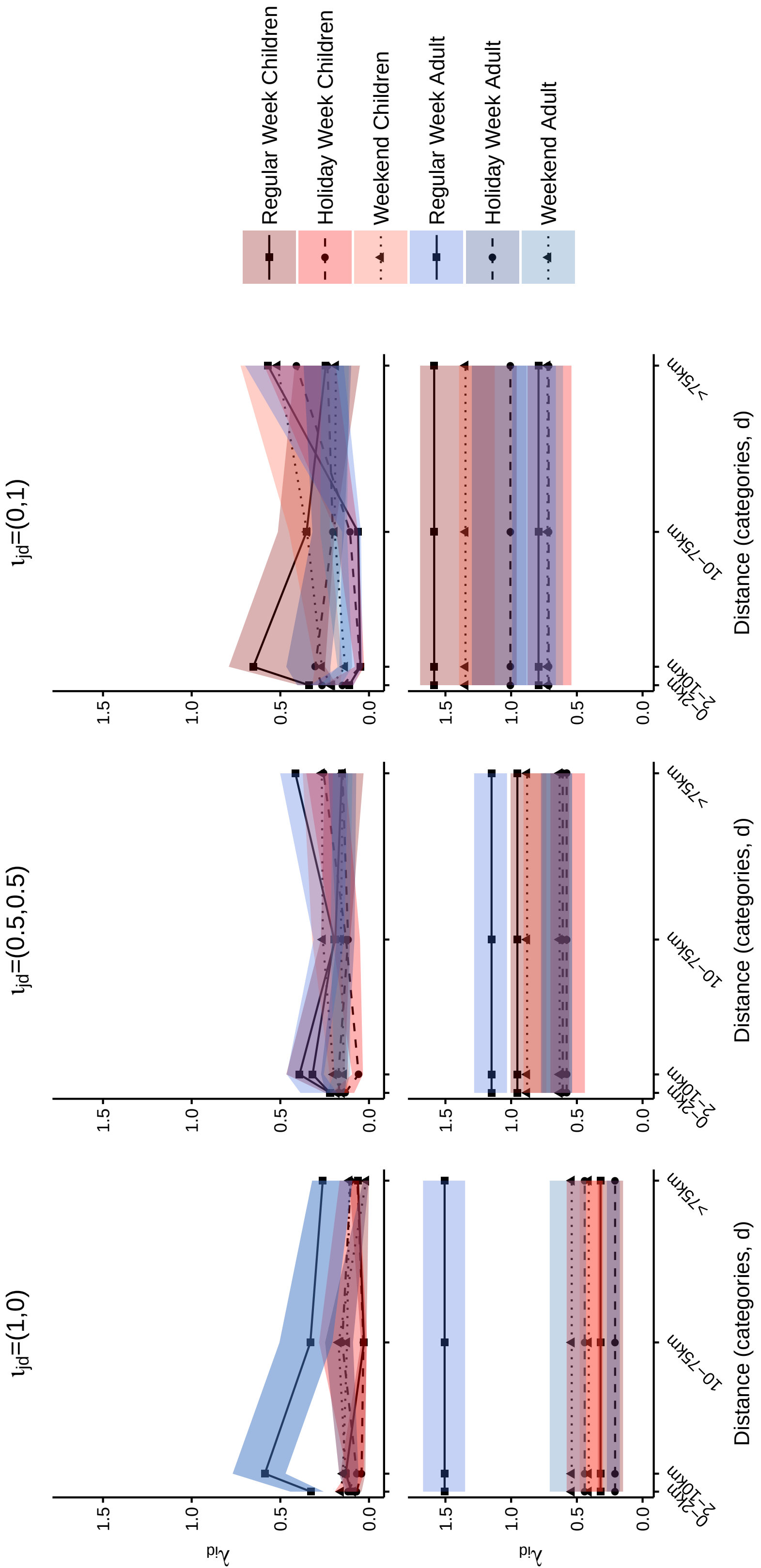


Figure 5.11: The conditional FOI (top row) and the constant matrix across distance approach (bottom row) for different relative incidences and different day types. Different line types and shapes of dots distinguish between day types; the colours differentiate between children ($[0, 18]$) and adults ($(18, 100]$). The left panel shows the cFOI in case only children are infectious; the middle panel shows the cFOI given the relative incidence is equal across ages and the right panel shows the cFOI when only adults are infectious. We calculated a rescaling factor at the level of contact matrices (Section 5.1.2.4) and applied it for ease of comparison. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap.

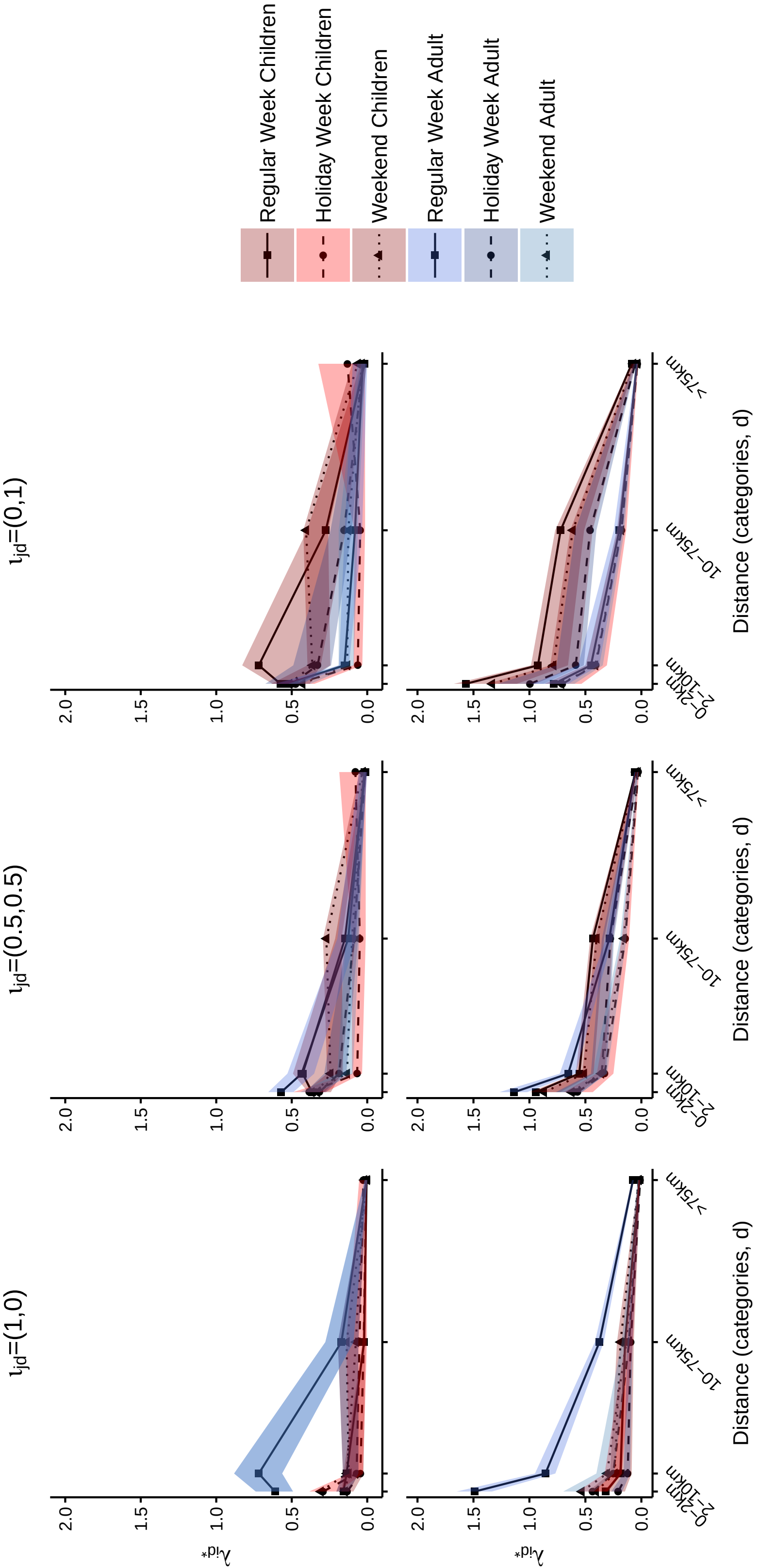


Figure 5.12: The unconditional FOI (top row) and the constant matrix across distance approach (bottom row) for different relative incidences and different day types. Different line types and shapes of dots distinguish between day types; the colours differentiate between children ($[0, 18]$) and adults ($(18, 100]$). The left panel shows the uFOI in case only children are infectious; the middle panel shows the uFOI given the relative incidence is equal across ages and the right panel shows the uFOI when only adults are infectious. We calculated a rescaling factor at the level of contact matrices (Section 5.1.2.4) and applied it for ease of comparison. The presented confidence intervals are based on an initial 900 runs of a nonparametric stratified bootstrap.

with changing contact matrices over distance show more discrepancies compared to the work of De Luca *et al.* (Submitted). The results of our degree distribution over distance indicate the need for jointly incorporating travel and contact behaviour from the same source.

Next, we presented a conditional approach to the social contact matrices and the force of infection. The results deviate from the previous results as we use a different mindset. As mentioned before, we can interpret the conditional approach for the force of infection in an individual context. Therefore, these results are more useful to construct agent-based models. Although, similarly we observe the type of day to play a crucial role in the effect as well as the distribution of infected individuals across age. Overall, using changing social contact behaviour seems to result in a more decreasing trend compared to the unconditional approach. The general trend, not considering the type of day appears to be increasing. However, this method is prone to a reduction in sample sizes across distance; thus one needs to be cautious in the interpretation of the results at a further distance. A possible solution could be in taking the last two categories together, however with the disadvantage of less detailed information.

Kucharski *et al.* (2015) look at travel patterns of children and quantify a difference between school term and holiday term for the longest distance travelled by children when following them a fortnight. They motivate their work given the limited knowledge on movement behaviour of children and report that various authors use the combination of social mixing information and movement data (Meyer and Held, 2017; Colizza *et al.*, 2007; Ferguson *et al.*, 2006; House *et al.*, 2011; Keeling and Rohani, 2007; Apolloni *et al.*, 2014; De Luca *et al.*, Submitted). Our work is not the first to report results on social contact behaviour and distance as Fu *et al.* (2012) and Danon *et al.* (2012) have collected this information as well. However, we are the first to analyse both aspects jointly to estimate the contact dispersal kernel. Furthermore, by the degree distribution over distance, we show the importance of considering spatial aspects in social contact data as well given contacts and movements to be dependent.

To study and quantify the effect of the dispersal of social contact behaviour, we used the infection pressure. Another approach of interest could be the expected dispersal of a newly emerging epidemic, reflecting where you will infect most people when you are infectious. We did not explicitly state these results since we can deduce them from the degree distribution over distance, as this reflects the distribution of contacts by distance and thus the possible people you can infect when travelling. However, a question with this approach is whether people travel when feeling ill. Nonetheless, when infections are merely asymptomatic, it offers first insights.

An assumption made by using the diary weights, to make the results representative of the population and for a year, is that the data collected during a particular period of the year are representative for a full year. Albeit the novelty of the presented methods and results, there are some limitations. To be able to study social contact behaviour across distance we linked social contact data to time use data. This linking led to pseudo-counts, as only probable links could be deduced assuming the contact diary to be filled in in chronological order. As participants did not fill in the time use question entirely, they were not considered and this lead to missingness. Furthermore, contacts have a smaller time scale compared to the 1-hour time blocks of the time use data, thus some contacts could not be linked to a distance, and hence missingness is introduced. In the presented analysis, we ignored this missingness and based the conclusions on all available information. Lacking locations and distances added additional missingness; however, this might also be present in other studies as people can forget to fill in some questions. Due to computational difficulties in the extension of the independent models to an NLMM, we accounted for the correlated nature of the data at hand by using nonparametric stratified bootstrapping.

Despite these limitations, this work is the first to show the spatial dispersal of social contact behaviour rigorously and the first to contrast the commonly assumed independence between movement data and social contact behaviour. We, therefore, believe this work to be valuable in the various range of epidemic models and as a tool to quantify social contact behaviour over distance. Given this work to be the first, it is desirable to apply the techniques to other social contact data with a distance measure. The missingness induced due to the differences in time scales could be avoided by either requesting a more detailed time use or requesting the distance for each contact. However, in both cases, the contact diaries are becoming very labour intensive to participants which might result in lower quality of the data. A possible solution would be the use of an app instead of the paper-based questionnaire, which can deduce the distance or can avoid repetition of particular questions by its set-up. Nonetheless, this has to date not yet been compared to the paper-based diary approach and has limitations of its own, but yet offers new research routes.

Furthermore, we considered a simple approach to quantify the impact of contact dispersal over distance, driven by certain assumptions. One of these assumptions being the age-distribution of infectious cases to be the same across distance and the other that q is constant across age and distance. To see the impact more rigorously, we should incorporate our results into an appropriate model, and this thus opens up further avenues of research. Another plausible approach would be conditioning on travel patterns given that they could imply differences in contact behaviour. Nevertheless, we limited ourselves in this work to a single interpretation. Another aspect of further research is the inclusion of gender in the models as the current approach is still lacking this factor. As differences in social contact behaviour are known (Dodd *et al.*, 2016; Béraud, 2015) and differences in the travel behaviour might furthermore be present, including this covariate in the described approach can offer additional insights.

Furthermore, the current survey is a cross-sectional survey which allows comparing it to the previous Flemish part of the POLYMOD project and thus challenging the time homogeneity of social contact behaviour (Kifle *et al.*, To be submitted).

5.2 Parallel study: Contacts and mobility in a metapopulation setting

This section is based on the submitted work: "De Luca, G., Van Kerckhove, K., Coletti, P., Poletto, C., Bossuyt, N., Hens, N., Colizza, V. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. Submitted to BMC Infectious Diseases.".

Parallel to the previous work, we combined data on social contacts and mobility in a metapopulation model with the aim to quantify the effect of regular school closure on seasonal influenza epidemics. School closure has the potential to reduce influenza propagation in the community by breaking major chains of transmission as children represent an epidemiological group of central importance for the transmission of influenza (Longini *et al.*, 1982; Viboud *et al.*, 2004; Baguelin *et al.*, 2013). It is expected to potentially delay the peak, and reduce the epidemic impact, at the peak time of the overall wave. Hence it is envisioned as a possible non-pharmaceutical intervention for pandemic mitigation, however not recommended by the WHO during the 2009 A/H1N1pdm epidemic and thus left to the decision of national and local authorities (World Health Organization Writing Group *et al.*, 2006; World Health Organization, 2009). A large body of literature exists on the topic; however contradictory evidence lead to no definitive emerging consensus (Cauchemez *et al.*, 2009b; Jackson *et al.*, 2013). Given its potentially crucial role in reducing the epidemic impact, Jackson *et al.* (2013) investigated school closure in the realm of pandemics. Few of the transmission models fitting the 2009 H1N1 pandemic (Merler *et al.*, 2011; Apolloni *et al.*, 2013; Fumanelli *et al.*, 2016; Ciavarella *et al.*, 2016; Eames *et al.*, 2012) are based on estimates for social mixing changes (Apolloni *et al.*, 2013; Eames *et al.*, 2012), as data collected during a pandemic is limited (Eames *et al.*, 2012; Jackson *et al.*, 2011), leaving other approaches to rely on assumptions about contacts that may critically affect the studies' findings. Applications to seasonal influenza may on the other hand count on a more accurate description of population mixing, given available social contact surveys (Chapter 2, Hens *et al.*, 2009a; Eames *et al.*, 2010, 2011; Béraud *et al.*, 2015). As such they offer the opportunity to perform data-driven modelling studies that mechanistically assess the role of school holidays on seasonal influenza. Interestingly, Hens *et al.* (2009a) also highlighted considerably large differences across countries in the way contacts change from term-time to school holidays, suggesting the need for country-specific estimates to accurately and reliably parameterize models.

Changes in mobility is another important aspect that researchers rarely integrate into school closure studies. Travel is known to be responsible for the spatial dissemination of influenza (Apolloni *et al.*, 2013; Grais *et al.*, 2003; Viboud *et al.*, 2006; Cooper *et al.*, 2006; Ferguson *et al.*, 2006; Epstein *et al.*, 2007; Fraser *et al.*, 2009; Balcan *et al.*, 2009a,b; Bajardi *et al.*, 2011). In addition to extraordinary travel drops in reaction to epidemics (Bajardi *et al.*, 2011; Abdullah *et al.*, 2004; Poletto *et al.*, 2014), mobility changes regularly occur during school holidays compared to term-time (Kucharski *et al.*, 2015; Ewing *et al.*, 2017). Moreover, important differences were highlighted in the mobility of children vs. adults and their associated variations, so that their coupling with social mixing changes occurring during holidays may have a considerable impact on the epidemic outcome (Apolloni *et al.*, 2013; Ewing *et al.*, 2017).

Here we aim to explicitly integrate social mixing and travel from data into a modelling framework to assess how variations induced by regular school closure may impact seasonal influenza epidemics. Two modelling studies were developed so far with similar objectives. Apolloni *et al.* (2013) used a stylised analytical approach to evaluate the role of age-dependent social mixing and travel behaviour on the conditions for epidemic spatial invasion (Apolloni *et al.*, 2013). The model can compare different contexts, with or without schools in terms, and also account for associated changes. However, they consider the contexts independently (no full school calendar) and the epidemic impact is evaluated only regarding conditions for spatial dissemination. Going beyond these limitations, more recently Ewing *et al.* (2017) introduced an age-specific spatial metapopulation model to assess how behavioural changes associated with winter holiday impact the flu season. They applied the model to the United States, and the model integrates data on travel behaviour, whereas mixing is assumed from estimates available from Europe and adapted to summer holiday changes measured in the UK during the 2009 pandemic (Eames *et al.*, 2012). Their findings identify changes in mixing patterns as the key element responsible for the epidemic effects induced by holidays.

Here we built a discrete stochastic age-specific spatial metapopulation model at the municipality level, based on demographic, mixing and mobility data of Belgium. Metapopulation models are composed of patches or subpopulations of the system, connected through a coupling process generally describing hosts mobility. We consider two age classes, children (c, below 19 years) and adults (a, 19 years or older), based on the modelling framework introduced in Apolloni *et al.* (2013). Infection dynamics occur inside each patch, driven by the contacts between and within these two classes and quantified from the Belgian data in the POLYMOD Study (Section 2.2) as follows:

$$\bar{\mathbf{C}} = \begin{pmatrix} \bar{C}_{cc} & \bar{C}_{ca} \\ \bar{C}_{ac} & \bar{C}_{aa} \end{pmatrix}, \tag{5.4}$$

where the element ij $(i = a, c,\ j = a, c)$ is an adaptation of the social contact rate matrix as described in Section 1.2.3.2 in order to obtain a normalized contact matrix which is scale-invariant, and calculated in the following way:

$$\bar{C}_{ij} = \frac{m_{ij}}{w_j} w_{tot}. \tag{5.5}$$

m_{ij} denotes the average number of contacts made by a participant in age class i with individuals in age class j, w_j is, as in Section 1.2.3.2, the population of age class j and w_{tot} is the total population of Belgium. We define $\bar{\mathbf{C}}$ at the national level and assume it is the same throughout the country, with the patch demography altering the number of contacts exclusively. The estimated matrices are given in Section 5.B.

The spatial spread occurs via the mobility of individuals as they couple patch p to patch q denoted by $f_{pq}(i)$ for age class i. We consider commuting data across patches from the 2001 Socio Economic Survey of the Belgian Census (INS [producer] and Directorate General Statistics and Economic Information (DGSEI) [distributor], 2006) to describe the regular mobility of individuals for school/work during a regular weekday. Data are not age-specific, so we extract the commuting fluxes per age class based on the probability of children (adults) of commuting on a given distance computed on the French commuting data (INSEE [producer] and Centre Maurice Halbwachs (CMH) [distributor], 1999). We base this inference on the assumption of a similar mobility behaviour across the two neighbouring countries. We do not consider air travel due to negligible internal air traffic within the country. Regarding the movement variations of children, they do not commute during weekends and holidays as schools are closed during those days. We consider adults to continue commuting during holiday weekdays, assuming that adults' time off of work to be homogeneously distributed throughout the year. Concerning adult mobility during the weekends, we estimated the travel fluxes reductions based on statistics available for France (SOeS [producer] and Centre Maurice Halbwachs (CMH) [distributor], 2008), based on the mentioned assumptions. The resulting reduction for adults on Saturday is 0.4053 and on Sunday 0.1473.

