Transnational University Limburg is a unique collaboration of two universities in two countries:
the University of Hasselt and Maastricht University.

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Modelling health and socio-demographic indicators of social contacts in the Flranders, Belgium

Supervisor :
Prof.dr. Christel FAES
Prof. dr. Niel HENS

Nicholas Tendongfor
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Modelling health and socio-demographic indicators of social contacts in the Flranders, Belgium

Supervisor :
Prof.dr. Christel FAES
Prof. dr. Niel HENS

Nicholas Tendongfor
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Dedication

This piece of work is dedicated to my beloved daughter

Anoboundem Presley Tendongfor

Acknowledgement

My profound gratitude goes to my promoters, Professors Niel Hens and Christel Faes for giving me the opportunity to exploit this topic. Their constant attention, supervision and recommendations were the driving force to the accomplishment of this work.

To the entire teaching staff of the Biostatistic/EPI program, I would say there would be no thesis without the knowledge and skills which you patiently and carefully imparted in me.

To Yimer Wasihun Kifle and Kim Van Kerckhove of the I-BioStat, for providing me with cleaned social contact study data and for their advice during the entire work.

My profound gratitude goes to my family for their unending love, care and support throughout this journey.

Finally, to my friends and colleagues of the Biostatistics program, I am thankful for meeting people like you. Your encouragement and cooperation throughout the study duration was an amazing platform for me to reach my goal.

Table of Content

Abstract

Background

Knowledge of social mixing patterns among populations is important in understanding the transmission of infectious diseases and their control. In this study, we model the association between individual health status, socio-demographic characteristics and the number of social contacts using data collected from a diary based social contact survey collected between September 2010 and February 2011 in Flanders, Belgium.

Method

The total number of social contact during a randomly assigned day was modelled using a Generalised additive model. The analysis consisted of fitting models to the whole data and diary based stratified data and selecting important predictors of social contacts.

Findings

Significant health indicators associated with the number of social contacts were daily activity, self-care, pains and anxiety. These indicators varied from one diary type to another. With the exception of self-care in elderly people, poorer health status negatively affected the number of social contacts. Socio-demographic indicators retained also varied from one diary type to another. The rate of social contacts increased significantly with household size, education level and in those who own animals, but decreased with age and during the weekends and holiday periods.

Conclusions

This study demonstrates that individual health status and socio-demographic characteristics influence social contact patterns. These finding could be exploited in the building of mathematical models for transmission of epidemics and the development and implementation of control strategies.

1. Introduction

Disease transmission is prerequisite for the survival of pathogens. Some infectious diseases, such as influenza and measles, which spread via air droplets, require close encounters between infected and uninfected individuals for efficient transmission (Salathé *et al.,* 2010). Social mixing patterns are useful in the estimation of transmission parameters for airborne infections and disease transmitted by direct contacts (Wallinga *et al*., 2006). Contacts between individuals happen several times per day, increasing the risk of disease transmission. The way people live, their demographic and socio-cultural characteristics influence their mixing patterns. During the recent Ebola epidemics in West Africa, the poor knowledge of social contact among populations, socio-cultural behaviour and population mobility exacerbated the epidemics (Chowell and Nishiura 2014). Previous studies on social contacts patterns described associations between household size, class size for children as well as professional contacts in adults (Mossong *et al*., 2008, Hens *et al*., 2009, Eames *et al*., 2011). Social contact data are valuable source of information when building mathematical models for the transmission of infectious diseases (Wallinga *et al.,* 2006; Ogunjimi *et al*., 2009; Goeyvaerts *et al.,* 2010).

In a recent study, Van Kerckhove *et al*., (2013) demonstrated an association between influenza illness and a reduction in the number of social contacts, underlying the importance of population health on their social mixing patterns. It is not known what could be the impact of health indicators such as mobility, self-care, anxiety, daily activities and general health perception on the mixing patterns in the population.

A social contact survey involving 1774 participants was conducted between September 2010 and February 2011 in the Flanders, Belgium. During this survey, information on participant's health status, socio-demography and the number of social contacts was collected. The analysis of the data from this survey revealed that social contacts vary with weather conditions (Willem *et al*., 2012) and an association between animal ownership and social contacts has also been described (Kifle *et al.,* 2015).