Figure 5.13 illustrates the combination of the mixing and mobility information. Both processes are modelled explicitly with a discrete and stochastic approach. As explained above, we base mixing and mobility on Belgian data, furthermore the model follows the time evolution of the 2008/2009 school calendar. This school calendar includes different holidays during which schools are closed. The holidays are, in chronological order: Fall holiday, the public holiday of the first of November, Christmas holiday, winter holiday, Easter holiday, three long weekends in May and June and the summer holidays. The specific dates of each of the holidays can be found in Section 5.B. The model includes 589 patches corresponding to the 589 municipalities of Belgium. As is evident from the presented changes in mobility and mixing, weekends and school holidays are explicitly accounted for in the model and are based on data. The infection disease progression is described through a Susceptible-Exposed-

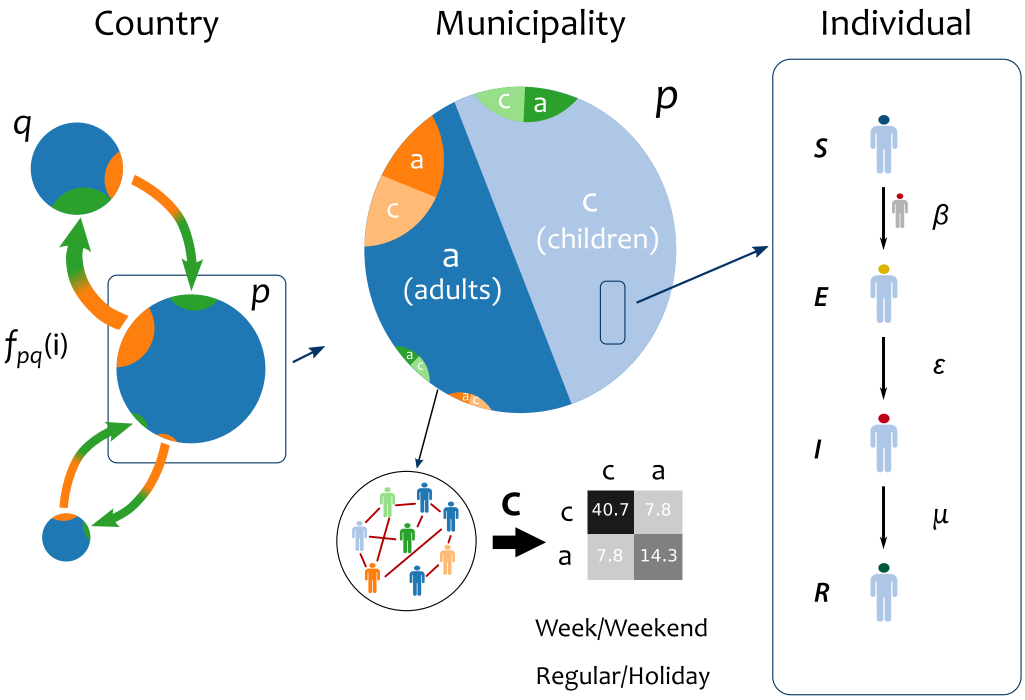


Figure 5.13: Schematic illustration of the spatial age-structured metapopulation model. The metapopulation modelling scheme is composed of three layers. At the country scale, we model Belgium as a set of patches (here indicated with q and p) corresponding to municipalities coupled through mobility of individuals $f_{pq}(i)$ of age class i at time t. Within each municipality, we divided the population into two age classes, children (c) and adults (a), whose mixing pattern we defined by the contact matrix $\boldsymbol{C}$. The figure reports as an example the contact matrix of a regular weekday (Eq. (5.B.1)). Mobility and mixing vary based on the calendar day (regular/holiday, week/weekend). A Susceptible-Exposed-Infectious-Recovered compartmental scheme models the influenza disease progression at the individual level, with β indicating the per-contact transmission rate, ϵ the rate from exposed to infectious state, μ the recovery rate. Figure from De Luca *et al.* (Submitted).

Infectious-Recovered (SEIR) model (Figure 5.13, Section 1.2.2), with a per-contact transmissibility rate β (Note: this is q from previous notations and not β_{ij}), an average latency period of ϵ^{-1}=1.1 days (Boëlle *et al.*, 2011; Carrat *et al.*, 2008) and an average duration of μ^{-1}=3 days (Boëlle *et al.*, 2011; Carrat *et al.*, 2008). A fraction of children (g_c) and adults (g_a) are considered immune to the disease at the beginning of the influenza epidemic, based on available knowledge on prior immunity and vaccination coverage in the country for the 2008/2009 influenza season and prior seasons ($g_c = 39.87\%$, $g_a = 53.19\%$) (Yang *et al.*, 2015; Beutels *et al.*, 2013; Tafforeau, 2008). The force of infection for a susceptible individual of age class i ($i = a, c$) in a given patch p is calculated similar to Eq. (1.12) in Section 1.2.3.3 but with slight adaptations:

$$\lambda(i, p, t) = \beta \sum_j \bar{C}_{ij}(t) \frac{I_j^p(t)}{N^p(t)}. \tag{5.6}$$

This quantity changes with time for two reasons. The first is the school calendar, distinguishing between regular weekdays, regular weekends, holiday weekdays, holidays weekends, and accounted for by $\bar{C}_{ij}(t)$. The second is the mobility of individuals. At a given day of the simulation, each patch p may include: non-commuting residents of p, commuters from neighbouring patches for school/work, commuting residents of p after school/work.

We simulate the model by discretizing time in time steps of 0.5 days, with one time-step corresponding to the activities of a work day (i.e. commuting and social mixing) and the other to the activities performed out of that timeframe (i.e. social mixing), as typically done in agent-based epidemic models (Merler *et al.*, 2011). Binomial processes describe the influenza transmission within each patch. In short, we extract the numbers of individuals in each age class entering the E, I and R compartment from a binomial distribution. Starting from the initial conditions set by ILI surveillance data, we perform 2000 stochastic runs for each model under study.

We calibrate the model to the 2008/2009 influenza season and describe it in more detail in Section 5.B. To assess the impact of variations of contacts and mobility due to school closure, we compare different models. We refer to the model based on the Belgian school calendar with changes in mobility and contacts as the *realistic* model, the *travel changes* model includes only mobility variations over the calendar and keeps social mixing as on a regular weekday; the *mixing changes* model includes only social mixing variations over the calendar and keeps mobility fixed as on a regular weekday; and the *regular weekday* model fixes both mobility and social mixing as on a regular weekday.

To assess the role of each school holiday period (not long weekends) we remove each period one by one. The names of these model are *w/o name holiday* and imply that regular weekdays and regular weekends replaces this holiday. Furthermore, we tested the *w/o holiday* model where regular week structures replace all holiday periods. Next, the *Christmas holiday extension* models consider a scenario where we extended the Christmas holiday with one week, before it started or after its end.

To assess the interplay between the timing of the epidemic and that of the holiday periods, we consider anticipation models with the start of the influenza epidemic anticipated 4 or 2 weeks before the onset of the *realistic* model calibrated on empirical data. We also consider delay models which are as the anticipation models but with a delay of 2 or 4 weeks. Hence in these models, only the onset of the epidemic is altered.

We analyse the spatial distribution of the force of infection determined by the demographic profile in space. To do so, we study the distribution of the reproductive number per patch R^p. We do this calculation for the four types of days considered regarding their variations of social mixing, namely regular weekday, regular weekend, holiday weekday, and holiday weekend before the start of the epidemic but taking prior immunity (g_c,g_a) into account.

We do the validation of the model at the district level using the peak difference (Section 5.B). We compare the various scenarios to the *realistic* model by using the following measures: the peak time difference per patch p ΔT^p, the peak incidence relative variation per patch p ΔI^p and the epidemic size relative variation per patch p $\Delta\sigma^p$. In Section 5.B we describe these measures more extensively. We synthesise using medians and 50%, 95% CIs at the patch levels.

Season 2008/2009 shows an ILI incidence that reaches its peak in week 5 of 2009, both in the Brussels district and on the national scale. The incidence is visibly slowed down during the Christmas holiday (Figure 5.14 (a) & (b)), suggesting that holiday periods may have a measurable effect on transmission. The model seems to reproduce the empirical influenza spreading pattern well. The peak time is found to be within one week of the empirically observed time for 76% of the districts, and within two weeks for 90% of them. Only two districts in the province of Luxembourg showed greater discrepancies (four weeks), being at the edge of the country (Figure 5.B.3).

The average patch reproductive number is estimated to be $R = 2.12$, corresponding to $\beta = 0.0850$ ($[0.0674, 0.0858]$ 95% CI) of the per-contact transmissibility obtained from the calibration procedure. The demographic profile of the population and the immunity profile of the patch lead to the variation of R^p at the patch level. Besides, it depends on the day type considered, whether regular or during a holiday, whether during the week or the weekend (Figure 5.14). Larger variations and higher values are obtained for a regular weekday, having the highest number of contacts, compared to less heterogeneous distributions and smaller R^p values in the other cases. The reproductive number per patch is lowest for the holiday weekend, corresponding to the reduced mixing. Changes in individual behaviour induced by weekends and holidays are found to considerably alter the epidemic dynamics leading to a considerable delay of the peak time (median of 3.7 weeks across patches, *regular weekday* model compared to the *realistic* model, Figure 5.15), and smaller peak time incidence (33% median relative change) and total epidemic size (11% median relative change). When considering the variations in social mixing or mobility in isolation, we find social mixing to be mainly responsible for the effects just described. The *travel changes* model is indeed comparable to the *regular weekday* model, whereas neglecting changes in mobility (*mixing changes* model) produces epidemic patterns very similar to the *realistic* model (zero median variations).

Cumulatively, all holidays concur to delay the peak time of 1.7 weeks and to reduce the epidemic size of approximately 2%, with a reduction of the peak incidence (4%, all median values across patches). Among all holiday periods, the Christmas holiday produces the largest effect, responsible for the overall reduction of the epidemic size and a peak delay of about one week. The early break of Fall leave has a negligible impact instead. Winter holiday leads to a minimal reduction of the epidemic size (median of 1%), but no effect on the peak timing or peak incidence. The impact of Easter holiday is negligible on all indicators. By comparing the effect of the *regular weekday* model (Figure 5.15) with the one of the *w/o holiday* model, both on the *realistic* model, we find that weekends have a major effect on slowing down the epidemic curve: a difference of $\Delta T^p = -3.7$, $[-3.9, -3.6]$

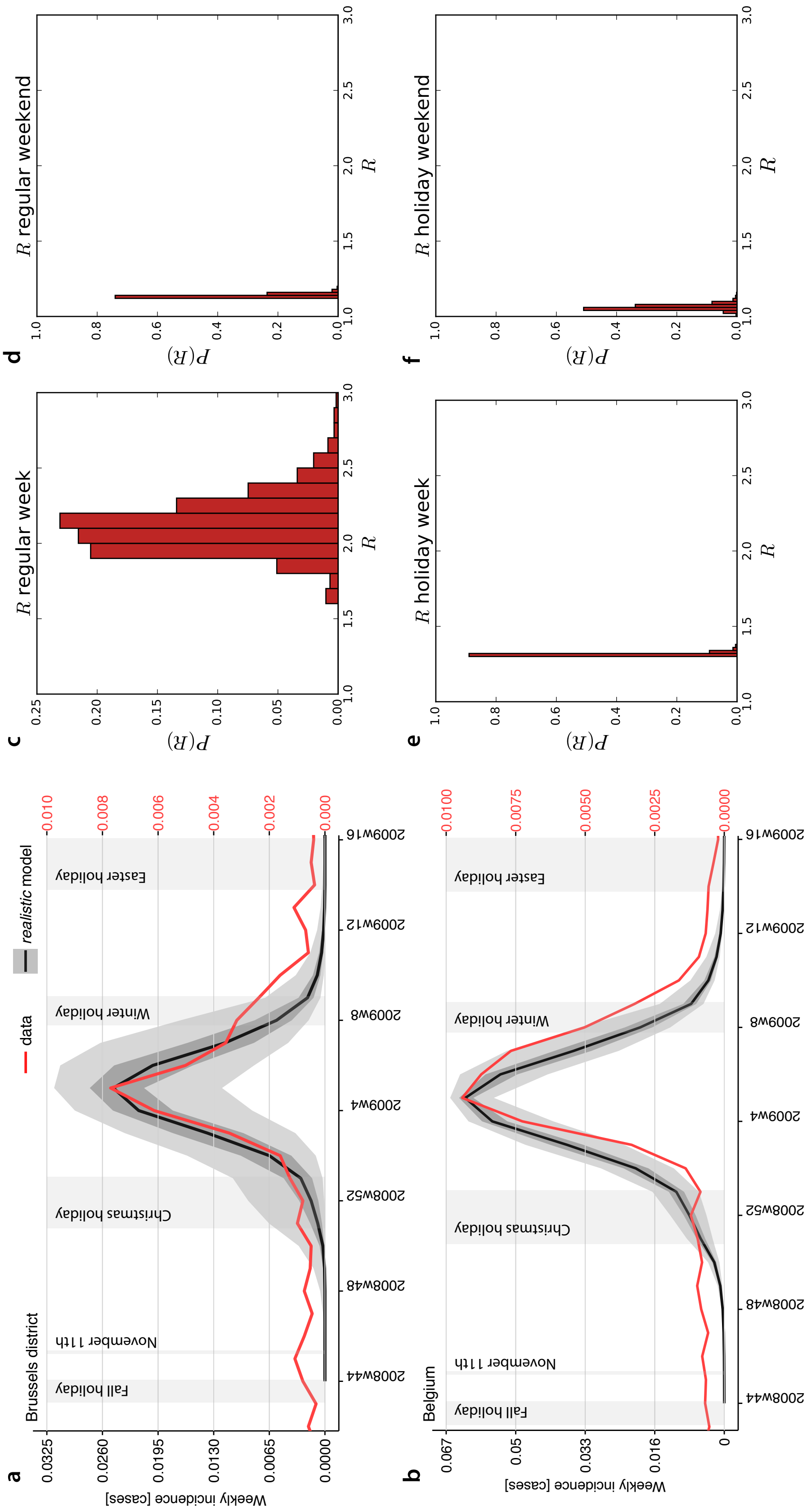


Figure 5.14: Calibration results. (a)-(b): Simulated and empirical incidence curves for the district of Brussels (panel a) and the entire Belgium (panel b). The incidence curve of Brussels is the sole empirical data used for the calibration of the model. Different incidence axis between simulated vs. empirical data refer to GP consultation rates that are not explicitly accounted for in our model, due to unavailable estimates. (c)-(f): Probability distribution of the values of the reproductive number R^p computed in each patch following the calibration. They refer to the different day types explored, i.e. belonging to a regular week (panel c), regular weekend (d), holiday week (e), holiday weekend (f). Figure from De Luca *et al.* (Submitted).

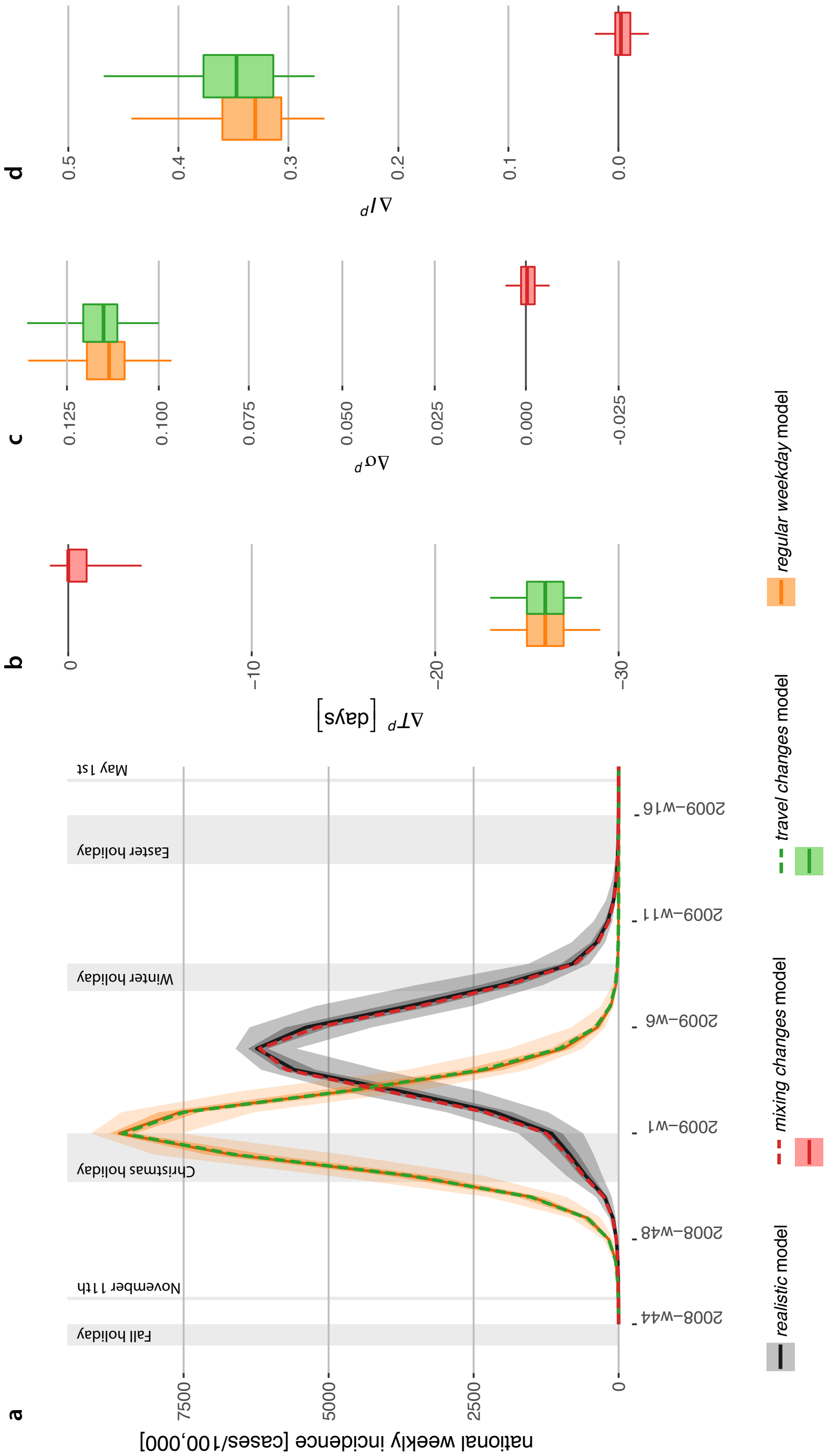


Figure 5.15: Role of social mixing vs. travel behaviour. (a): Simulated weekly incidence profiles for influenza in Belgium. The *realistic* model is compared to the *travel changes* model, the *mixing changes* model, the *regular weekday* model. Median curves are shown for all cases, along with 50% confidence intervals (dark shade) and 95% CI (light shade), for the *realistic* and *regular weekday* model (they are not shown for the other models for the sake of visualization). (b)-(c)-(d): Peak time difference ($\Delta T^p = T^p_{scenario} - T^p_{realistic\ \mathrm{model}}$), relative variation of epidemic size ($\Delta\sigma^p = \left(\sigma^p_{scenario} - \sigma^p_{realistic\,\mathrm{model}}\right)/\sigma^p_{realistic\,\mathrm{model}}$), and relative variation of peak incidence ($\Delta I^p = \left(I^p_{scenario} - I^p_{realistic\,\mathrm{model}}\right)/I^p_{realistic\,\mathrm{model}}$), respectively, across the three experimental scenarios. Boxplots refer to the distributions across patches. Figure from De Luca *et al.* (Submitted).

weeks when we consider no weekends compared to $\Delta T^p = -1.7$, $[-1.9, -1.2]$ weeks when we include them.

The extension before Christmas holidays does not impact the resulting epidemic. If we consider the additional week of leave after the break, no changes to the epidemic timing are observed, however size and peak incidence decrease of 4% (median values), respectively.

The anticipation and delay models allow us to model various epidemic profiles characterised by several degrees of early and late seasons, isolating the calendar effect but keeping the same season-specific conditions. We make all comparisons with the *realistic* model. The earliest epidemic ($-4w$ model) shows the strongest impact, reporting a median anticipation of more than one week on the *realistic* model (once discounted for the earlier start) and a median reduction of the peak incidence of about 10%. All other epidemics are rather similar to the realistic one, except for the $-2w$ model reporting a considerable reduction of the peak incidence (median of approximately 13% across patches). Besides, it is important to note that, differently from previous effects, the anticipation or delay of the season leads to a considerably larger variation of the simulated epidemic indicators across patches, signalled by larger confidence intervals.

In summary, we have shown that a spatial metapopulation model for the transmission of the 2008/2009 influenza season in Belgium, combining contact and mobility data and their changes during weekends and holiday periods, can reproduce the empirical pattern observed with a reasonably good agreement. This suggests that data-driven mixing and mobility are crucial ingredients to capture spatial dynamics (Merler *et al.*, 2011; Apolloni *et al.*, 2013; Eames *et al.*, 2012; Viboud *et al.*, 2006; Balcan *et al.*, 2009a; Ewing *et al.*, 2017; Brownstein *et al.*, 2006; Crépey and Barthélemy, 2007; Charaudeau *et al.*, 2014). The local aspect of this model is captured by the mixing which is patch-dependent. This local aspect induces reproduction numbers close to 1 or not at all depending on the patch. Mobility transfers the epidemic to other patches, hence resulting in peak times which can differ with one week to the empirical peak time for most patches. However, patches on the border display a delay in peak time, predominantly due to the lack of mobility from other countries in our model, for the sake of simplicity. We expect this border effect to be increasingly negligible for larger countries.

The resulted model shows that the Christmas break results in temporary reductions in the influenza transmission, as reported by the sentinel surveillance and observed in other empirical studies (Cauchemez *et al.*, 2008; Ewing *et al.*, 2017). We have confirmed the results in Ewing *et al.* (2017) being a delay of the epidemic when mixing changes are adapted and no noticeable effect of travel changes compared to a model without school closure. In contrast to this work, we have shown that an essential mitigation of the epidemic impact at peak time also occurs.

We have shown a more present effect of mixing changes, induced by a reduction in mixing between children, which results in a decrease in transmission potential. We expect travel not to act on the transmission potential as it is merely a coupling force and have seen this by the smaller, even negligible, effect of travel changes (as in Ewing *et al.*, 2017). In addition to information from influenza surveillance systems, which are on weekly basis, we observe school closure during weekends to periodically dampening the transmission.

In the 2008/2009 influenza season, we have found that the Christmas holiday, occurring during the growing phase of the epidemic, has the largest impact in timing and epidemic size compared to the other holidays. By using different scenarios, we have found that the Christmas break would have the strongest impact on a very early season when school closing would occur at (before) or around the epidemic peak. Our investigation shows that it is hard to infer the role of holiday timing from a few examples and other breaks beyond the Christmas break (Ewing *et al.*, 2017) may have an important mitigating impact. Furthermore, the effect of holidays in a season is different from just taking the sum of the effects of each holiday period on its own, so the full calendar needs to be considered. Our findings help to shed light on previous empirical findings showing no clear pattern for the effects of school closure on peak incidence or total epidemic size, comparing closure before and after the peak (Jackson *et al.*, 2013).

In addition to school breaks already occurring in the calendar, we have also explored a possible extension of Christmas holiday of one week. Previous work analyzing the length of school closure found that two weeks or more appear to be enough to result in a recognizable effect (Wu *et al.*, 2010; Heymann *et al.*, 2004; Cauchemez *et al.*, 2008; Fujii *et al.*, 2002; Baguelin *et al.*, 2010), whereas shorter closures may not be beneficial or may not have an obvious impact (Danis *et al.*, 2004; Cowling *et al.*, 2008; Johnson *et al.*, 2008; Rodriguez *et al.*, 2009; Calatayud *et al.*, 2010). Our synthetic results show that the extension would be advantageous only if implemented after the Christmas break, with a mitigation of the peak incidence and a small peak delay of a few days. Here we decided to test the extensions as a regular closure given that – in a broader context – authorities in Belgium are currently discussing whether to modify the school calendar for pedagogical reasons: the aim would be to reduce summer holidays and redistribute holiday periods throughout the year (Glorieux, I. and Vandeweyer, J., 2011). We have found that an extension of the Christmas holiday would be beneficial in the management of the influenza season potentially mitigating its epidemic impact.

We obtained our findings on seasonal influenza, and modelling works also recovered the results on peak delay on synthetic influenza pandemics considering reactive school closure (Merler *et al.*, 2011). The straightforward extension of our conclusions to the pandemic case faces, however, several challenges. First, effects induced by school closure may be specific to the particular epidemic profile, and therefore they may lead to different results depending on the pandemic under consideration (Cauchemez *et al.*, 2009b). Second, the nature of the school closure may alter the behaviour of individuals during that period. In our study, we have considered holidays that are regularly planned in the school calendar and associated to specific social activities (e.g. vacation trips, family visits and others), for which contact data are available (Hens *et al.*, 2009a). One can envision school closure during an influenza pandemic as a proactive or reactive measure to the ongoing outbreak. Not being planned, it is expected to have a stronger disruptive impact on the social mixing of individuals on the short term compared to regular closure. On the other hand, Cauchemez *et al.* (2008, 2009b) argued that a prolonged closure may limit the reduction of contacts on the long course, because of costs and logistics, and reduction in compliance rate. Having shown here that changes in social mixing represent the single element critically responsible for the impact of school closure on the epidemic outcome, we note that modelling results on school closure in the case of a pandemic would strongly be affected by assumptions considered for mixing changes, in the absence of data.

While a large body of literature has recently focused on behavioural changes during an epidemic (Funk *et al.*, 2010; Perra *et al.*, 2011; Rizzo *et al.*, 2013; Funk *et al.*, 2015; Verelst *et al.*, 2016), still little is known to quantify them (Jackson *et al.*, 2011; Hens *et al.*, 2009a; Eames *et al.*, 2010, 2011, 2012; Zhang *et al.*, 2011, 2012; Chen and You, 2015; Luh *et al.*, 2016). Our work focused on Belgium, as a rather detailed survey was conducted in the country to estimate contact rates in the population of different age classes at different periods of the calendar year (Hens *et al.*, 2009a). These estimates constitute the input data to parameterize our spatial modelling framework. Modelling approaches to study epidemics in settings where no data exist often assume that mixing would reduce following school closure and import estimates available from other circumstances or epidemiological contexts (Merler *et al.*, 2011; Fumanelli *et al.*, 2016; Ewing *et al.*, 2017). This assumption may lead to several issues. Contacts and their changes along the calendar may be country-specific (Mossong *et al.*, 2008), thus affecting epidemic results when applied to a different context. Finally, such overall reduction does not allow to parametrize a contact matrix entirely. Such evidence does not support the parameterization of mixing changes from different countries and/or epidemic situations (e.g. seasonal vs. pandemic) (Ewing *et al.*, 2017). The reduction is expected to be heterogeneous across mixing groups, because of compensatory behaviours (e.g. children drastically reduce child-to-child contacts but increase child-to-adult contacts during holidays) (Hens *et al.*, 2009a; Eames *et al.*, 2011). Assumptions on the relative role of specific age classes in the absence of data may lead to biases in the modelled epidemic outcome, especially for epidemics reporting large differences in attack rates in children vs. adults. Our work highlights the need to expand our knowledge of contacts and associated changes induced by social activity or by the epidemic itself, to better parameterize models and provide reliable and accurate results for epidemic management.

Our study has a set of limitations; firstly the host population is divided into two classes only. While a larger heterogeneity is known for the distribution of contacts across age classes (Mossong *et al.*, 2008), our approach still accounts for the significant role of children vs. adults in the spread of the disease. Moreover, the validation analysis shows that considering children and adults and the associated mixing and travel behaviour is enough to reproduce the spatio-temporal unfolding of the epidemic to a good accuracy.

The study is focused on one season only, the 2008/2009 influenza season. We can clearly include additional seasons in the analysis, however, our choice aimed at discounting season-specific effects to avoid uncertainties and discordance found in previous works. Also, we argue that the main effect behind the observed impact is in the interplay between the incidence profile and holidays timing, all other aspects being equal. To fully assess this issue, we have systematically explored earlier and later epidemics than the 2008/2009 season, thus synthetically accounting for other (similar) influenza seasons.

Travel statistics informed mobility changes from commuting during regular weekdays to non-regular travel during weekends. We lack however specific data on travel behaviour for adults during school holidays. We have therefore assumed that adults would continue commuting during holiday weekdays. While we expect that a fraction of adults would stop commuting at least for a few days during breaks as they take time off work, we expect this change in travel fluxes (compensated by additional trips to visit families (Ewing *et al.*, 2017)) to have a negligible effect on the simulated epidemic. More drastic changes to travel, i.e. fully neglecting travel changes as in the *mixing changes* model, indeed did not alter the resulting epidemic.

To conclude, we have shown, with a data-driven spatial metapopulation model calibrated on the 2008/2009 influenza season in Belgium, that regular school closure considerably slows down influenza epidemics and mitigate their impact on the population, because of changes in social mixing that are empirically measured. This finding may help the management of epidemics and lessen the pressure on the public health infrastructure. The effect is due to both school holidays and weekend closures, the latter periodically dampening transmission. Variations in travel behaviour do not lead instead to visible effects. The observed impact strongly depends on the timing of the school closure, and to a lesser extent on its duration. Christmas holiday is the school break playing the most important role in mitigating the epidemic course, though variations are observed depending on the influenza season (e.g. early vs. late epidemic). The addition of one week after the Christmas holiday may represent an additional strategy to delay the epidemic peak further and mitigate its impact.

5.A Additional details to the social contact dispersal

5.A.1 Process to link contact data with time use data

Some participants (197) failed to report at least once a location in their time use data and sometimes also the distance was not reported. This makes a proper linking impossible and hence these participants will be removed from the current data and analysis.

Step 1: Short list of TUD

By taking consecutive time blocks with the same location and distance together in one larger time block, double information is reduced. As a result a shorter format of the Time Use Data is obtained with new information for each entry. This reduction in information facilitates the linking process, as one can easily deduce the number of times a certain location is visited. Furthermore, whenever people entered home as location but did not fill in the distance or entered a distance category different from 0-1 km from home, the distance was changed to 0-1 km from home.

Step 2: Split contacts per location

Contacts occurring at multiple locations are split in different entries (*subcontacts*), one for each location where the contact occurred. Hence, we obtain the same contact information multiple times with one unique location. To keep track of the number of contacts, a contact id (cnt_id) is inserted, hence repeated for each of the duplicated lines. For programming reasons the following variables are added:

- number of times in TUD - gives the number of times the location occurs in the time use data during the entire day,
- number of locations in TUD - gives per contact the number of locations mentioned in the time use data. Hence this number can be smaller than the number of locations for a contact.

Step 3: Removal of certain groups of participants

During the process in Step 4 we assume that participants filled in the contacts in a chronological way, as requested. However, for 8 participants this was not true and thus 7 were removed at this step, manually linked and added later on (Step 5). The other one is removed from the data as he/she did not fill in the diary in a logical way.

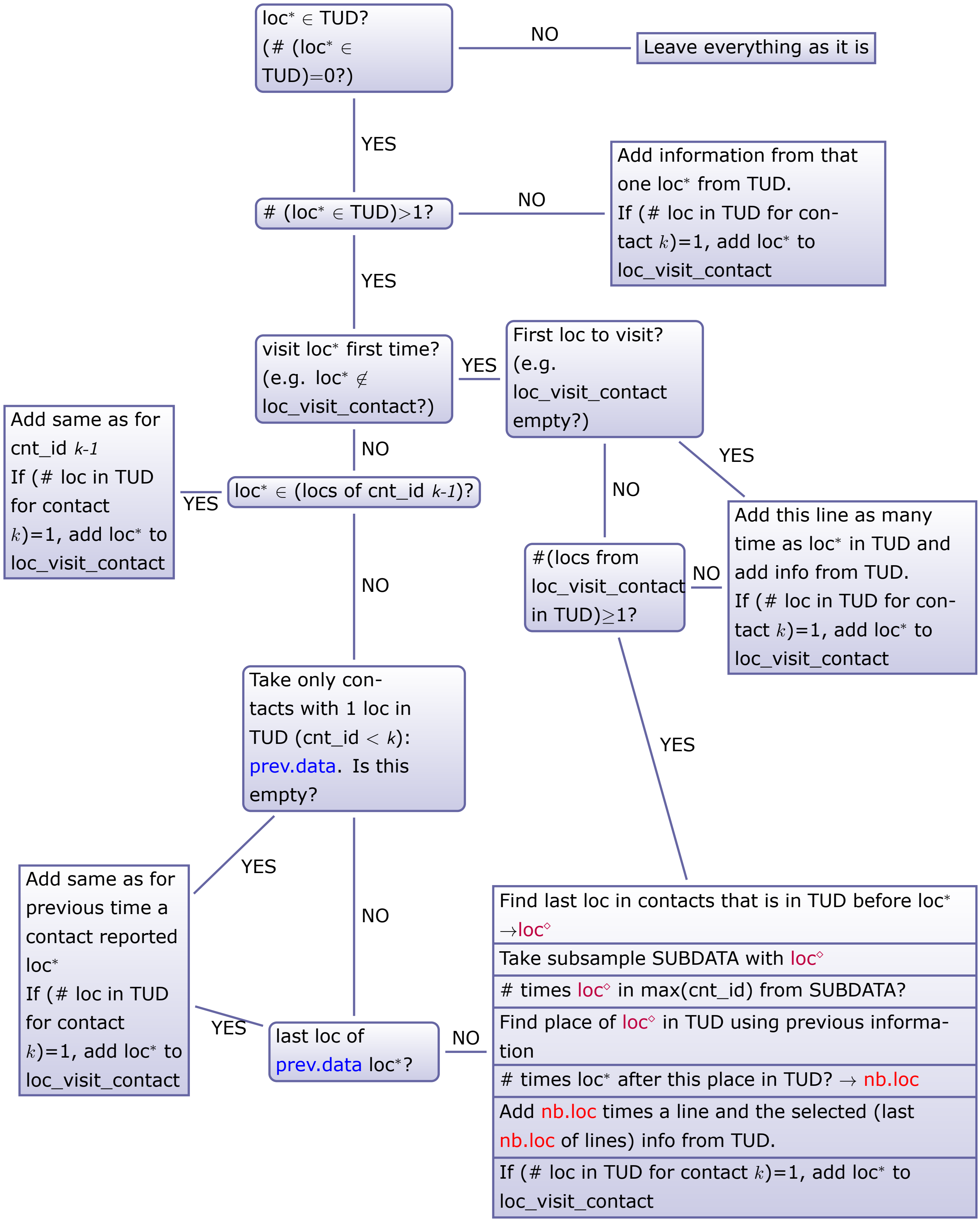
Step 4: Linking process

The linking process for every location (loc*) of every contact k, is described in the following flow chart, using the following notations.

- loc* - the location for which you want to make a link (for a contact)
- TUD - time use data
- cnt_id - unique ID per contact
- loc$^\diamond$ - last location, before loc*, that is in time use data for a specific participant

- prev.data - data before this (location of a) contact with locations occurring only once in the time use data
- SUBDATA - subset of the contact data for a participant which happened before the current contact at location $\text{loc}^{\diamond}$
- nb.loc - number of times loc^{*} can be found in the time use data after a certain point

Initialize loc_visit_contact to be empty at the start of every participant i, meaning that for that participant no locations has been visited yet. For each line of the contact, hence a *subcontact*, a first check occurs: Is loc^{*} work and is this added due to too many work contacts? If yes, all available information about work from the time use is entered to the data. If no, the process runs as described in the following flow-chart.


loc* $\in$ TUD? (# (loc* $\in$ TUD)=0?)
NO
Leave everything as it is
YES
(loc* $\in$ TUD)>1?
NO
Add information from that one loc* from TUD. If (# loc in TUD for contact k)=1, add loc* to loc_visit_contact
YES
visit loc* first time? (e.g. loc* $\notin$ loc_visit_contact?)
YES
First loc to visit? (e.g. loc_visit_contact empty?)
YES
Add this line as many time as loc* in TUD and add info from TUD. If (# loc in TUD for contact k)=1, add loc* to loc_visit_contact
NO
#(locs from loc_visit_contact in TUD)$\geq$1?
NO
YES
NO
loc* $\in$ (locs of cnt_id k-1)?
YES
Add same as for cnt_id k-1 If (# loc in TUD for contact k)=1, add loc* to loc_visit_contact
NO
Take only contacts with 1 loc in TUD (cnt_id $<$ k): prev.data. Is this empty?
YES
Add same as for previous time a contact reported loc* If (# loc in TUD for contact k)=1, add loc* to loc_visit_contact
NO
last loc of prev.data loc*?
YES
NO
Find last loc in contacts that is in TUD before loc* $\rightarrow$loc$^\diamond$
Take subsample SUBDATA with loc$^\diamond$
times loc$^\diamond$ in max(cnt_id) from SUBDATA?
Find place of loc$^\diamond$ in TUD using previous information
times loc* after this place in TUD? $\rightarrow$ nb.loc
Add nb.loc times a line and the selected (last nb.loc of lines) info from TUD.
If (# loc in TUD for contact k)=1, add loc* to loc_visit_contact

Step 5: Manually linking

For 7 subjects the linking process was done manually following the same process, but allowing for the fact that perhaps travel could occur earlier than in the time use data, due to a short stay there. After the linking by hand, the new data is added to the data obtained in Step 4.

Step 6: Calculate denominator

To ensure that each original contact only contributes as 1 contact in future analysis, weights are given to the *subcontacts*. The denominator of the weights is the sum of the time lags for a certain contact. For those 4583 (out of 30128) unique contacts with locations not in the time use data this information is not available and hence they are ignored in this sum.

Step 7: Calculate weight

In order to account for the uncertainty in the linking process, weights are given for each *subcontact*. This weight is the time lag at the location of the *subcontact* over the denominator of Step 6. For locations not in the time use data, the weight is set to 0.

5.A.2 Distance travelled for all participants

Table 5.A.1: Patterns travelled over distance based on time use data of all participants (n=1773). The row frequencies and percentages show the number and proportion of people exhibiting a particular travel behaviour over distance throughout the day.

	D	Frequency	Percent
1	(1, 0, 0, 0)	367	20.69
2	(1, 0, 0, 1)	23	1.30
3	(1, 0, 1, 0)	299	16.85
4	(1, 0, 1, 1)	31	1.75
5	(1, 1, 0, 0)	612	34.50
6	(1, 1, 0, 1)	12	0.68
7	(1, 1, 1, 0)	342	19.28
8	(1, 1, 1, 1)	14	0.79
9	(0, 1, 0, 0)	5	0.28
10	(0, 1, 0, 1)	0	0
11	(0, 1, 1, 0)	0	0
12	(0, 1, 1, 1)	1	0.06
13	(0, 0, 0, 0)	0	0
14	(0, 0, 0, 1)	11	0.62
15	(0, 0, 1, 0)	7	0.39
16	(0, 0, 1, 1)	1	0.06
17	(NA, NA, NA, NA)	49	2.76

5.A.3 Observed mean of the distance-degree density by age

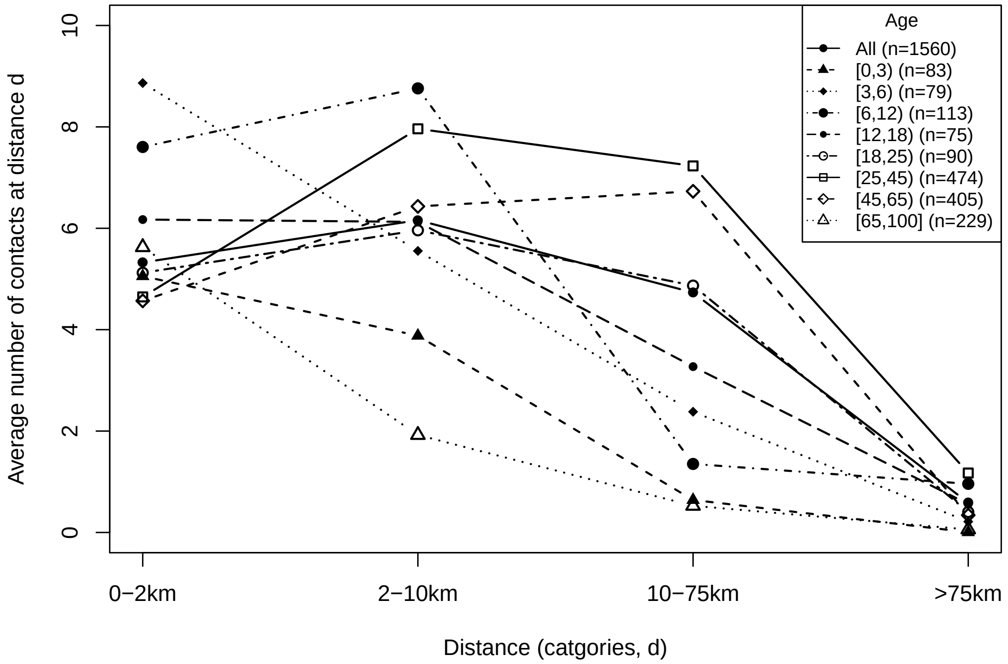


Figure 5.A.1: The average number of contacts at distance d ($\bar{Y}_{d\cdot}$) are connected by (full) lines for all participants. The other line types distinguish for the various age classes. Discrepancies are present for the different age classes.