In this report we model the impact of health status and socio-demographic characteristics of participants on the number of social contacts with the aim of identifying possible relationships between health status, socio-demographic indicators and the number of social contacts. The first section describes the data used in the study, followed by the methods of data analysis. The last section deals with presentation and discussion of the results.

2. Data

The data used in this study came from a social contact survey conducted in the Flemish region of Belgium from September 2010 until February 2011. In total 1714 participants were involved in the study. Participants were recruited by random digit dialing on mobile phones and landlines. Only one person per household was included in the survey. All participants were asked to fill in a paper diary recording their contacts during one randomly assigned day. In addition to their daily contacts, participants were also asked to fill in information about their health status (mobility, pains, daily activities, anxiety, self-care and health status grading). The diaries also captured socio-demographic information on all participants.

A contact was recorded when a participant engaged in a direct conversation with someone else at most three meters away or physical contact if a participant touched someone else (e.g. shaking hands) even without speaking. In addition to inter-human contacts, participants were requested to complete enquiries about human-animal interactions (animal ownership and animal touching).

Three types of diaries adapted to the age of participants (0-12 years, 13-60 years and 60+ years) were used. The diary for children less than 13 years and for elderly people (> 60 years) was completed by proxy (relatives, caregivers and teachers). Sampling days were nearly uniformly distributed between all days of the week. The number of contacts was defined as the total number of contacts including professional contacts (physical and non-physical) reported by a participant during the assigned day. **Table 1** shows the health and socio-demographic predictors used in the models.

Table 1: Health and socio-demographic predictors

Variable name	Abbreviation	Code
1.Health predictors		
Mobility	mobility	1 = No problems 2 = Some problems 3 = Bedridden
Self-care	selfcare	1 = No problems 2 = Some problems 3 = Unable
Daily activity	dailyactivity	1 = No problems 2 = Some problems 3 = Unable
Pains	pains	1 = No pain 2 = Moderate pains 3 = Severe pains
Anxiety	anxiety	1 = Not anxious 2 = Moderately anxious 3 = Very anxious
Health grading	hgrading	1 – 100
2. Demographic predictors		
Gender	gender	1 = Male; 2 = Female
Age group	agegroup	1 = 0 – 5 years 2 = 6 – 11 years 3= 12 – 17 years 4 = 18 – 44 years 5 = 45 – 64 years 6 = 65+ years
Province	province	1 = Flemish Brabant 2 = Antwerp 3 = Limburg 4 = West Flanders 5 = East Flanders
Education level	edulevel	1 = No formal education 2 = Never/Primary 3 = Vocational 4 = Lower technical 5 = Lower secondary 6 = Higher technical 7 = Upper Secondary 8 = Higher non-University 9 = University
Smoking status	smoke	1 = Smoker; 2 = Ex-smoker; 3 = Non-smoker
Alcohol consumption	aclcohol	1 = Yes; 2 = No
3. Temporality indicators		
Week period	weekperiod	1 = School day; 2 = Weekend
Period	period	2 = Holiday period; 1 = School period
4. Animal ownership and touching indicators		
Animal ownership	ownanimal	2 = No; 1 = Yes
Touching animal	touchanimal	2 = No1 = Yes

3. Methodology

3.1 Modeling the number of social contacts.

In this study, we exploited three models of count data (Poisson, Negative binomial and Generalized additive models) to investigate the association between the number of social contacts, health and demographic predictors. The Poisson regression and the Negative binomial model models were not appropriate because the response variable exhibited over-dispersion and they could not also account for the non-linearity of the health grading predictor. The Generalized additive model (Hastie and Tibshirani, 1986) which employs a class of equations called "smoothers" was the most appropriate model for the data. It provides a flexible method for identifying nonlinear covariate effects in exponential family models. The model has the following structure:

$$g(\mu) = b_0 + f_i(x_{i1}) + f_i(x_{i2}) + \cdots + fm(x_{ip})$$

$$y \sim Exponential\ family(\mu)$$

$$\mu = E(y)$$

where $f_i(x)$ are smooth functions of the covariates $\boldsymbol{x_i}$ and $g(\mu)$ is the link function. y is the observed value and is assumed to be of some exponential family distribution and b_0 is the intercept.

Different smoothing functions (**Appendix 3**) were used on the predictor “health grading”.

The GAM model was implemented in the R package “mgcv” (Simon Wood, 2006).