5.A.4 Degree distribution over distance for different situations (Tables)

Table 5.A.2: The total number of contacts over the various distances stratified by age class and week/weekend.

Age class	Week/ Weekend	Nb. data-lines	Nb. part	Nb. cnt	0-1km	2-9km	10-74km	75km or more
$[0,3[$	week	1024	60	648.30	325.80	286.68	35.41	0.42
$[3,6[$	week	1623	55	1121.00	583.16	392.79	143.00	2.05
$[6,12[$	week	2672	83	1717.08	698.81	898.95	68.96	50.35
$[12,18[$	week	1569	55	1016.67	359.76	411.87	203.05	42.00
$[18,25[$	week	1657	70	1194.00	310.07	460.29	403.13	20.50
$[25,45[$	week	12608	349	8293.41	1789.18	3050.19	2977.33	476.71
$[45,65[$	week	9753	316	6032.20	1453.46	2063.01	2419.36	96.37
$[65,100[$	week	2399	188	1542.87	1019.80	397.93	111.14	14.00
$[0,3[$	weekend	265	23	147	93.78	35.43	17.79	0
$[3,6[$	weekend	403	24	223	117.17	45.90	45.11	14.82
$[6,12[$	weekend	789	30	393.17	160.45	90.92	83.76	58.04
$[12,18[$	weekend	345	20	193	103.14	47.57	42.29	0
$[18,25[$	weekend	487	20	278	150.92	76.10	35.13	15.85
$[25,45[$	weekend	2827	125	1707.72	422.52	739.78	464.18	81.24
$[45,65[$	weekend	2182	89	1322.86	406.13	554.68	319.52	42.53
$[65,100[$	weekend	603	41	385	313.86	57.73	13.41	0

Table 5.A.3: The total number of contacts over the various distances stratified by age class, week/weekend and regular/holiday.

Age class	Week/ Weekend	Regular/ Holiday	Nb. data-lines	Nb. part	Nb. cnt	0-1km	2-9km	10-74km	75km or more
All	week	regular	26848	928	17512.45	5365.73	6118.64	5506.01	514.07
All	weekend	regular	5227	234	3048.74	1161.41	1057.05	710.28	118.00
All	week	holiday	6466	257	4062.08	1174.30	1843.07	855.37	188.33
All	weekend	holiday	2677	141	1604	606.57	591.05	310.91	94.47
$[0, 3[$	week	regular	917	51	570.30	295.08	244.53	30.28	0.41
$[3, 6[$	week	regular	1333	42	935	514.39	353.24	67.37	0
$[6, 12[$	week	regular	2321	66	1527.95	600.89	874.69	52.38	0
$[12, 18[$	week	regular	1374	43	902.67	292.59	402.03	203.05	5
$[18, 25[$	week	regular	1271	55	970	255.95	354.64	359.41	0
$[25, 45[$	week	regular	10033	282	6803.67	1443.13	2319.55	2608.08	431.91
$[45, 65[$	week	regular	7940	248	4745.81	1226.42	1341.59	2113.06	62.74
$[65, 100[$	week	regular	1659	141	1057.06	737.30	228.37	72.39	14
$[0, 3[$	week	holiday	107	9	78	30.72	42.15	5.13	0
$[3, 6[$	week	holiday	290	13	186	68.78	39.55	75.63	2.05
$[6, 12[$	week	holiday	351	17	189.13	97.92	24.27	16.58	50.36
$[12, 18[$	week	holiday	195	12	114	67.17	37	9.83	0
$[18, 25[$	week	holiday	386	15	224	54.13	105.65	43.72	20.50
$[25, 45[$	week	holiday	2576	68	1490.74	346.05	730.64	369.25	44.81
$[45, 65[$	week	holiday	1815	70	1288.39	227.04	721.42	306.30	33.63
$[65, 100[$	week	holiday	746	53	491.82	282.50	169.56	38.76	0
$[0, 3[$	weekend	regular	151	15	90	62.86	20.36	6.77	0
$[3, 6[$	weekend	regular	282	13	147	68.50	27.86	40.82	9.82
$[6, 12[$	weekend	regular	489	18	231.17	119.66	55.61	32.67	23.23
$[12, 18[$	weekend	regular	172	11	102	54.65	30.35	17	0
$[18, 25[$	weekend	regular	386	15	238	136.73	55.47	33.04	12.76
$[25, 45[$	weekend	regular	1789	78	1071.72	272.39	402.16	323.98	72.19
$[45, 65[$	weekend	regular	1500	54	882.86	220.75	419.51	242.60	0
$[65, 100[$	weekend	regular	458	30	286	225.86	45.73	13.41	0
$[0, 3[$	weekend	holiday	114	8	57	30.92	15.07	11.02	0
$[3, 6[$	weekend	holiday	121	11	76	48.67	18.05	4.29	5
$[6, 12[$	weekend	holiday	300	12	162	40.79	35.31	51.09	34.80
$[12, 18[$	weekend	holiday	173	9	91	48.49	17.22	25.29	0
$[18, 25[$	weekend	holiday	101	5	40	14.20	20.63	2.09	3.09
$[25, 45[$	weekend	holiday	1039	48	637	150.13	337.62	140.21	9.05
$[45, 65[$	weekend	holiday	682	35	440	185.38	135.16	76.93	42.53
$[65, 100[$	weekend	holiday	147	13	101	88.00	12.00	0	0

5.A.5 Underlying densities for the social contact dispersal over distance

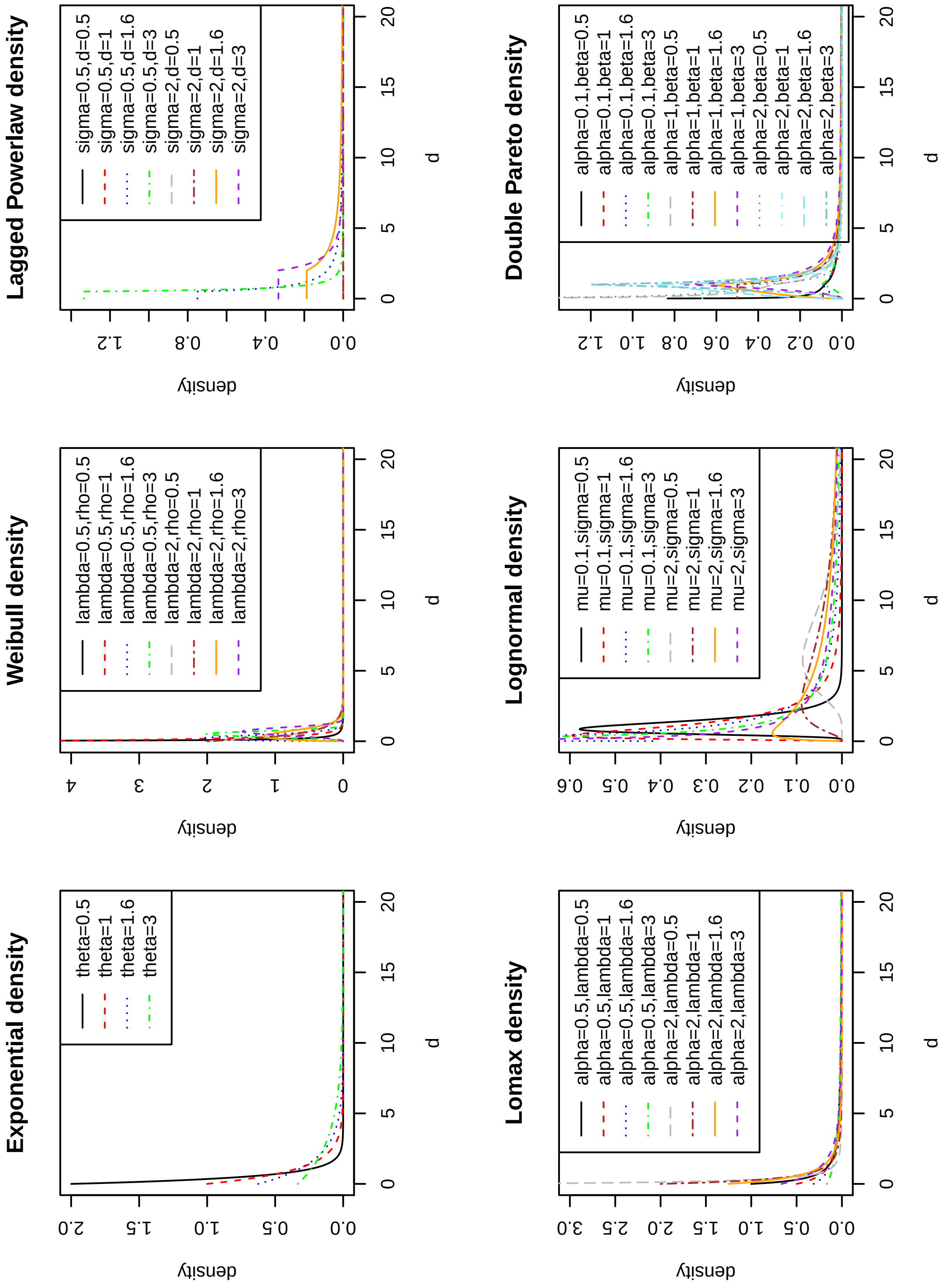


Figure 5.A.2: An overview of the considered continuous distance densities $f(d^*, \boldsymbol{\theta}_t)$. Different parameter values and combinations thereof are shown. The options of these parameters are indicated in the legends and differ for the various densities.

5.A.6 Overdispersion estimates

Table 5.A.4: Overdispersion estimates based on the best models.

Distance category	Lognormal model (age & week)	Saturated model (age*week*reg)
0-2 km	1.56	1.58
2-10 km	0.39	0.40
10-75 km	0.22	0.24
over 75 km	0.014	0.02

5.B Additional details to the metapopulation model

5.B.1 Estimated contact matrices

The contact matrices, based on the data, are given by:

$$\bar{C}^{\textbf{ reg weekday}} = \begin{pmatrix} 40.71 & 7.84 \\ 7.84 & 14.25 \end{pmatrix} \text{ for a regular weekday,} \tag{5.B.1}$$

$$\bar{C}^{\textbf{ reg weekend}} = \begin{pmatrix} 12.51 & 6.00 \\ 6.00 & 10.85 \end{pmatrix} \text{ for a regular weekend,} \tag{5.B.2}$$

$$\bar{C}^{\textbf{ hol weekday}} = \begin{pmatrix} 14.02 & 7.28 \\ 7.28 & 12.29 \end{pmatrix} \text{ for a holiday weekday,} \tag{5.B.3}$$

$$\bar{C}^{\textbf{ hol weekend}} = \begin{pmatrix} 10.89 & 7.20 \\ 7.20 & 8.59 \end{pmatrix} \text{ for a holiday weekend.} \tag{5.B.4}$$

5.B.2 Dates of the school holidays

The school holidays in the 2008/2009 school year, in which schools are closed, were:

- Fall holiday: from October 25 to November 2, including the public holiday of the first and second of November;
- Public holiday of November 11;
- Christmas holiday: from December 20, 2008 to January 4, 2009;
- Winter holiday: from February 21 to March 1;
- Easter holiday: from April 4 to April 17;
- Long weekend: from May 1 to May 3, around the public holiday of May 1;
- Long weekend: from May 21 to May 24, around the public holiday of May 21;
- Long weekend: from May 30 to June 1, around the public holiday of June 1.

From July 1 to August 31 schools are closed for the summer holidays.

5.B.3 Calibration and validation of the model

The model is calibrated to the 2008/2009 influenza season, for which we have surveillance data collected by the Belgian Scientific Institute of Public Health (Van Casteren *et al.*, 2010). Data report new episodes of ILI registered each week by the network of sentinel general practitioners. For every episode, additional information is reported: age group (<5, 5-14, 15-64, 65-84, 85+), hospitalization, antiviral treatment, vaccination status, municipality of residence. Surveillance data on the new number of cases are aggregated at the district level (including several municipalities) to reduce signal noise.

We perform the calibration on Brussels district only, i.e. by comparing the simulated incidence profile of Brussels to the incidence ILI data for that district. We do not consider calibrating the model also in the remaining districts, as these are used for validation.

The model is seeded with the first non-zero incidence value provided by surveillance data per district and accounted for possible sampling biases. We use a bootstap/particle filter Weighted Least Square (WLS) with 20 particles to calibrate our model fixing the epidemiological parameters described for the influenza dynamics and obtain the per-contact transmissibility β. Calibration is performed on normalized incidence curves to discount effects due to unknown GP consultation rates. Additional details can be found in the Supplementary Information of De Luca *et al.* (Submitted).

To validate the model, we look in particular at the *peak difference* per district d per stochastic run r:

$$\Delta T^d(r) = T^d(r) - T^d_{ILI} \tag{5.B.5}$$

where $T^d(r)$ is the peak time of weekly incidence of run r in district d and T^d_{ILI} is the incidence peak reported from surveillance data in the same district. Medians per patch over $2 \cdot 10^3$ stochastic runs are computed.

5.B.4 Comparison measures for the different model scenarios

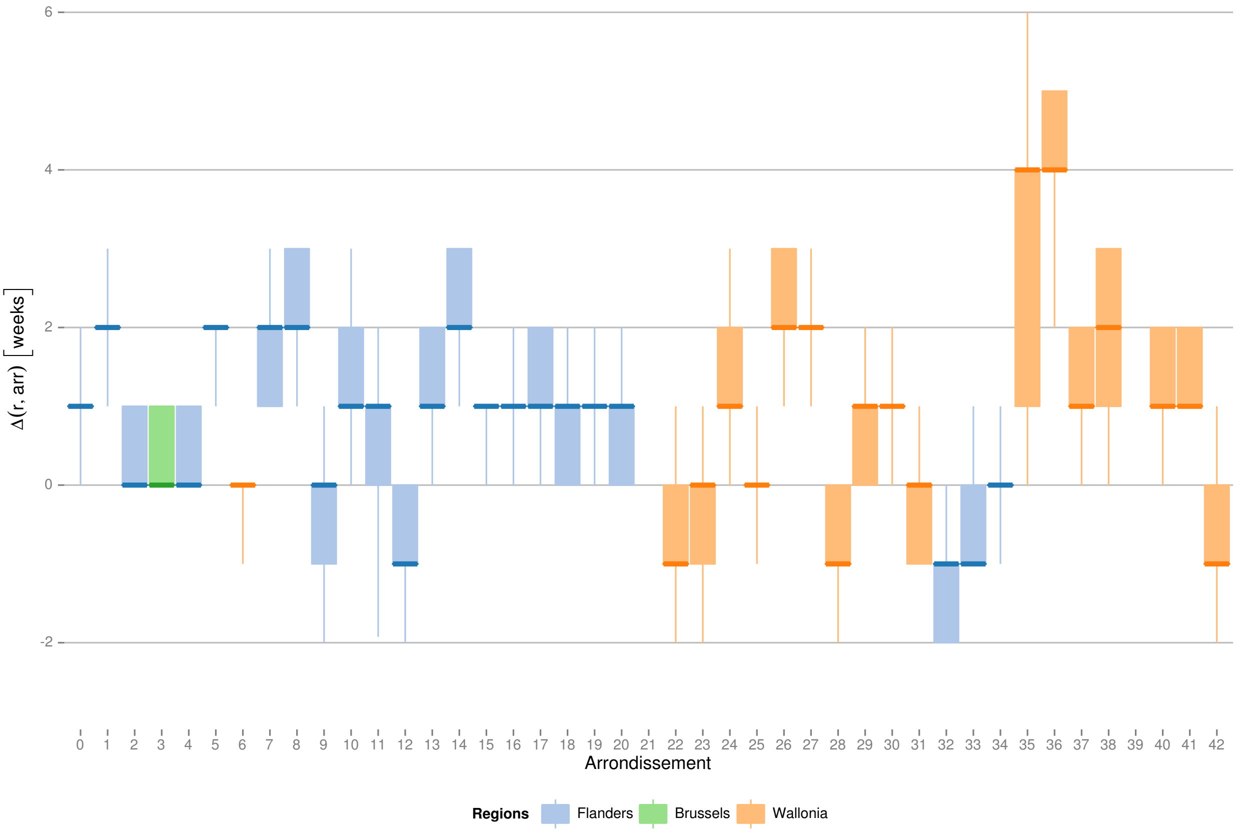
- the *peak time difference* per patch $\Delta T^p = T^p_{scenario} - T^p_{realistic\text{ model}}$, with T^p the median peak time of the incidence curve in patch p computed on all stochastic runs (for both the scenario under study and the *realistic* model);

- the *peak incidence relative variation* per patch $\Delta I^p = (I^p_{scenario} - I^p_{realistic\,\text{model}}) / I^p_{realistic\,\text{model}}$, with I^p the median incidence value at peak time in patch p;

- the *epidemic size relative variation* per patch $\Delta \sigma^p = (\sigma^p_{scenario} - \sigma^p_{realistic\,\text{model}}) / \sigma^p_{realistic\,\text{model}}$, with σ^p the median epidemic size in patch p.

5.B.5 Additional validation results

Figure 5.B.3 shows the comparison in the peak timing between simulations calibrated with values of Table **??** and surveillance data.



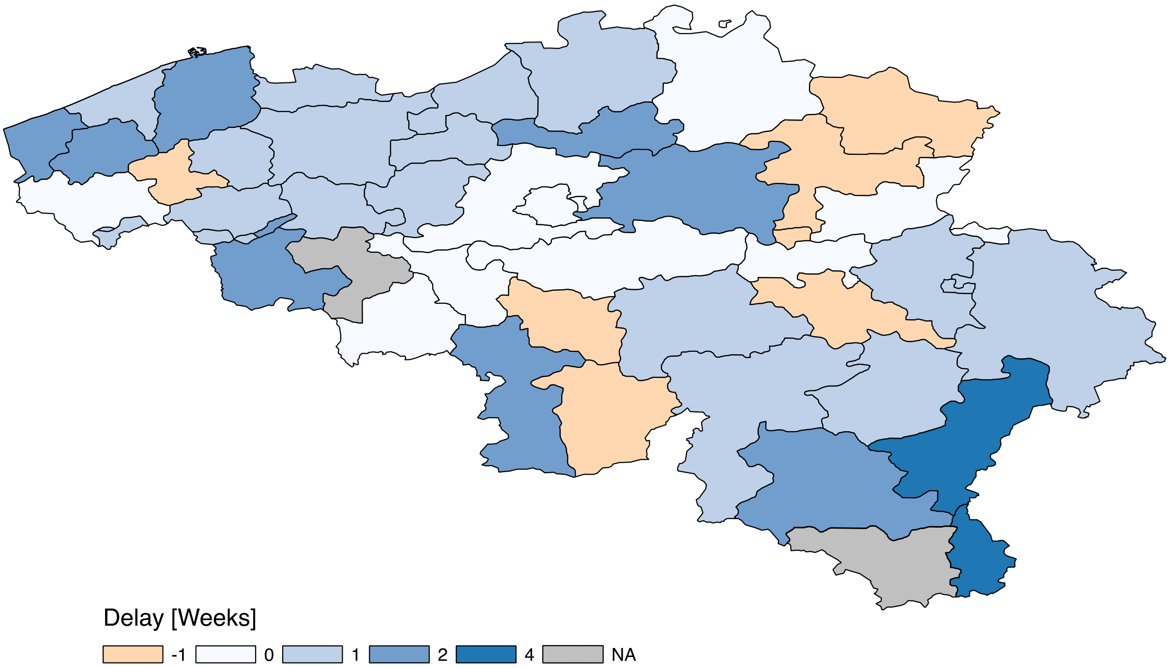


Figure 5.B.3: Top: Boxplot of the peak time difference ΔT^d per district between simulations and empirical data. Numbers represent Belgium districts, see Table 5.B.5 for corresponding names. Bottom: Geographical map of the median peak time difference per district. Figure from Supplementary Materials of De Luca *et al.* (Submitted).

Table 5.B.5 presents the list of district names and associated IDs used in the study.

Table 5.B.5: List of district names and associated IDs.

ID	District	alternative name	ID	District	alternative name
0	Antwerpen	Anvers	22	Charleroi	
1	Mechelen	Malines	23	Mons	Bergen
2	Turnhout		24	Moeskroen	Mouscron
3	Brussel	Bruxelles	25	Soignies	Zinnik
4	Halle-Vilvoorde	Hal-Vilvorde	26	Thuin	
5	Leuven	Louvain	27	Tournai	Doornik
6	Nivelles	Nijvel	28	Huy	Hoei
7	Brugge	Bruges	29	Liège	Luik
8	Diksmuide	Dixmude	30	Verviers	
9	Ieper	Ypres	31	Waremme	Borgworm
10	Kortrijk	Courtrai	32	Hasselt	
11	Oostende	Ostende	33	Maaseik	
12	Roeselare	Roulers	34	Tongeren	Tongres
13	Tielt		35	Arlon	Aarlen
14	Veurne	Furnes	36	Bastogne	Bastenaken
15	Aalst	Alost	37	Marche-en-Famenne	
16	Dendermonde	Termonde	38	Neufchâteau	
17	Eeklo		39	Virton	
18	Gent	Gand	40	Dinant	
19	Oudenaarde	Audenarde	41	Namur	Namen
20	Sint-Niklaas	Saint-Nicolas	42	Philippeville	
21	Ath	Aat			

Chapter 6

Avenues for Future Research

To gain knowledge about the effects of intervention strategies, mathematical models should be as realistic and at the same time as parsimonious as possible. Social contact behaviour relevant for infectious disease transmission can prove valuable to inform transmission parameters in mathematical models. Throughout this thesis we discussed various factors which interplay with social contact behaviour. Quantifying these effects on social contact behaviour and on the spread of infectious diseases, allows us to improve mathematical models of infectious diseases. As a result, we can obtain models better able in grasping the effects of various intervention strategies.

In Chapter 3, we studied the interplay between illness and social contact behaviour and quantified the change in social contact behaviour as a result of infection with 2009 A/H1N1pdm influenza. We observed the R_0, under an equal infectivity assumption, for symptomatic individuals to be a quarter of the value compared to when they would have been asymptomatic. This implied a change in the expected distribution of new infections whereby the concentration in incidence in children when healthy is shifted to a concentration in incidence in young adults when symptomatic. Additionally, we observed the change in the number of contacts between ill and healthy to be influenced by location, where we saw a reduction at work/school and leisure but none for the contacts at home. Two approaches were considered to estimate the proportion of infections attributed to symptomatic individuals. The first approach assumed no change in social contact behaviour during the infectious period and found the distribution of ILI incidence to be consistent with the majority of transmissions due to people with overt symptoms of ILI, e.g. two-thirds of infections are caused by symptomatic persons. In the second approach, we extended the first approach, allowing for a change in contact behaviour during the course of the infection in a non-preferential transmission model. Additionally, we extended the non-preferential transmission model to a preferential transmission model which accounts for the effect of the status (symptomatic or asymptomatic) of the infector. The data supported the last model and thus support a preferential transmission assumption stating that the development of ILI symptoms is dependent on the symptomatic or asymptomatic status of the infector. We furthermore showed the possibility to estimate the proportion of symptomatic infections or the relative infectiousness of symptomatic cases compared to asymptomatic cases in the non-preferential model. Thus, with prior information on one of these parameters, the other can be estimated from incidence data. The second approach considered a Bayesian approach and took underreporting into account; however, this was not entirely estimable from the data at hand. To be able to estimate the underreporting, data on underreporting in at least one age category is necessary. What's more, these underreporting factors can possibly account for factors other than the intention to visit the general practitioner such as the ability to better fit the data because of working with a hidden layer. Additionally, some of the parameters were kept fixed to literature values, which impacts the obtained estimates. Next, we were able to show a reduction in the total number of cases when individuals would stay home immediately after symptom onset for certain scenarios. For more subtle scenarios of home isolation, individual-based models can offer a solution and thus remains an avenue for future research. These conclusions are all based on data that are not obtained from a representative sample, but they are able to capture epidemiologically relevant behaviour and the behaviour of healthy participants was in line with the behaviour estimated from previous surveys. Furthermore, it is possible that asymptomatic cases are not feeling in the best shape and might experience different contact behaviour than healthy individuals. As the data are collected during the 2009 A/H1N1pdm influenza epidemic, it is unclear how the results apply to other influenza strains or other diseases. In light of the other influenza strains and other diseases, we conducted and described a pilot study in Chapter 3 as well. This study aimed at collecting information regarding seasonal influenza similarly as the initial survey and additionally to quantify the baseline and change in behaviour as a result of chronic infections. The pilot study used more extensive questionnaires than the first survey anno 2009-2010, as the questionnaires from the pilot study resembled the diaries from the Flemish survey anno 2010-2011. General practitioners as well as participants experienced them to be rather complex, as a result some general practitioners selected their patients based on their intelligence. This bias can lead to a non-representative sample and given the aim to reach a representative sample, it is of importance to obtain a solution when conducting a similar large-scale survey. Despite the size of this pilot study, we found a significant difference in the number of encounters between people with a chronic condition and healthy people at the 5% significance level. Furthermore, the number of contacts of people with a chronic condition and an ILI were significantly different from the number of contacts of healthy people. Besides, we observed the a borderline non-significance when hypothesising that people when ill have less contacts compared to when they are healthy. This last observation is in line with the results of the initial survey. These results thus offer a potential topic for future research, as extending the survey to a general population will contribute more insights. One of the interesting aspects is to disentangle the difference in social interactions for those with an ILI and a chronic condition compared to healthy individuals, as it is of interest to know whether the difference is driven by the ILI infection, the chronic condition or a combination of both. To extend the survey to the general population, one should think of simplifying the diaries or considering an app-based approach. Besides, as the set-up is influenced by general practitioners, it can be insightful to organise a focus-group with general practitioners and as such be able to account for their issues in the study design and the design of the questionnaires. Furthermore, a collaboration with public health authorities might result in setting up the survey across the general practitioner sentinel network, which might offer a more representative sample of participants and thus of patients/participants. Related to the diary approach, an alternative for future surveys, as touched upon previously, can be the use of an app-based survey. The recent technology of apps allows to set up surveys on multiple platforms, e.g. internet and smartphones, and to include data quality checks along data collection. Furthermore, the app might help in reducing the burden to participants. Lastly, an app-based survey can be easily extended to other social contact surveys perhaps in the context of other infectious diseases. Nevertheless, the approach has not been applied currently and it remains unclear whether this will be able to obtain comparable results as with the paper-based surveys. The gap in knowledge on this difference thus offers the potential for further investigation. We previously touched upon the need to quantify changes as a result of other infectious diseases as extending the (non-)preferential transmission model allows an estimate of the proportion asymptomatic infections, which in its turn influences the interventions under consideration. Likewise, changes in contact behaviour as a result of a chronic condition are as important, given the increased risk for those patients. Furthermore, to these patients non-pharmaceutical interventions can be more beneficial and thus it is important to be able to quantify their impact accurately.

In Chapter 4, we confirmed previous results quantifying differences in the number of contacts for week and weekend as well as for holiday periods (Hens *et al.*, 2009b) using the Flemish survey. Furthermore, we quantified an effect of age and household size and a significant effect of animal ownership, gender and provinces. Household size and age influence animal ownership, where we observed that participants below five years old were less likely to own animals compared to those older than five years, next, living alone resulted in a lower chance to own animals compared to those in larger households. The probability of touching an animal depended on age and animal ownership, including their interaction. Based on our proposed method to characterise the relative probability of a major zoonotic outbreak, we observed children below ten years and adults between 25 and 45 years of age to be more likely to cause an outbreak. We additionally presented evidence that weather might influence social mixing patterns, where the duration of contacts seemed to depend on the weather. Concrete, we observed a significant increase in the long duration contacts on regular weekdays with low temperature, low precipitation or low absolute humidity of the air. This last observation is also found in time use surveys as people tend to congregate longer inside when the weather is bad (Mccurdy and Graham, 2003; Graham and Mccurdy, 2004). In the context of the within-household networks, we estimated a larger proportion of complete networks compared to previous work based on ego-centric social networks (Potter *et al.*, 2011; Potter and Hens, 2013). On top of that, we found contacts between father and children to be less likely compared to contacts between father and mother, mother and children and between children (apart from older siblings), which is in line with work based on pertussis in households (de Greeff *et al.*, 2012). One of the ambitions was to challenge the random mixing assumption within households; however, initially we did not find any substantial differences based on a two-level mixing model, thus indicating the random mixing assumption to be reasonable. However, accounting for within-household density is of importance, based on the results. Additionally, results from a more extreme setting suggest that informing mixing between household members with social contact data could impact modelling efforts in certain settings. This chapter addressed a variety of research questions, with new results but also new methodologies. Kifle *et al.* (2016) proposed a new method to estimate the probability of a major zoonotic outbreak. Thus, it is of interest to apply the method to similar and other situations to confirm and test the results and methodology. Besides, the method imposes possibly strong assumptions such as the ability of the disease to spread from animals to humans and futhermore spread between humans, hence studying the impact of these assumptions can offer potential insights. Studying which type of contacts are of high potential will offer insights for possible non-pharmaceutical interventions. Willem *et al.* (2012) were the first to study the relationship between weather conditions and social contact behaviour. In this case, it would be of interest to perform a social contact survey which allows for a more thorough investigation; this survey should thus consider to sample a sufficient number of participants for the various types of days but also with specific weather conditions. One proposed solution is to consider many different days to collect information and hope to reach sufficient variations in weather conditions, besides increasing the number of people participating on weekends and holiday periods. Another solution can come from sampling the same individuals repeatedly or over a long period, however this might suffer from fatigue of the participants as Hens *et al.* (2009b) noticed in their survey. Yet, in an app-based context the number of recurrent questions can be limited and participants can be followed more easily over a longer period. When aiming for specific settings, such as within schools, using radio-frequency identification (RFID) can be applied and increase the accuracy compared to the contact diary approach (Smieszek *et al.*, 2014). In this chapter, we were able to quantify the within-household networks based on social contact data from households. However, we did not include the observed relation-specific heterogeneity in the duration of contacts, which might be relevant for some diseases. To incorporate this in future models, one can extend the ERGMs to be valued, where the value of the edge gives the duration of the contacts. Exploring the different type of contacts within households which are important for within-household transmission are furthermore avenues for research, possibly in line with previous research in the context of individual contact behaviour (Goeyvaerts *et al.*, 2010; Melegaro *et al.*, 2004; Ogunjimi *et al.*, 2009). The results can offer insights in targeted interventions within households. As was the case for the individual survey, the data also contains information on the time use of the participants. However, up-to-date this has not been used but can help to develop temporal within-household networks in which one takes the formation and dissolution of contact within the households into account. It would be of interest to investigate the implications of this temporal networks on the spread of infections.

Lastly, in Chapter 5 we presented two parallel surveys concerned with the interplay between time use, mobility and contact behaviour on the one hand and the interaction between mobility and contact behaviour on the other hand. In the first study, we were able to link time use data to social contact data from the Flemish survey anno 2010-2011, which resulted in distance information for the contacts. Based on this data we were able to investigate the social contact dispersal function and the change in social contact behaviour over distance and the influence on the spread of infections. We found the dispersal function to be deviating from the transmission dispersal functions used in the work from Meyer and Held (2014). However, we found subtle difference in the impact on the spread of infections from incorporating social contact matrices varying across distance. When considering constant social contact matrices across differences, one is unable to describe these subtleties. In the second study, we were able to describe the incidence patterns of seasonal influenza during the 2008/2009 season using a metapopulation model including mixing changes and mobility changes according to the school calendar of 2008/2009. The results indicate the importance of changes of mixing over the changes in mobility and that Christmas holidays have the most profound effect on the spread of influenza by mitigating the epidemic at peak time. We were able to quantify the effect of implementing one additional week of school closure to the Christmas holiday and found that only an extra week after the holiday was effective in reducing the epidemic spread. Considering early and late seasons, from 4 weeks earlier or later compared to the 2008/2009 season, the early seasons showed a more remarkable effect of the holidays. Both studies offer significant contributions to the field, but also suffer from limitations. In the first study, the linking process introduced missingness, furthermore we proposed a method to estimate the social contact dispersal for the first time. Considering smaller time blocks in the time use data or by requesting the distance for each contact we might obtain a reduction in the missingness, although these suggestions result in a larger burden on participants. Thus, an alternative might be in the app-based surveys, where the location of people might be tracked and possibly from the contacts as well. Concerning the method, it would be good to apply the method in similar contexts to test its applicability and to confirm the results. However, the method might also be extended to other areas whereby one aims to estimate an underlying continuous distribution when only categorical data is observed. As indicated, time use data and contact data were both available, in this case, we used them to infer the distance of social contacts. However, Zagheni *et al.* (2008) and De Cao *et al.* (2014) developed and applied a method to estimate social contact matrices from time use data or from the combination of independent time use data and social contact data. In the current situation, we have both for the same participants; thus, one can apply the social contact methodology as applied in this thesis as well as the methods from Zagheni *et al.* (2008) and De Cao *et al.* (2014) and compare the results. Based on these results, one can potentially investigate extensions of the time use method to obtain more alike social contact matrices. In this section, we considered a simple approach to quantify the effect of changing social contact matrices across distance. An interesting route for future research would aim to include this information into an agent-based model, or the model proposed by Meyer and Held (2017) and quantify the effects using them. In the context of the metapopulation model, we only considered two age classes, but as indicated before we were still able to reproduce the spatio-temporal unfolding of the epidemic. Next, we only aimed to replicate one season, but we accounted for potential anticipation or delays in the seasons as the effects are mainly inherent to the incidence profiles and holiday timings. Furthermore, there was a lack of data on the travel behaviour of adults during school holidays. Collaborations with research institutes or authorities in the field of mobility might help in inferring more detailed information and possibly confirm the appropriateness of the approach taken here. As this metapopulation model was able to reproduce the season fairly well, a future project could test the models' ability in other seasons. Next, one can investigate the model's ability to reproduce hospitalisations in the various hospitals around the country. In the case of good reproducibility, one can aim to predict the timings of excessive hospitalisations and thus help authorities in their decisions to the public to reduce hospitalisations or offer appropriate support to ill people.

Besides the specific limitations which offer possible avenues for future research, there are some general aspects not considered in this thesis. One such aspect is the gender of the participant, and even of the contact, which can have an influence on the number of contacts as well as on the changes described in this work. As was indicated by (Béraud, 2015), gender has an impact on infections themselves, on the consultation rates and the impact can even differ by infection, but it also has an impact on social contact behaviour. Thus, when zooming in on the interplays we have investigated here, we might suspect gender to play a role as well. It is therefore of interest to investigate the gender aspect more thoroughly. Another general consideration is related to the comparison of social contact matrices. In Chapter 3 and 4 we compared social contact behaviour by using the R_0 ratio as proposed by Hens *et al.* (2009b). Albeit its usefulness, this results in a single value to compare social contact matrices. In an initial analysis on changes as a result of illness, we explored the Euclidean distance between leading eigenvectors and the sum of absolute differences between contact matrices. However, they do not necessarily lead to epidemiologically relevant differences. On the other hand, from a theoretical point of view one would expect the ratio of incidences, i.e. leading eigenvectors, to equal one for each age class. However, comparing the obtained ratio to one for each age class results in a multiple testing problem which should be accounted for. Iozzi *et al.* (2010) consider two assortativity measures which compare the assortativity to the proportionate mixing case; their Q index has been employed by Melegaro *et al.* (2017) to compare different contact matrices. However, this also results in one single value for the two matrices. We therefore suggest to investigate the proposed measures and explore novel measures which can quantify epidemiologically relevant differences.

A major aspect of influence in this thesis is the use of the paper-based diaries to which we restricted the work. These diaries are able to capture transmission-relevant mixing sufficiently well to explain transmission of infections (Melegaro *et al.*, 2004; Ogunjimi *et al.*, 2009; Goeyvaerts *et al.*, 2010), can capture information retrospectively, are easy to administer and in a wide range of environments and settings (Read *et al.*, 2012). Nevertheless, the method suffers from several shortcomings as well. Read *et al.* (2012) reported these shortcomings to be the study fatigue when collecting longitudinal information on participants' contacts, the difficulty to quantify potential biases in participant recall and reporting, and limiting the reported encounters to a subset of all social encounters in which transmission can occur by the definition of a contact. Furthermore, Smieszek (2009) compared the contact diary approach (web-based) to the use of RFID in a closed community and observed that the two methods cannot be used interchangeable for informing models of infectious disease dynamics. The contact diary approach is however not easy-to-administer in the context of young children at school as proxies might limit themselves to long duration contacts. Hence, in those settings the RFID technology might offer a solution. Using the contact-diary approach we tackled in this thesis elements of the challenges which are present when incorporating dynamics of behaviour in infectious disease models, described by Funk *et al.* (2015). Their first challenge describes setting a baseline of behaviour in terms of contacts as well as related to disease prevention and control. Chapter 3 presents changes towards the baseline of contact behaviour, whereby additional insights are to be gained in the future. Nonetheless, the impact of behaviour towards prevention of disease and control is of importance as well and research related to this aspect might offer insights which can potentially be related to the social contact behaviour. To quantify the impact of interventions accurately, this aspect should thus be taken into account. However, as indicated by Funk *et al.* (2015) how and to what extent this should be considered in models is open for investigation. Likewise for the other challenges they present, each of them has an impact on the methods applied, e.g. if people change their health-seeking behaviour during the course of an outbreak, reaching them to perform a social contact survey might not be easy. As suggested, digital sources can offer solutions but likewise digital surveys can reach a broader audience geographically but not necessarily in age. The topics touched upon here and the other challenges depicted in Funk *et al.* (2015) are important areas for further research as they allow an important transition from theoretical models to relevant models for policy decisions.

Bibliography

Abdullah, A. S. M., Thomas, G. N., McGhee, S. M. and Morisky, D. E. (2004) Impact of severe acute respiratory syndrome (SARS) on travel and population mobility: implications for travel medicine practitioners. *Journal of Travel Medicine*, **11**, 107–111.

Addy, C. L., Longini Jr, I. M. and Haber, M. (1991) A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics*, **47**, 961–974.

Agresti, A. (2002) *Categorical Data Analysis.* Hoboken: Wiley & Sons, second edn.

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Alizon, S., Hurford, A., Mideo, N. and Van Baalen, M. (2009) Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of Evolutionary Biology*, **22**, 245–259.

Apolloni, A., Poletto, C. and Colizza, V. (2013) Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic. *BMC Infectious Diseases*, **13**, 176–193.

Apolloni, A., Poletto, C., Ramasco, J. J., Jensen, P. and Colizza, V. (2014) Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, **11**, 3–28.

Baguelin, M., Flasche, S., Camacho, A., Demiris, N., Miller, E. and Edmunds, W. J. (2013) Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Medicine*, **10**, e1001527.

Baguelin, M., Hoschler, K., Stanford, E., Waight, P., Hardelid, P., Andrews, N. and Miller, E. (2011) Age-specific incidence of A/H1N1 2009 influenza infection in England from sequential antibody prevalence data using likelihood-based estimation. *PLoS One*, **6**, e17074.

Baguelin, M., Van Hoek, A. J., Jit, M., Flasche, S., White, P. J. and Edmunds, W. J. (2010) Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. *Vaccine*, **28**, 2370–2384.

Bailey, N. T. (1957) *The mathematical theory of epidemics*. London: Charles Griffin & Co.

Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V. and Vespignani, A. (2011) Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS One*, **6**, e16591.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J. and Vespignani, A. (2009a) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, **106**, 21484–21489.