A stepwise procedure was used to select the final model. The significance of the covariates was checked using the WALD test statistics whereas the overall significance of the predictors in the model was checked using an ANOVA test. Significant predictors and two-way interactions were retained in the final model. The final model selection was based on AIC criterion (Akaike 1973).

Model diagnosis was done using graphical displays implemented in the gam.check R function. The variance inflation factor (VIF) was used to check for multi-collinearity among predictors (http://scg.sdsu.edu/logit_r). Outlying observations were investigated using Cook’s distance (Cook, 1977).

The total number of social contacts in each household was weighted to account for overrepresentation of households with large number of people or underrepresentation of households with fewer people. Information on the age structure and household size was used to calculate the weights. Census data from 2001 Belgium population was used as reference.

3.2 Distribution of frequency of social contact in health indicators

In order to gain insight of the distribution of the frequency of social contacts in in different health indicators, we verify if the occurrence of events followed a power law (heavy tail) distribution. We compared the empirical cumulative distribution of the frequency of social contacts with the power law simulated data using the Goodness of fit statistic test (Colin S. Gillespie, 2015). The empirical distribution was generated using the lognormal and exponential distribution for health grading and the lognormal distribution for the other health indicators.

3.3 Software

The data was analysed using the statistical analysis software R version 3.2.2. A significance level of 5% was used for statistical decision making.

4. Results

4.1. Exploratory data analysis

Figure 1 below presents the plot of the distribution of the number of social contacts in the sampled population. A greater fraction of the population made between 1 and 20 contacts per day.

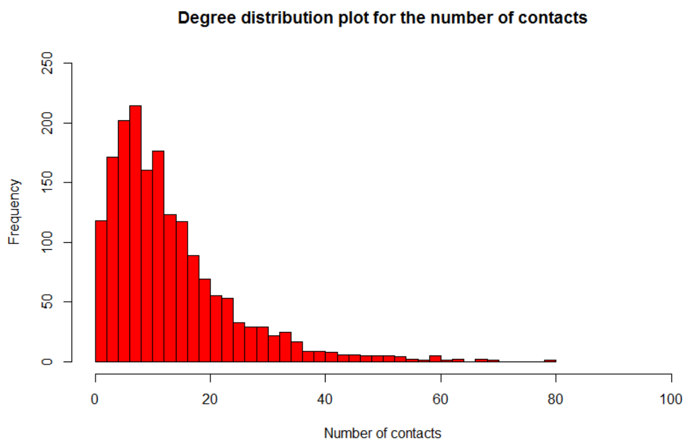
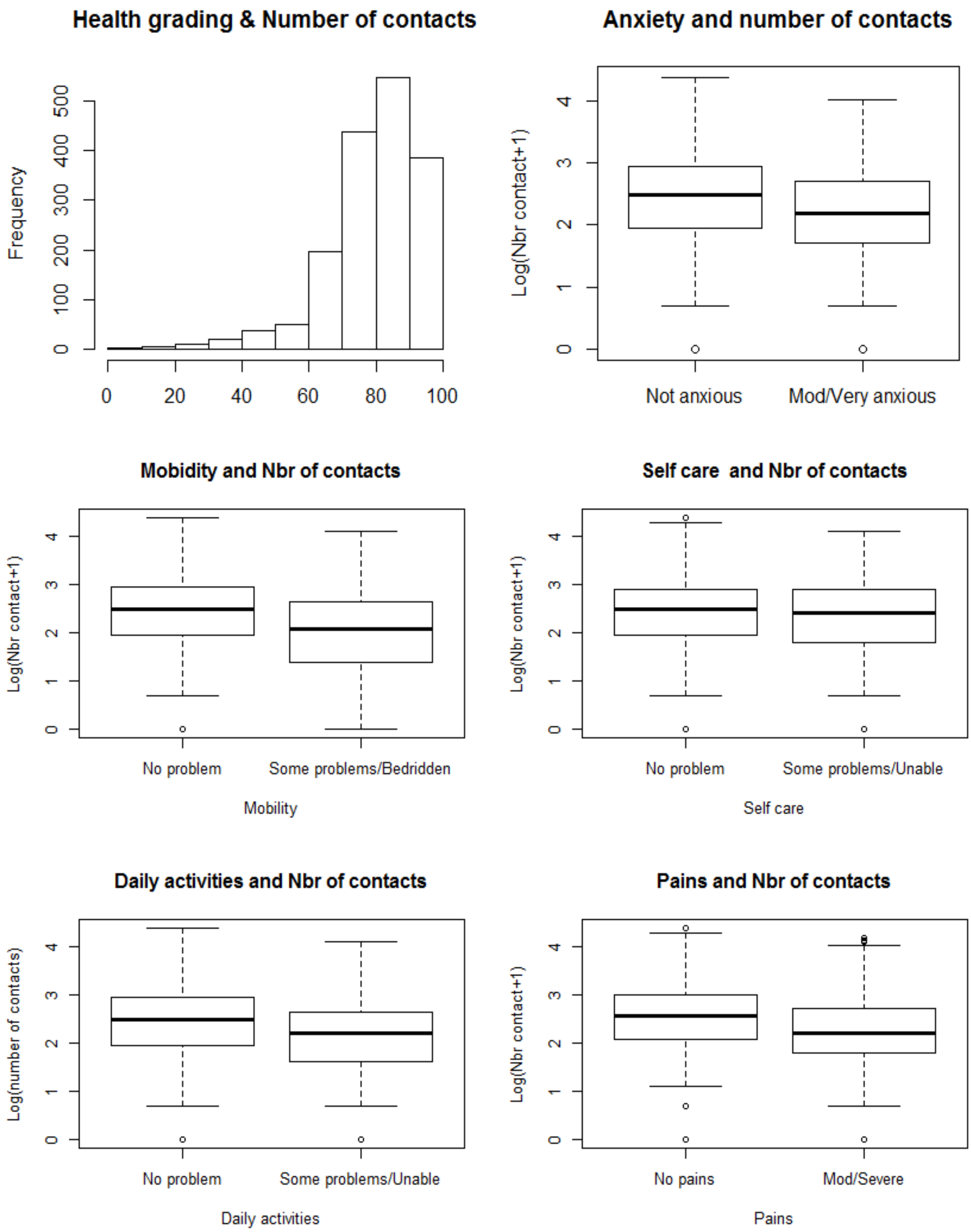