Balcan, D., Hu, H., Goncalves, B., Bajardi, P., Poletto, C., Ramasco, J. J., Paolotti, D., Perra, N., Tizzoni, M., Van den Broeck, W., Colizza, V. and Vespignani, A. (2009b) Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine*, **7**, 45–56.

Ball, F. and Britton, T. (2007) An epidemic model with infector-dependent severity. *Advances in Applied Probability*, **39**, 949–972.

Ball, F., Mollison, D. and Scalia-Tomba, G. (1997) Epidemics with two levels of mixing. *The Annals of Applied Probability*, **7**, 46–89.

Ball, F. and Neal, P. (2002) A general model for stochastic SIR epidemics with two levels of mixing. *Mathematical Biosciences*, **180**, 73–102.

Becker, N. G. (1989) *Analysis of Infectious Disease Data.* London: Chapman and Hall.

Béraud, G., Kazmercziak, S., Beutels, P., Levy-Bruhl, D., Lenne, X., Mielcarek, N., Yazdanpanah, Y., Boëlle, P.-Y., Hens, N. and Dervaux, B. (2015) The french connection: The first large population-based contact survey in france relevant for the spread of infectious diseases. *PLoS One*, **10**, e0133203.

Beutels, P., Shkedy, Z., Aerts, M. and Van Damme, P. (2006) Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection*, **134**, 1158–1166.

Beutels, P., Vandendijck, Y., Willem, L., Goeyvaerts, N., Blommaert, A., Van Kerckhove, K., Bilcke, J., Hanquet, G., Neels, P., Thiry, N., Liesenborgs, J. and Hens, N. (2013) Seasonal influenza vaccination: prioritizing children or other target groups? Part II: cost-effectiveness analysis. *KCE Report, Health Technology Assessment*, **204**.

Bilcke, J., Hens, N. and Beutels, P. (2017) Quality-of-life: a many-splendored thing? Belgian population norms and 34 potential determinants explored by beta regression. *Quality of Life Research*, **26**, 2011–2023.

Boëlle, P.-Y., Ansart, S., Cori, A. and Valleron, A.-J. (2011) Transmission parameters of the A/H1N1 (2009) influenza virus pandemic: a review. *Influenza and Other Respiratory Viruses*, **5**, 306–316.

Brankston, G., Gitterman, L., Hirji, Z., Lemieux, C. and Gardam, M. (2007) Transmission of influenza A in human beings. *The Lancet Infectious Diseases*, **7**, 257–265.

Béraud, G. (2015) *Modelling infectious agent transmission using social mixing data*. Ph.D. thesis, Université de Lille & Hasselt University.

Breslow, N. and Clayton, D. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Brownstein, J. S., Wolfe, C. J. and Mandl, K. D. (2006) Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Medicine*, **3**, e401.

Calatayud, L., Kurkela, S., Neave, P., Brock, A., Perkins, S., Zuckerman, M., Sudhanva, M., Bermingham, A., Ellis, J., Pebody, R., Catchpole, M., Heathcock, R. and Maquire, H. (2010) Pandemic (H1N1) 2009 virus outbreak in a school in London, April–May 2009: an observational study. *Epidemiology and Infection*, **138**, 183–191.

Carrat, F., Vergu, E., Ferguson, N. M., Lemaitre, M., Cauchemez, S., Leach, S. and Valleron, A.-J. (2008) Time lines of infection and disease in human influenza: a review of volunteer challenge studies. *American Journal of Epidemiology*, **167**, 775–785.

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F. and Vespignani, A. (2010) Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One*, **5**, e11596.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. and Boëlle, P. Y. (2004) A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Statistics in Medicine*, **23**, 3469–3487.

Cauchemez, S., Donnelly, C., Reed, C., Ghani, A., Fraser, C., Kent, C., FInelli, L. and Ferguson, N. (2009a) Household transmission of 2009 pandemic influenza (H1N1) virus in the United States. *The New England Journal of Medicine*, **361**, 2619–2627.

Cauchemez, S., Ferguson, N. M., Wachtel, C., Tegnell, A., Saour, G., Duncan, B. and Nicoll, A. (2009b) Closure of schools during an influenza pandemic. *The Lancet Infectious Diseases*, **9**, 473–481.

Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. and Ferguson, N. M. (2008) Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, **452**, 750–754.

Chao, D., Halloran, M. and Longini, I. (2010a) School opening dates predict pandemic influenza A(H1N1) outbreaks in the United States. *The Journal of Infectious Diseases*, **202**, 877–880.

Chao, D. L., Halloran, M. E., Obenchain, V. J. and Longini Jr, I. M. (2010b) FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Computational Biology*, **6**, e1000656.

Charaudeau, S., Pakdaman, K. and Boëlle, P.-Y. (2014) Commuter mobility and the spread of infectious diseases: application to influenza in France. *PLoS One*, **9**, e83002.

Chen, S.-C. and You, Z.-S. (2015) Social contact patterns of school-age children in Taiwan: comparison of the term time and holiday periods. *Epidemiology and Infection*, **143**, 1139–1147.

Ciavarella, C., Fumanelli, L., Merler, S., Cattuto, C. and Ajelli, M. (2016) School closure policies at municipality level for mitigating influenza spread: a model-based evaluation. *BMC Infectious Diseases*, **16**, 576–586.

Cleaveland, S., Laurenson, M. and Taylor, L. (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 991–999.

Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J. and Vespignani, A. (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Medicine*, **4**, e13.

Cooper, B. S., Pitman, R. J., Edmunds, W. J. and Gay, N. J. (2006) Delaying the international spread of pandemic influenza. *PLoS Medicine*, **3**, e212.

Cowling, B. J., Lau, E. H., Lam, C. L., Cheng, C. K., Kovar, J., Chan, K. H., Peiris, J. M. and Leung, G. M. (2008) Effects of school closures, 2008 winter influenza season, Hong Kong. *Emerging Infectious Diseases*, **14**, 1660–1662.

Crépey, P. and Barthélemy, M. (2007) Detecting robust patterns in the spread of epidemics: a case study of influenza in the United States and France. *American Journal of Epidemiology*, **166**, 1244–1251.

Danis, K., Fitzgerald, M., Connell, J., Conlon, M. and Murphy, P. (2004) Lessons from a pre-season influenza outbreak in a day school. *Communicable Disease and Public Health*, **7**, 179–183.

Danon, L., Ford, A. P., House, T., Jewell, C. P., Keeling, M. J., Roberts, G. O., Ross, J. V. and Vernon, M. C. (2011) Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, **2011**, Article ID 284909.

Danon, L., House, T. A., Read, J. M. and Keeling, M. J. (2012) Social encounter networks: collective properties and disease transmission. *Journal of the Royal Society Interface*, rsif20120357.

Danon, L., Read, J. M., House, T. A., Vernon, M. C. and Keeling, M. J. (2013) Social encounter networks: characterizing great britain. *Proceedings of the Royal Society B (Biological Sciences)*, **280**, 20131037.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press.

De Cao, E., Zagheni, E., Manfredi, P. and Melegaro, A. (2014) The relative importance of frequency of contacts and duration of exposure for the spread of directly transmitted infections. *Biostatistics*, **15**, 470–483.

De Luca, G., Van Kerckhove, K., Coletti, P., Poletto, C., Bossuyt, N., Hens, N. and Colizza, V. (Submitted) The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium.

Del Valle, S., Hyman, J., Hethcote, H. and Eubank, S. (2007) Mixing patterns between age groups in social networks. *Social Networks*, **29**, 539–554.

Diekmann, O., Heesterbeek, H. and Britton, T. (2013) *Mathematical Tools for Understanding Infectious Disease Dynamics*. New Jersey: Princeton University Press.

Diekmann, O., Heesterbeek, J. and Metz, J. (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, **28**, 65–382.

Dodd, P. J., Looker, C., Plumb, I. D., Bond, V., Schaap, A., Shanaube, K., Muyoyeta, M., Vynnycky, E., Godfrey-Faussett, P., Corbett, E. L., Beyers, N., Ayles, H. and White, R. G. (2016) Age- and sex-specific social contact patterns and incidence of mycobacterium tuberculosis infection. *American Journal of Epidemiology*, **183**, 156–166.

Dorjee, S., Poljak, Z., Revie, C. W., Bridgland, J., McNab, B., Leger, E. and Sanchez, J. (2013) A review of simulation modelling approaches used for the spread of zoonotic influenza viruses in animal and human populations. *Zoonoses Public Health*, **60**, 383–411.

Eames, K., Tilston, N., White, P., Adams, E. and Edmunds, W. (2010) The impact of illness and the impact of school closure on social contact patterns. *Health Technology Assessment*, **4**, 249–294.

Eames, K. T., Tilston, N. L., Brooks-Pollock, E. and Edmunds, W. J. (2012) Measured dynamic social contact patterns explain the spread of H1N1v influenza. *PLoS Computational Biology*, **8**, e1002425.

Eames, K. T., Tilston, N. L. and Edmunds, W. J. (2011) The impact of school holidays on the social mixing patterns of school children. *Epidemics*, **3**, 103–108.

Edmunds, W., Kafatos, G., Wallinga, J. and Mossong, J. (2006) Mixing patterns and the spread of close-contact infectious diseases. *Emerging Themes in Epidemiology*, **3**.

Edmunds, W., O'Callaghan, C. and Nokes, D. (1997) Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 949–957.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Epstein, J. M., Goedecke, D. M., Yu, F., Morris, R. J., Wagener, D. K. and Bobashev, G. V. (2007) Controlling pandemic flu: the value of international air travel restrictions. *PLoS One*, **2**, e401.

European Centre for Disease Prevention and Control (2017a) Factsheet about seasonal influenza, `https://ecdc.europa.eu/en/seasonal-influenza/facts/factsheet`. Accessed 23 August 2017.

European Centre for Disease Prevention and Control (2017b) Factsheet Varicella, `http://ecdc.europa.eu/en/healthtopics/varicella_infection/Pages/factsheet_general_public.aspx`. Accessed 16 May 2017.

Ewing, A., Lee, E. C., Viboud, C. and Bansal, S. (2017) Contact, travel, and transmission: The impact of winter holidays on influenza dynamics in the United States. *The Journal of Infectious Diseases*, **215**, 732–739.

Farrington, C., Kanaan, M. and Gay, N. (2001) Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *Applied Statistics*, **50**, 251–292.

Ferguson, N. M., Cummings, D. A. T., Fraser, C., Cajka, J. C., Cooley, P. C. and Burke, D. S. (2006) Strategies for mitigating an influenza pandemic. *Nature*, **442**, 448–452.

Flahault, A., Letrait, S., Blin, P., Hazout, S., Menares, J. and Valleron, A.-J. (1988) Modelling the 1985 influenza epidemic in France. *Statistics in Medicine*, **7**, 1147–1155.

Flasche, S., Hens, N., Boëlle, P., Mossong, J., van Ballegooijen, W., Nunes, B., Rizzo, C., Popovici, F., Santa-Olalla, P., Hrubá, F., Parmakova, K., Baguelin, M., van Hoek, A., Desenclos, J., Bernillon, P., Cámara, A., Wallinga, J., Asikainen, T., White, P. and Edmunds, W. (2011) Different transmission patterns in the early stages of the influenza A(H1N1)v pandemic: A comparative analysis of 12 European countries. *Epidemics*, **3**, 125–133.

Fraser, C., Donnelly, C., Cauchemez, S., Hanage, W., Van Kerhove, M., Hollingsworth, T., Griffin, J., Baggaley, R., Jenkins, H., Lyons, E., Jombart, T., Hinsley, W., Grassly, N., Balloux, F., Ghani, A. and Ferguson, N. (2009) Pandemic potential of a strain of influenza A(H1N1): Early findings. *Science*, **324**, 1557–1561.

Fu, Y.-c., Wang, D.-W. and Chuang, J.-H. (2012) Representative contact diaries for modeling the spread of infectious diseases in taiwan. *PLoS One*, **7**, e45113.

Fuhrmann, C. (2010) The effects of weather and climate on the seasonality of influenza: What we know and what we need to know. *Geography Compass*, **4**, 718–730.

Fujii, H., Takahashi, H., Ohyama, T., Hattori, K. and Suzuki, S. (2002) Evaluation of the school health surveillance system for influenza, Tokyo, 1999-2000. *Japanese Journal of Infectious Diseases*, **55**, 97–99.

Fumanelli, L., Ajelli, M., Merler, S., Ferguson, N. M. and Cauchemez, S. (2016) Model-based comprehensive analysis of school closure policies for mitigating influenza epidemics and pandemics. *PLoS Computational Biology*, **12**, e1004681.

Funk, S., Bansal, S., Bauch, C. T., Eames, K. T., Edmunds, W. J., Galvani, A. P. and Klepac, P. (2015) Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*, **10**, 21–25.

Funk, S., Salathé, M. and Jansen, V. A. (2010) Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, rsif.2010.0142.

Geilhufe, M., Held, L., Skrøvseth, S. O., Simonsen, G. S. and Godtliebsen, F. (2014) Power law approximations of movement network data for modeling infectious disease spread. *Biometrical Journal*, **56**, 363–382.

Geyer, C. J. and Thompson, E. A. (1992) Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**, 657–699.

Glorieux, I. and Vandeweyer, J. (2011) Gezin en school. De kloof voorbij, de grens gezet?, `http://www.vlor.be/publicatie/gezin-en-school-de-kloof-voorbij-de-grens-gezet`. Accessed 24 May 2017.

Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P. and Beutels, P. (2010) Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society Series C*, **59**, 255–277.

Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., Van Kerckhove, K., Willem, L., Aerts, M., Beutels, P. and Hens, N. (In preparation) Household members do not contact each other at random: Implications for infectious disease modelling.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T. and Morris, M. (2008) A statnet tutorial. *Journal of Statistical Software*, **24**, 1–27.

Graham, S. and Mccurdy, T. (2004) Developing meaningful cohorts for human exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, **14**, 23–43.

Grais, R. F., Hugh Ellis, J. and Glass, G. E. (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, **18**, 1065–1072.

de Greeff, S. C., de Melker, H. E., Westerhof, A., Schellekens, J. F., Mooi, F. R. and van Boven, M. (2012) Estimation of household transmission rates of pertussis and the effect of cocooning vaccination strategies on infant pertussis. *Epidemiology*, **23**, 852–860.

Greene, C. and Levy, J. (2006) Immunocompromised people and shared human and animal infections: zoonoses, sapronoses and anthroponoses. *Infectious Diseases of the Dog and Cat*, **3**, 1051–1068.

Grefenstette, J. J., Brown, S. T., Rosenfeld, R., DePasse, J., Stone, N. T., Cooley, P. C., Wheaton, W. D., Fyshe, A., Galloway, D. D., Sriram, A., Guclu, H., Abraham, T. and Burke, D. S. (2013) FRED (A Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health*, **13**, 940–953.

Grijalva, C. G., Goeyvaerts, N., Verastegui, H., Edwards, K. M., Gil, A. I., Lanata, C. F., Hens, N. and for the RESPIRA PERU project (2015) A household-based study of contact networks relevant for the spread of infectious diseases in the highlands of Peru. *PLoS One*, **10**, e0118457.

Halloran, M. E., Longini, I. M., Nizam, A. and Yang, Y. (2002) Containing bioterrorist smallpox. *Science*, **298**, 1428–1432.

Handcock, M., Hunter, D., Butts, C., Goodreau, S. and Morris, M. (2003) ERGM: Fit, simulate and diagnose exponential-family models for networks, Version 2.1. *url: http://statnetproject. org*.

Handcock, M., Hunter, D. R., Butts, C. T., Goodreau, S., Krivitsky, P., Bender-deMoll, S. and Morris, M. (2014) statnet: Software tools for the statistical analysis of network data. *The Statnet Project (http://www. statnet. org). R package version*.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M. and Morris, M. (2008) statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, **24**, 1548Ű7660.

Hanneke, S., Fu, W. and Xing, E. P. (2010) Discrete temporal models of social networks. *Electronic Journal of Statistics*, **4**, 585–605.

Held, L., Hofmann, M., Höhle, M. and Schmid, V. (2006) A two-component model for counts of infectious diseases. *Biostatistics*, **7**, 422–437.

Held, L., Höhle, M. and Hofmann, M. (2005) A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, **5**, 187–199.

Held, L. and Paul, M. (2012) Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, **54**, 824–843.

Hens, N., Ayele, G. M., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J. W. and Beutels, P. (2009a) Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infectious Diseases*, **9**, 187–199.

Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Damme, P. V. and Beutels, P. (2009b) Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, **9**, 5–22.

Hens, N., Shkedy, Z., Aerts, M., Faes, C., Van Damme, P. and Beutels, P. (2012) *Modeling Infectious Disease Parameters Based on Serological and Social Contact Data. A Modern Statistical Perspective.* New York: Springer.

Hens, N., Wienke, A., Aerts, M. and Molenberghs, G. (2009c) The correlated and shared gamma frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, **28**, 2785–2800.

Heymann, A., Chodick, G., Reichman, B., Kokia, E. and Laufer, J. (2004) Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *The Pediatric Infectious Disease Journal*, **23**, 675–677.

Hoang, V. T. *et al.* (In preparation) A systematic review on social contact surveys relevant for transmission of respiratory diseases.

van Hoek, A. J., Andrews, N., Campbell, H., Amirthalingam, G., Edmunds, W. J. and Miller, E. (2013) The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young. *PLoS One*, **8**, e76180.

House, T., Baguelin, M., Van Hoek, A. J., White, P. J., Sadique, Z., Eames, K., Read, J. M., Hens, N., Melegaro, A., Edmunds, W. J. and Keeling, M. J. (2011) Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 2753–2760.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M. and Morris, M. (2008) ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, **24**, nihpa54860.

INS [producer] and Directorate General Statistics and Economic Information (DGSEI) [distributor] (2006) Enquête socio-économique générale 2001 (ESEG2001) [electronic files] . Available upon request at distributor. URL `http://statbel.fgov.be/fr/statistiques/collecte\_donnees/recensement/2001/`. Brussels.

INSEE [producer] and Centre Maurice Halbwachs (CMH) [distributor] (1999) Récensement de la population 1999 : tableaux mobilités [electronic files]. Available upon request at distributor. URL `http://www.reseau-quetelet.cnrs.fr/`. Paris.

Iozzi, F., Trusiano, F., Chinazzi, M., Billari, F. C., Zagheni, E., Merler, S., Ajelli, M., Del Fava, E. and Manfredi, P. (2010) Little Italy: an agent-based approach to the estimation of contact patterns- fitting predicted matrices to serological data. *PLoS Computational Biology*, **6**, e1001021.

Jackson, C., Mangtani, P., Vynnycky, E., Fielding, K., Kitching, A., Mohamed, H., Roche, A. and Maguire, H. (2011) School closures and student contact patterns. *Emerging Infectious Diseases*, **17**, 245Ű247.

Jackson, C., Vynnycky, E., Hawker, J., Olowokure, B. and Mangtani, P. (2013) School closures and influenza: systematic review of epidemiological studies. *BMJ Open*, **3**, e002149.

Johnson, A. J., Moore, Z. S., Edelson, P. J., Kinnane, L., Davies, M., Shay, D. K., Balish, A., McCarron, M., Blanton, L., Finelli, L., Averhoff, F., Bresee, J., Engel, J. and Fiore, A. (2008) Household responses to school closure resulting from outbreak of influenza B, North Carolina. *Emerging Infectious Diseases*, **14**, 1024–1030.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L. and Daszak, P. (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Keeling, M. J. and Eames, K. T. D. (2005) Networks and epidemic models. *Journal of the Royal Society Interface*, **2**, 295–307.

Keeling, M. J. and Rohani, P. (2007) *Modeling Infectious Diseases in Humans and Animals*. New Jersey: Princeton University Press.

Kifle, Y. W., Goeyvaerts, N., Van Kerckhove, K., Willem, L., Kucharski, A., Faes, C., Leirs, H., Hens, N. and Beutels, P. (2016) Correction: Animal ownership and touching enrich the context of social contacts relevant to the spread of human infectious diseases. *PLoS One*, **11**.

Kifle, Y. W., Vercruysse, S., Beutels, P. and Hens, N. (To be submitted) Macroscopic changes in social contact patterns over time relevant to inform the transmission process of respiratory infections.

Kolaczyk, E. (2009) *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.

Krivitsky, P. N. (2012) Exponential-family random graph models for valued networks. *Electronic Journal of Statistics*, **6**, 1100–1128.

Krivitsky, P. N. and Handcock, M. S. (2014) A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 29–46.

Kucharski, A. J., Conlan, A. J. and Eames, K. T. (2015) School's out: seasonal variation in the movement patterns of school children. *PLoS 0ne*, **10**, e0128070.

Lawson, A. B. (2009) *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Boca Raton: CRC press.

Lin, C.-J., Deger, K. A. and Tien, J. H. (2016) Modeling the trade-off between transmissibility and contact in infectious disease dynamics. *Mathematical Biosciences*, **277**, 15–24.

Lofgren, E., Fefferman, N. H., Naumov, Y. N., Gorski, J. and Naumova, E. N. (2007) Influenza seasonality: Underlying causes and modeling theories. *Journal of Virology*, **81**, 5429–5436.

Lomax, K. (1954) Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, **49**, 847–852.

Longini, I. M., Koopman, J. S., Monto, A. S. and Fox, J. P. (1982) Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology*, **115**, 736–751.

Longini Jr, I. M. and Koopman, J. S. (1982) Household and community transmission parameters from final distributions of infections in households. *Biometrics*, **38**, 115–126.

Lowen, A., Mubareka, S., Steel, J. and Palese, P. (2007) Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens*, **3**, e151.

Luh, D.-L., You, Z.-S. and Chen, S.-C. (2016) Comparison of the social contact patterns among school-age children in specific seasons, locations, and times. *Epidemics*, **14**, 36–44.

Mccurdy, T. and Graham, S. E. (2003) Using human activity data in exposure models: analysis of discriminating factors. *Journal of Exposure Analysis and Environmental Epidemiology*, **13**, 294–317.

Melegaro, A., Del Fava, E., Poletti, P., Merler, S., Nyamukapa, C., Williams, J., Gregson, S. and Manfredi, P. (2017) Social contact structures and time use patterns in the manicaland province of Zimbabwe. *PLoS One*, **12**, e0170459.

Melegaro, A., Gay, N. and Medley, G. (2004) Estimating the transmission parameters of pneumococcal carriage in households. *Epidemiology and Infection*, **132**, 433–441.

Melegaro, A., Jit, M., Gay, N., Zagheni, E. and Edmunds, J. (2011) What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. *Epidemics*, **3**, 143–151.

Merler, S., Ajelli, M., Pugliese, A. and Ferguson, N. M. (2011) Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. *PLoS Computational Biology*, **7**, e1002205.

Meyer, S. and Held, L. (2014) Power-law models for infectious disease spread. *The Annals of Applied Statistics*, **8**, 1612–1639.

Meyer, S. and Held, L. (2017) Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, **18**, 338–351.

Meyers, L. A., Pourbohloul, B., Newman, M. E., Skowronski, D. M. and Brunham, R. C. (2005) Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology*, **232**, 71–81.

Mikolajczyk, R. T., Akmatov, M. K., Rastin, S. and Kretzschmar, M. (2008) Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology and Infection*, **136**, 813–822.

Miller, E., Hoschler, K., Hardelid, P., Stanford, E., Andrews, N. and Zambon, M. (2010) Incidence of 2009 pandemic influenza A H1N1 infection in England: a cross-sectional serological study. *The Lancet*, **375**, 1100–1108.

Mniszewski, S. M., Del Valle, S. Y., Stroud, P. D., Riese, J. M. and Sydoriak, S. J. (2008) EpiSimS simulation of a multi-component strategy for pandemic influenza. In: *Proceedings of the 2008 Spring simulation multiconference*, 556–563.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.

Morris, M., Handcock, M. S. and Hunter, D. R. (2008) Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software*, **24**, 1548–7660.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M. and Edmunds, J. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, **5**, 381–391.

van Noort, S. P., Águas, R., Ballesteros, S. and Gomes, M. G. M. (2012) The role of weather on the relation between influenza and influenza-like illness. *Journal of Theoretical Biology*, **298**, 131–137.

Ogunjimi, B., Hens, N., Goeyvaerts, N., Aerts, M., Damme, P. V. and Beutels, P. (2009) Using empirical social contact data to model person to person infectious disease transmission: An illustration for varicella. *Mathematical Biosciences*, **218**, 80–87.

Paul, M. and Held, L. (2011) Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Statistics in Medicine*, **30**, 1118–1136.

Paul, M., Held, L. and Toschke, A. M. (2008) Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, **27**, 6250Ű6267.

Perra, N., Balcan, D., Gonçalves, B. and Vespignani, A. (2011) Towards a characterization of behavior-disease models. *PLoS One*, **6**, e23084.

Poletto, C., Gomes, M. F., Pastore y Piontti, A., Rossi, L., Bioglio, L., Chao, D. L., Longini, I. M., Halloran, M. E., Colizza, V. and Vespignani, A. (2014) Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Euro Surveillance*, **19**, pii=20936.

Potter, G. E., Handcock, M., Longini, I. M. and Halloran, E. (2011) Estimating within-household contact networks from egocentric data. *Annals of Applied Statistics*, **5**, 1816–1838.

Potter, G. E. and Hens, N. (2013) A penalized likelihood approach to estimate within-household contact networks from egocentric data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **62**, 629–648.

Potter, G. E., Smieszek, T. and Sailer, K. (2015) Modeling workplace contact networks: The effects of organizational structure, architecture, and reporting errors on epidemic predictions. *Network Science*, **3**, 298–325.

Public Health England (2010) Weekly Epidemiological Updates Archive, `http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/PandemicInfluenza/H1N1PandemicArchive/SIEpidemiologicalData/SIEpidemiologicalReportsArchive/influswarchiveweeklyepireports/`. Accessed 20 December 2010.

Read, J. M., Edmunds, W. J., Riley, S., Lessler, J. and Cummings, D. A. T. (2012) Close encounters of the infectious kind: methods to measure social mixing behaviour. *Epidemiology and Infection*, **140**, 2117–2130.

Read, J. M., Lessler, J., Riley, S., Wang, S., Tan, L. J., Kwok, K. O., Guan, Y., Jiang, C. Q. and Cummings, D. A. (2014) Social mixing patterns in rural and urban areas of southern China. *Proceedings of the Royal Society of London B: Biological Sciences*, **281**, 20140268.

Reed, W. J. (2003) The pareto law of incomes - an explanation and an extension. *Physica A: Statistical Mechanics and its Applications*, **319**, 469–486.

Reumers, S., Polders, E., Janssens, D., Declercq, K. and Wets, G. (2016) Onderzoek verplaatsingsgedrag Vlaanderen 5.1 (2015-2016). *Tech. rep.*, Instituut voor Mobiliteit.

Rizzo, C., Fabiani, M., Amlôt, R., Hall, I., Finnie, T., Rubin, G. J., Cucuiu, R., Pistol, A., Popovici, F., Popescu, R., Vainola, J., Auranen, K., Leach, S., Declich, S. and Pugliese, A. (2013) Survey on the likely behavioural changes of the general public in four European countries during the 2009/2010 pandemic. In: *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*, 23–41. New York: Springer.

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007) An introduction to exponential random graph (p*) models for social networks. *Social Networks*, **29**, 173–191.

Rodriguez, C. V., Rietberg, K., Baer, A., Kwan-Gett, T. and Duchin, J. (2009) Association between school closure and subsequent absenteeism during a seasonal influenza epidemic. *Epidemiology*, **20**, 787–792.

Rohani, P., Zhong, X. and King, A. A. (2010) Contact network structure explains the changing epidemiology of pertussis. *Science*, **330**, 982–985.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W. and Jones, J. H. (2010) A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 22020–22025.

Santermans, E., Van Kerckhove, K., Azmon, A., Edmunds, W. J., Beutels, P., Faes, C. and Hens, N. (2017) Structural differences in mixing behavior informing the role of asymptomatic infection and testing symptom heritability. *Mathematical Biosciences*, **285**, 43–54.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.

Shaman, J. and Kohn, M. (2009) Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 3243–3248.

Shaman, J., Pitzer, V., Viboud, C., Grenfell, B. and Lipsitch, M. (2010) Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biology*, **2**, e1000316.

Singh, S., Schneider, D. J. and Myers, C. R. (2013) The structure of infectious disease outbreaks across the animal-human interface. *arXiv preprint arXiv:1307.4628*.

Smieszek, T. (2009) A mechanistic model of infection: why duration and intensity of contacts should be included in models of disease spread. *Theoretical Biology and Medical Modelling*, **6**, 25–34.

Smieszek, T., Barclay, V. C., Seeni, I., Rainey, J. J., Gao, H., Uzicanin, A. and Salathé, M. (2014) How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infectious Diseases*, **14**, 136–148.

SOeS [producer] and Centre Maurice Halbwachs (CMH) [distributor] (2008) Transports et déplacements (ENTD) – 2008 [electronic file]. Available upon request at distributor. URL `http://www.reseau-quetelet.cnrs.fr/`. Paris.

Steel, J., Palese, P. and Lowen, A. (2011) Transmission of a 2009 pandemic influenza virus shows a sensitivity to temperature and humidity similar to that of an H3N2 seasonal strain. *Journal of Virology*, **85**, 1400–1402.

Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., Van den Broeck, W., Régis, C., Lina, B. and Vanhems, P. (2011) High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One*, **6**, e23176.

Stein, M. L., van der Heijden, P. G., Buskens, V., van Steenbergen, J. E., Bengtsson, L., Koppeschaar, C. E., Thorson, A. and Kretzschmar, M. E. (2015) Tracking social contact networks with online respondent-driven detection: who recruits whom? *BMC Infectious Diseases*, **15**, 522–533.

Strauss, D. and Ikeda, M. (1990) Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, **85**, 204–212.

Tafforeau, J. (2008) Enquête de santé par téléphone 2008. La vaccination. *Tech. rep.*, Wetenschappelijk Instituut voor Volksgezondheid (WIV-ISP), `https://his.wiv-isp.be/fr/Documentspartages/VA_FR_2008.pdf`.

Tamerius, J., Nelson, M., Zhou, S., Viboud, C., Miller, M. and Alonso, W. (2011) Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental Health Perspectives*, **119**, 439–445.

Taylor, L. H., Latham, S. M. and Mark, E. (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **356**, 983–989.

Van Casteren, V., Mertens, K., Antoine, J., Wanyama, S., Thomas, I. and Bossuyt, N. (2010) Clinical surveillance of the influenza A (H1N1) 2009 pandemic through the network of sentinel general practitioners. *Archives of Public Health*, **68**, 62–67.

Van Kerckhove, K., Hens, N., Edmunds, W. J. and Eames, K. T. D. (2013) The impact of illness on social networks: implications for transmission and control of influenza. *American Journal of Epidemiology*, **178**, 1655–1662.

Verelst, F., Willem, L. and Beutels, P. (2016) Behavioural change models for infectious disease transmission: a systematic review (2010–2015). *Journal of The Royal Society Interface*, **13**, rsif.2016.0820.

Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A. and Grenfell, B. T. (2006) Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, **312**, 447–451.

Viboud, C., Boëlle, P.-Y., Cauchemez, S., Lavenu, A., Valleron, A.-J., Flahault, A. and Carrat, F. (2004) Risk factors of influenza transmission in households. *British Journal of General Practice*, **54**, 684–689.

Voith, V. L. (2009) The impact of companion animal problems on society and the role of veterinarians. *Veterinary Clinics of North America: Small Animal Practice*, **39**, 327–345.

Vynnycky, E. and White, R. (2010) *An Introduction to Infectious Disease Modelling*. New York: Oxford University Press.

Wallace, J. and Hobbs, P. (2006) *Atmospheric Science: An Introductory Survey*. New York: Academic press, second edn.

Wallinga, J., Teunis, P. and Kretzschmar, M. (2006) Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American Journal of Epidemiology*, **164**, 936–944.

Wasserman, S. and Faust, K. (1994) *Social Network Analysis: Methods and Applications*. New York: Cambridge university press.

Wetenschappelijk Instituut voor Volksgezondheid (2017) Influenza: High-risk groups, `https://www.wiv-isp.be/en/topics/influenza/high-risk-groups#who-is-influenza-dangerous-for-`. Accessed 15 June 2017.

Willem, L., Van Kerckhove, K., Chao, D. L., Hens, N. and Beutels, P. (2012) A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. *PLoS One*, **7**, e48695.

Woolhouse, M. E. and Gowtage-Sequeria, S. (2005) Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, **11**, 1842Ű1847.

World Health Organization (2009) Reducing transmission of pandemic (H1N1) 2009 in school settings. A framework for national and local planning and response., `http://www.who.int/csr/resources/publications/reducing_transmission_h1n1_2009.pdf`. Accessed 24 May 2017.

World Health Organization (2017a) Influenza (Seasonal), `http://www.who.int/mediacentre/factsheets/fs211/en/`. Accessed 15 June 2017.

World Health Organization (2017b) Measles, `http://www.who.int/mediacentre/factsheets/fs286/en/`. Accessed 16 May 2017.

World Health Organization Writing Group, Bell, D., Nicoll, A., Fukuda, K., Horby, P., Monto, A., Hayden, F., Wylks, C., Sanders, L. and van Tam, J. (2006) Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerging Infectious Diseases*, **12**, 88–94.

Wu, J. T., Cowling, B. J., Lau, E. H., Ip, D. K., Ho, L.-M., Tsang, T., Chuang, S.-K., Leung, P.-Y., Lo, S.-V. and Liu, Shao-Haeiand Riley, S. (2010) School closure and mitigation of pandemic (H1N1) 2009, Hong Kong. *Emergerging Infectious Diseases*, **16**, 538–541.

Yang, W., Lipsitch, M. and Shaman, J. (2015) Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, **112**, 2723–2728.

Zagheni, E., Billari, F., Manfredi, P., Melegaro, A., Mossong, J. and Edmunds, W. (2008) Using time use data to parameterize models for the spread of close-contact infectious diseases. *American Journal of Epidemiology*, **168**, 1082–1090.

Zhang, T., Fu, X., Kwoh, C. K., Xiao, G., Wong, L., Ma, S., Soh, H., Lee, G. K. K., Hung, T. and Lees, M. (2011) Temporal factors in school closure policy for mitigating the spread of influenza. *Journal of Public Health Policy*, **32**, 180–197.

Zhang, T., Fu, X., Ma, S., Xiao, G., Wong, L., Kwoh, C. K., Lees, M., Lee, G. K. K. and Hung, T. (2012) Evaluating temporal factors in combined interventions of workforce shift and school closure for mitigating the spread of influenza. *PLoS One*, **7**, e32203.

Samenvatting

Om het aantal gevallen van infectieziekten, zoals influenza (griep), mazelen en varicella, te verminderen, worden interventiemaatregelen ingevoerd, waaronder vaccinatie of het isoleren van de besmette individuen. Alvorens deze maatregelen te implementeren, is het belangrijk om de vermindering in kosten en de lasten voor de gemeenschap te kwantificeren. Het is belangrijk om hierbij gebruik te maken van realistische, doch eenvoudige modellen. Dit kunnen onder andere wiskundige modellen zijn die trachten het verspreidingsproces van de ziekte te visualiseren en te berekenen aan de hand van vergelijkingen. Sociaal contact gedrag dat relevant is voor infectieziekten, blijkt waardevol te kunnen zijn om het verspreidingsproces te voorspellen. Informatie over dit gedrag wordt verzameld door middel van sociale contact studies. Die studies staan toe om te beschrijven wie met wie contact heeft en dus mogelijk de ander besmet. Momenteel gebruiken slechts enkele wiskundige modellen deze sociale contact gegevens en nog minder laten toe dat dit gedrag verandert. Bijgevolg zijn deze modellen niet volledig realistisch. In deze thesis kwantificeren we veranderingen in sociaal contact gedrag ten gevolge van ziekte, tijdsbesteding en mobiliteit. Onze resultaten kunnen het realisme van de modellen verbeteren, waardoor deze de effecten van interventies beter kunnen voorspellen en aldus de lasten en de kosten van de ziekten verminderen. In wat volgt, weiden we uit over de link tussen verspreiding en sociale contact studies, startend met drie voorbeelden van infectieziekten die via de lucht worden overgedragen.

Seizoensgebonden influenza, ook wel seizoensgriep genoemd, is een acute ademhalingsinfectie die veroorzaakt wordt door influenza virussen. Volgens het Europese Centrum voor Ziekteverspreiding en Controle veroorzaakt deze ziekte jaarlijks 4-50 miljoen symptomatische gevallen in de Europese Unie met 15 000 tot 70 000 geassocieerde sterften. Ouderen, jonge kinderen en mensen met chronische aandoeningen lopen het grootste risico om ernstige complicaties te ontwikkelen die de dood tot gevolg kunnen hebben. Mazelen is een andere, zeer besmettelijke, ziekte die meestal bij kinderen voorkomt. Volgens de Wereldgezondheidsorganisatie is dit één van de hoofdoorzaken van sterften bij kleine kinderen, niettegenstaande er een veilig en kosten-effectieve vaccinatie voorhanden is. Een andere kinderziekte, varicella, beter bekend als waterpokken, is een milde ziekte die toch complicaties kan veroorzaken bij zuigelingen, volwassenen, zwangere vrouwen en immuundeficiënte personen. De ziekte komt bij volwassenen voor in de vorm van gordelroos. Verspreiding kan beperkt worden door contact met besmette personen te mijden.

Uit deze drie voorbeelden zien we mogelijke lasten en kosten ten gevolge van infecties en eventuele complicaties. Verder halen we hieruit maatregelen om infectie te verminderen of te vermijden zoals vaccinatie en isolatie.

Wanneer men echter interventies wenst te implementeren in nieuwe doelgroepen of nieuwe interventies wenst te bestuderen, zou het onethisch zijn om deze zonder meer in te voeren; wiskundige modellen bieden hiervoor een oplossing. De modellen hebben als doel om de complexe realiteit op een gestructureerde wijze voor te stellen en hierin kunnen verschillende scenario's beschouwd worden. De modellen geven als het ware het verspreidingsproces van een ziekte weer. Dit verspreidingsproces kan op verschillende manieren optreden, zoals via direct of indirect contact, via druppels; dat wil zeggen door op korte afstand te praten, door hoesten of niezen; of via de lucht. De verspreiding van de drie beschreven ziekten kan ondergebracht worden bij direct contact of via druppels. De transmissies worden weergegeven in een matrix die informatie bevat over wie de infectie van wie kan krijgen, de zogenaamde WAIFW-matrix, voluit de Who-Acquires-Infection-From-Whom-matrix. Twee decennia geleden vond de eerste sociale contact studie plaats waarbij deelnemers hun contacten noteerden gedurende een willekeurig aangewezen dag. Van elk contact werd bijkomende informatie verzameld, zoals de leeftijd, het geslacht en de duur van de ontmoeting. De onderzoekers suggereerden dat deze gegevens gebruikt kunnen worden om realistische WAIFW matrices te verkrijgen. Sindsdien vonden verschillende sociale contact studies plaats en toonde men voor bepaalde ziektes, o.a. varicella, aan dat deze gegevens nuttig zijn. Verschillende methoden, zoals radio-frequentie identificatie en webgebaseerde vragenlijsten werden beschouwd om de data te verzamelen. In deze thesis maken we enkel gebruik van dagboekjes op papier. Het doel van deze thesis is om veranderingen in sociaal contact gedrag onder invloed van verschillende factoren te kwantificeren.

Na een inleidend hoofdstuk beschrijven we in Hoofdstuk 2 vier sociale contact studies. Het opzet van de studies is grotendeels gelijkaardig: men vraagt aan deelnemers om gedurende een (willekeurige) dag een boekje in te vullen dat bestaat uit twee of meer delen. Een eerste deel bevraagt de achtergrond van de deelnemer (leeftijd, geslacht, opleiding,...) en een tweede de contacten. De deelnemers vullen een lijn in voor iedereen die ze ontmoeten. Hierin wordt volgende informatie gevraagd: leeftijd (of leeftijdsgrenzen), geslacht, aanraking of niet, duur van het contact, plaats van het contact, frequentie van het contact (bv. dagelijks of wekelijks). De eerste studie, POLYMOD, is een grootschalige studie die plaatsvond in acht Europese landen en die o.a. gebruikt wordt als (historische) referentie voor de andere studies. De tweede studie, is een studie die in Groot-Brittannië plaatsvond tijdens de Mexicaanse griep in 2009. Hierbij werd aan de deelnemers gevraagd om een eerste boekje in te vullen terwijl ze ziek waren en kregen ze twee weken later een nieuw boekje toegestuurd in de veronderstelling dat ze genezen waren. In deze studie gaven de deelnemers bij hun achtergrond informatie ook informatie over hun ziekte en of ze reeds genezen waren. De derde studie gebeurde in Vlaanderen in 2009-2010 en omvat twee parallelle delen met verschillende deelnemers. Het eerste deel bevroeg individuele personen en het tweede deel bevroeg volledige huishoudens. De boekjes waren gebaseerd op deze van de eerste studie, maar bevatten extra vragen zoals een gezondheidsscore van de deelnemer, de tijdsbesteding en het al dan niet hebben of aanraken van dieren. Verder werd er ook gevraagd om voor ieder contact aan te geven of deze persoon al dan niet lid was van het huishouden. De laatste studie is een pilootstudie, ICSOC, die een gelijkaardige opzet heeft als de tweede beschreven studie. In dit geval werden mensen, via hun huisarts, gevraagd om deel te nemen indien ze seizoensgriep hadden of wanneer ze aan een chronische aandoening leden. De boekjes waren gebaseerd op die van de Vlaamse studie, mits enkele beperkingen in vragen en extra vragen over de ziekte zelf, dit laatste gelijkaardig aan de initiële Engelse studie.

In Hoofdstuk 3 bestuderen we de wisselwerking tussen ziekte en sociaal contact gedrag. In een eerste deel van Hoofdstuk 3 maken we gebruik van de studie in Groot-Brittannië om de verandering in sociaal contact gedrag als gevolg van infectie met A/H1N1pdm influenza te kwantificeren. We observeren dat het basis reproductiegetal (dit geeft weer hoeveel infecties een infectieus persoon kan veroorzaken in een volledig vatbare populatie) van personen in een symptomatische fase een vierde is van de waarde deze personen indien ze asymptomatisch zijn, en dit onder de aanname van gelijke infectiviteit. De verandering in sociaal contact gedrag veroorzaakt een verschuiving in de verwachte verdeling van nieuwe infecties; wanneer we uitgaan van contact gedrag van gezonde personen worden voornamelijk kinderen geïnfecteerd, terwijl dit voornamelijk jongvolwassenen zijn indien we uitgaan van contact gedrag tijdens ziekte. Verder observeren we een verandering in het aantal contacten tussen ziek en gezond die beïnvloed wordt door de plaats van het contact. Zo zien we een vermindering in werk/school- en vrijetijdscontacten, maar geen verschil in het aantal contacten die thuis plaatsvonden.

We beschouwen twee benaderingen om de proportie infecties veroorzaakt door symptomatische individuen, te schatten. In de eerste benadering nemen we aan dat er geen verandering in het contact gedrag gebeurt tijdens de infectieuze periode. We concluderen dat de incidentie van influenza-gerelateerde infecties overeenkomt met een meerderheid aan transmissies veroorzaakt door mensen met symptomen, met name twee derde van de infecties zijn veroorzaakt door symptomatische personen. In de tweede benadering, breiden we de eerste benadering uit door een verandering in het contact gedrag toe te laten tijdens de duur van de infectie door middel van een niet-preferentieel transmissiemodel. Bijkomend breiden we het niet-preferentieel transmissiemodel uit tot een preferentieel transmissiemodel wat het effect van de status van de besmette persoon (symptomatisch of asymptomatisch) in rekening brengt. We tonen aan dat het mogelijk is om de proportie van symptomatische gevallen of de relatieve infectiviteit van symptomatische gevallen t.o.v. asymptomatische gevallen te schatten in het niet-preferentieel transmissiemodel. We maken hierbij gebruik van een Bayesiaanse aanpak en brengen onderrapportering in rekening, maar om de overrapportering te kunnen schatten zijn bijkomende gegevens nodig in minstens één leeftijdscategorie. Bovendien kunnen deze onderrapporteringsfactoren andere factoren dan de intentie om een dokter te bezoeken in rekening brengen en zorgen ze ervoor dat de data beter beschreven zal worden, door te werken met een verborgen laag. We beschouwen sommige parameters als vast op basis van waarden uit de literatuur hetgeen een impact heeft op de verkregen schattingen. We tonen ook een vermindering in het aantal gevallen wanneer mensen thuis zouden blijven onmiddellijk na het vertoon van symptomen en dit voor verschillende scenario's. Om subtielere scenario's te bestuderen, kunnen individu-gebaseerde modellen een oplossing bieden, hetgeen een piste voor toekomstig onderzoek is. Deze conclusies worden gestaafd met de gegevens uit de studie in Groot-Brittannië, die geen representatieve steekproef van de populatie is maar wel epidemiologisch relevant contact gedrag beschrijft. Uit een vergelijking tussen het gedrag van gezonde mensen en het gedrag geschat op basis van de POLYMOD studie blijkt een gelijkaardig gedrag waarneembaar.

De data in Groot-Brittannië werden verzameld tijdens de A/H1N1pdm influenza epidemie, het is echter onduidelijk hoe deze resultaten van toepassing zijn voor andere influenza stammen of andere ziekten. Met het oog op de andere influenza stammen en de andere ziekten, voerden wij een pilootstudie uit, namelijk ICSOC. De resultaten van deze studie beschrijven we tevens in Hoofdstuk 3. Het doel van deze studie was om informatie te verzamelen zoals de initiële studie maar dan voor seizoensgebonden influenza en om bijkomend de basislijn en de verandering in gedrag door chronische infecties te kwantificeren. De ervaring van de huisartsen en de deelnemers leert ons dat de boekjes nogal ingewikkeld waren, waardoor sommige huisartsen hun patiënten selecteerden op basis van intelligentie. Deze vertekening kan leiden tot een niet-representatieve steekproef en aangezien het doel was om een representatieve steekproef te bekomen, is het van belang om een oplossing te voorzien wanneer een gelijkaardige studie op grote schaal wordt uitgevoerd. Ondanks de beperkte grootte van deze pilootstudie, vinden we een verschil in het aantal contacten van mensen met een chronische aandoening en gezonde mensen. Bovendien is het aantal contacten van mensen met een chronische aandoening en een influenza-gerelateerde infectie verschillend van het aantal contacten van een gezonde persoon. Een nipt niet-significant resultaat wordt gevonden wanneer we de hypothese testen dat mensen minder contacten hebben wanneer ze ziek zijn (influenza-gerelateerde aandoening) dan wanneer ze gezond zijn. Deze observatie stemt overeen met de resultaten van de initiële studie. Deze resultaten tonen dus een potentieel onderwerp voor verder onderzoek aangezien een studie op grotere schaal meer zal bijdragen tot inzichten. Eén van de interessante aspecten is het ontrafelen van verschillen in sociale contacten van mensen met een influenza-gerelateerde aandoening of een chronische aandoening en van gezonde individuen, om zo te weten te komen of dit verschil gedreven is door de influenza-gerelateerde infectie, de chronische aandoening of een combinatie van beide. Om dit onderzoek uit te breiden zou men een vereenvoudiging van de dagboekjes of het gebruik van een app-gebaseerde benadering kunnen overwegen. Aangezien de opzet van de studie beïnvloed wordt door huisartsen, kan het inzichten opleveren om focusgroepen te organiseren met huisartsen en aldus hun bezorgdheden mee te nemen in de opzet van de studie en in het ontwerp van de dagboekjes. Bovendien kan een samenwerking met de instelling van volksgezondheid resulteren in het opzetten van de studie over het peilpraktijk netwerk van huisartsen, hetgeen een meer representatieve steekproef kan geven van deelnemers en dus patiënt/deelnemers. Met betrekking tot het gebruik van papieren dagboekjes, zou het gebruik van app-gebaseerde boekjes een alternatief kunnen vormen. De recente technologie van apps laat toe om vragenlijsten op te stellen die gebruik maken van verschillende platformen, zoals internet en smartphones, en laat toe om kwaliteitstesten van de data uit te voeren tijdens de inzameling van de data. Bovendien kan de app de lasten voor de deelnemer verminderen. Tot slot kan een app-gebaseerde studie eenvoudig uitgebreid worden voor andere sociale contact studies, mogelijks in de context van andere infectieziekten. Desondanks wordt deze benadering momenteel nog niet toegepast en is het onduidelijk of deze methode tot gelijkaardige resultaten zal leiden als de studies met papieren dagboekjes. Deze lacune in kennis biedt potentieel voor verder onderzoek. Zoals eerder aangegeven is het nodig om de veranderingen in contact gedrag ten gevolge van andere infectieziekten te kwantificeren, omdat het uitbreiden van de (niet-)preferentiële transmissiemodellen toelaat om de proportie asymptomatische infecties te schatten, hetgeen een invloed zal hebben op de interventies onder beschouwing.

In Hoofdstuk 4 behandelen we drie onderzoeksthema's door gebruik te maken van de Vlaamse contact studie. Zo bevestigen we resultaten van andere studies over het verschil in het aantal contacten tussen weekdagen, en weekenddagen en vakantieperioden. Tevens vonden we een effect van leeftijd en huishoudgrootte en een effect van de eigendom van dieren, geslacht en provincies. Eén deel van het hoofdstuk beschrijft de wisselwerking tussen eigendom of aanraking van dieren en het sociaal contact gedrag. Huishoudgrootte en leeftijd beïnvloeden het al dan niet hebben van dieren, waarbij we observeren dat deelnemers jonger dan vijf jaar minder kans hebben op het bezitten van dieren in vergelijking met deelnemers ouder dan vijf jaar. Verder resulteert alleen wonen op een lagere kans tot het bezitten van dieren in vergelijking met grote huishoudens. De kans om dieren aan te raken hangt af van leeftijd en eigendom van een dier en hun interactie. Op basis van de door ons voorgestelde methode kunnen we de relatieve kans op een grote zoönotische uitbraak bepalen, waarbij we vinden dat kinderen jonger dan tien jaar en volwassenen tussen 25 en 45 jaar meer kans hebben om een uitbraak te veroorzaken.

In het tweede deel tonen we dat het weer een invloed kan hebben op sociaal contact gedrag,waarbij de duur van de contacten lijkt af te hangen van het weer. Concreet observeren we een stijging in de contacten van lange duur op reguliere weekdagen met een lage temperatuur, lage hoeveelheid neerslag of lage absolute luchtvochtigheid. Die laatste observatie is ook gevonden in tijdsbestedingsstudies aangezien mensen langer binnenshuis samenkomen wanneer het slecht weer is.

Het derde deel van het hoofdstuk beschrijft de contact netwerken binnen huishoudens. We schatten een grotere proportie volledige netwerken, i.e. iedereen binnen het gezin heeft met elkaar contact, in vergelijking met voorgaand werk dat steunde op egocentrische sociale contact netwerken. Daarenboven vonden we dat contacten tussen vader en kinderen minder waarschijnlijk zijn in vergelijking met contacten tussen vader en moeder, moeder en kinderen en tussen kinderen (behalve voor oudere broers en zussen), hetgeen in lijn is met het werk op basis van kinkhoest in huishoudens. Eén van de ambities van dit deel van het hoofdstuk was om de aanname van willekeurige mixing binnen huishoudens uit te dagen; echter vinden we initieel geen substantieel verschil op basis van een mixing model met twee niveaus, aldus een indicatie dat de willekeurige mixing aanname redelijk kan zijn. Bijkomend tonen de resultaten van een extremer scenario dat gebruik maken van sociale contact data om het contact gedrag binnen huishoudens te informeren, bepaalde modelleerpogingen kan beïnvloeden onder verschillende situaties.

Dit hoofdstuk behandelde een verscheidenheid aan onderzoeksvragen met nieuwe resultaten alsook nieuwe methodologieën. De voorgestelde methode om de kans op een grote zoönotische uitbraak te bepalen is nieuw en het zou interessant zijn als deze methode toegepast kan worden in gelijkaardige of andere situaties om de resultaten te bevestigen en de methode te testen. Daarenboven maakt de methode sterke aannames, waaronder de aanname dat de infectie kan overgaan van dieren op mensen en vervolgens zich verder kan verspreiden onder mensen, zodat het bestuderen van de invloed van deze aannames kan leiden tot nieuwe inzichten. Het bestuderen van het type contacten dat een groot risico met zich meedraagt, biedt inzichten die belangrijk zijn voor mogelijk niet-farmaceutische interventies. De studie over de wisselwerking tussen weersomstandigheden en sociaal contact gedrag is de eerste in zijn soort en het kan daarom interessant zijn om een sociale contact studie op te zetten die een dieper onderzoek in het onderwerp mogelijk maakt. Deze studie beschouwt best een voldoende groot aantal deelnemers voor de verschillende types van dagen (week, weekend, vakantie) alsook voor de verschillende weersomstandigheden. Eén mogelijkheid om dit te realiseren is om op veel verschillende dagen informatie te verzamelen en tevens een voldoende variatie trachten te bereiken wat betreft het weer. Een andere oplossing houdt in om individuen herhaaldelijk of over een langere periode te bevragen, maar dit kan leiden tot studie moeheid van de deelnemers. Echter, in een app-gebaseerde context kan het aantal herhaalde vragen beperkt worden en kunnen deelnemers eenvoudiger gevolgd worden over een langere periode. Voor specifieke situaties, zoals binnen scholen, kan het gebruik van radio-frequentie identificatie toegepast worden omdat die de nauwkeurigheid verhogen in vergelijking met de dagboekjes werkwijze. Binnen het onderdeel over de huishoudens brachten we de relatie-specifieke heterogeniteit in de duur van de contacten niet in rekening, iets wat relevant kan zijn voor sommige ziekten. Dit kan in de toekomst gebeuren door in de gebruikte modellen een waarde toe te kennen aan de linken tussen de leden van het huishoud. Het exploreren van de verschillende types van contacten binnen huishoudens die verantwoordelijk kunnen zijn voor verspreiding binnen huishoudens, is tevens toekomstmuziek, mogelijks overeenkomstig met voorgaand onderzoek in de context van individueel contact gedrag. De resultaten hiervan kunnen inzichten bieden die nuttig zijn bij gerichte interventies binnen huishoudens. Tot slot kan er in deze context ook gebruik gemaakt worden van de gegevens in verband met de tijdsbesteding van deelnemers om temporele modellen te ontwikkelen. Deze laatste modellen brengen in rekening wanneer contacten gelegd worden en afgelopen zijn. Het is interessant om te onderzoeken wat de implicaties van deze temporele modellen zijn op de verspreiding van ziektes.

Tot slot presenteren we in Hoofdstuk 5 twee parallelle studies over de wisselwerking tussen tijdsbesteding, mobiliteit en contact gedrag enerzijds en de interactie van mobiliteit en contact gedrag anderzijds. In de eerste studie leggen we een link tussen de tijdsbestedingsgegevens en de sociale contact gegevens van de Vlaamse studie, hetgeen resulteert in informatie over de afstanden van thuis waarop contacten plaatsvonden. Gebruikmakend van deze gegevens onderzoeken we de dispersie functie van sociale contacten over afstand en de verandering in sociaal contact gedrag over afstand en de invloed op de verspreiding van infecties. We vinden dat de dispersiefunctie afwijkt van deze gevonden in het werk van anderen. Wij vinden subtiele verschillen in de impact op de verspreiding van infecties wanneer we sociale contactmatrices gebruiken die varieren over de afstand. Wanneer men gebruik maakt van constante contactmatrices over afstand, slaagt men er niet in om deze subtiliteiten te beschrijven. In de tweede studie slagen we erin om het incidentie patroon van seizoensgebonden griep tijdens het 2008/2009 seizoen te beschrijven door variaties in sociaal contact gedrag en in mobiliteit volgens de schoolkalender van 2008/2009 in te werken in een metapopulatie model. Uit de resultaten blijkt de verandering in contact gedrag het belangrijkste te zijn en de kerstvakantie blijkt het meeste effect te hebben op de verspreiding van influenza door de epidemie op het piekmoment te verminderen. We vinden dat een extra week vakantie na de kerstvakantie effectief zou zijn in het beperken van de epidemische verspreiding. Uit variaties in de timing van de epidemie blijkt dat in vroege seizoenen de vakanties een groter effect hebben. Beide studies leveren een significante bijdrage aan het onderzoeksdomein, maar hebben tevens hun beperkingen. Het proces om linken te leggen tussen tijdsbestedingsgegevens en sociale contact gegevens resulteerde in ontbrekende gegevens en de voorgestelde methode om de dispersie van sociale contacten te schatten is nieuw. Oplossingen om het aantal ontbrekende gegevens te beperken, bestaan erin om de tijdsblokken van de tijdsbestedingsstudie te verkleinen, maar dit leidt tot meer last bij de deelnemers. Ook hier kan een app-gebaseerde vragenlijst een alternatief bieden aangezien hiermee de locatie van de deelnemers en mogelijks de contacten gevolgd kan worden. Wat betreft de methode zou het goed zijn om deze toe te passen in gelijkaardige contexten om de toepasbaarheid te testen en tevens de resultaten te bevestigen. Daarnaast kan de methode ook uitgebreid worden naar andere onderzoeksdomeinen waar men tracht een onderliggende continue verdeling te schatten als er slechts categorische gegevens geobserveerd worden. De data voorhanden staan bovendien toe om andere bestaande technieken om contact matrices te schatten op basis van tijdsbestedingsgegevens te valideren, te vergelijken en eventueel uit te breiden. Het effect van variërende contactmatrices over afstand op de verspreiding van infecties werd hier slechts eenvoudig beschouwd. Toekomstig onderzoek kan deze informatie trachten te verwerken in individu-gebaseerde modellen of in het model voorgesteld door Meyer and Held (2017). In het metapopulatie model beschouwden we slechts twee leeftijdsklassen maar waren we hiermee wel in staat om de verspreiding in plaats en tijd weer te geven. Verder was het doel slechts één seizoen te repliceren maar brachten we mogelijk vooruitlopen of vertragingen in de seizoenen in rekening. Aannames wat betreft de mobiliteitsveranderingen gebeurden met gegevens uit Frankrijk. Samenwerkingen met onderzoeksinstellingen in het domein van mobiliteit kunnen helpen om meer gedetailleerde gegevens te gebruiken of om de gebruikte methode te bevestigen. Een mogelijk toekomstig onderzoeksproject kan het model testen in andere seizoenen. Indien dit goed verloopt, kan men een stap verder gaan door te trachten de hospitalisaties in de verschillende ziekenhuizen in het land te reproduceren. Nog een stap verder kan men trachten voorspellingen te maken van de tijden waarop er een overvloed aan hospitalisaties gebeurt. Dit kan een hulp zijn voor overheden en instanties om beslissingen te nemen over hoe ze het aantal hospitalisaties kunnen verminderen of hoe ze zieke mensen gepast kunnen helpen.

Buiten de specifieke beperkingen die reeds vernoemd werden, zijn er algemene limitaties in de onderzoeken, die mogelijke richtingen voor toekomstig onderzoek bieden. In het geheel van deze thesis is het effect van geslacht niet bestudeerd, maar men verwacht dat dit een effect heeft, zowel op het contact gedrag als op de mobiliteit, de tijdsbesteding en het verloop van de infecties. Daarenboven kunnen nieuwe maten om contactmatrices te vergelijken een meerwaarde bieden als zij trachten epidemiologisch relevante aspecten weer te geven en bij voorkeur niet enkel in de vorm van één getal. Dit is echter niet evident aangezien samenvatten in een vector gepaard kan gaan met problemen van meervoudige testen. Tot slot is een grote aanname hier dat we gebruikmaken van de dagboekjes op papier. Hoewel deze voldoende voordelen hebben, zijn er tevens ook nadelen aan verbonden die we in Hoofdstuk 6 bespreken. Deze thesis focust zich voornamelijk op veranderingen in contact gedrag en de effecten hiervan op de verspreiding van ziektes. Dit alles is belangrijk vanuit het oogpunt om ziektes te beperken door gebruik te maken van interventies. Echter zijn er ook andere factoren die een invloed kunnen hebben op het al dan niet uitvoeren van interventies die wij niet belicht hebben. Zo is verder onderzoek naar het gedrag van mensen naar preventie en controle van infectieziekten van belang. Ook het gebruiken van de gevonden effecten in modellen kwam hier niet aan bod, en dient er nagedacht te worden hoe en in welke mate dit best zou gebeuren. Dit zijn belangrijke richtingen voor verder onderzoek omdat zij een belangrijke overgang van theoretische modellen naar relevante modellen voor beleidsmakers toestaan.