Figure 1: Distribution of the number of social contacts in the sampled population.

Figure 2 shows the plots of health indicators and the number of social contacts. The number of social contacts seems to reduced with decrease in mobility, inability to take care of self, severity of pains and inability to carry out daily activities. The number of social contacts appeared to increase with good health grading. Anxiety seems to have no effect on the number of social contacts. There appear to be some outliers in the number of social contacts in the health status indicators.

The number of social contacts seems to increase with household size, education level, animal ownership and decrease with age and smoking status. It seems to vary little with gender, province, touching of animals (**Figure 3**). There appear to be few outlaying observations in the number of social contacts in socio-demographic indicators. There seem to be a reduction in the number of social contacts during holiday and weekend periods (**Figure 4**).



Figur 2: Health status and number of social contacts in the Flanders, 2010-2011.

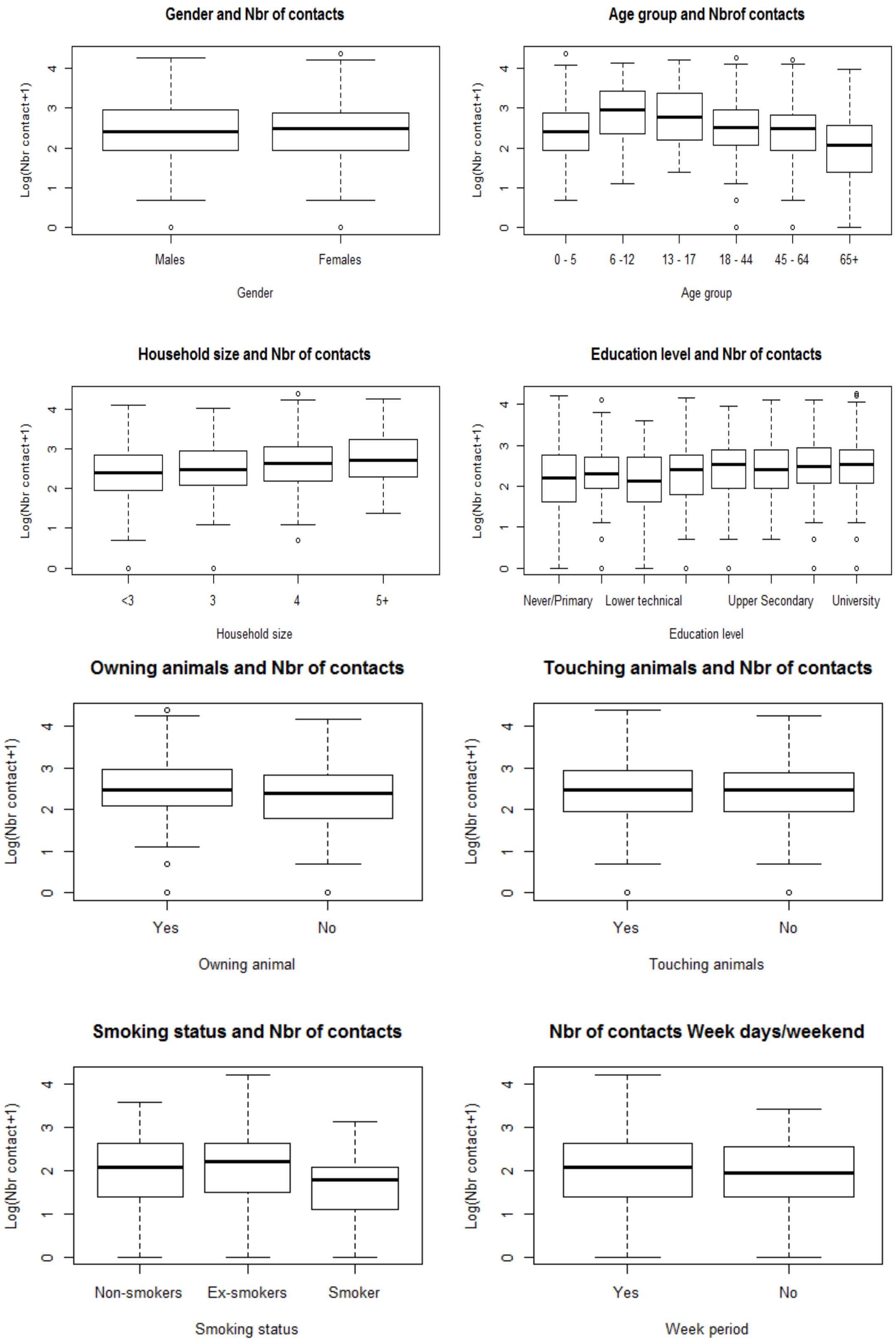
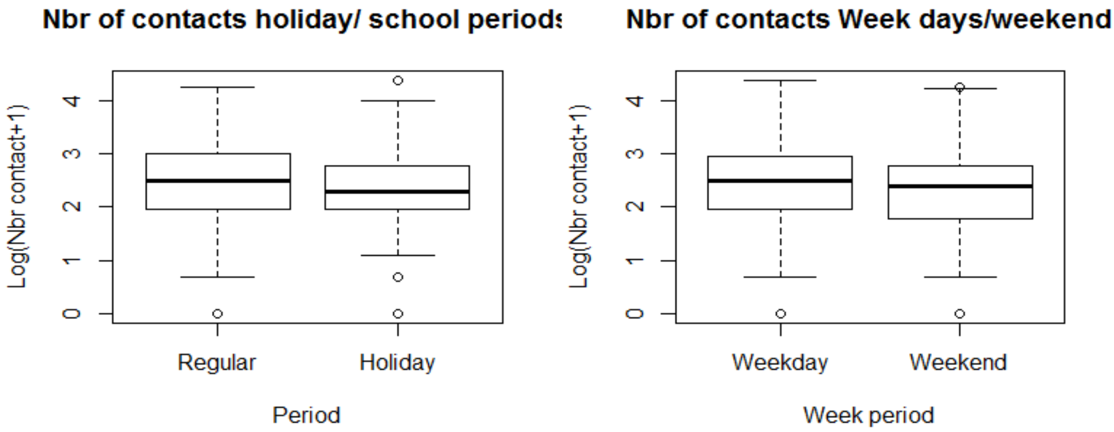


Figure 3: Plots of socio-demographic indicators and the number of social contacts in the Flemish population.



Figur 4: Number of social contacts in different day types (regular/holiday and weekend/ work day).

4.2 Modelling the number of social contacts

The Poisson regression model was not appropriate for the data because of over-dispersion (μ= 13.48 δ^2= 116.79). The Negative binomial was not also appropriate because it could not account for non-linearity in the health grading predictor. The GAM model fitted the data better with an AIC of 8306.1. **Figures 9** and **10** (**Appendix 2**) show model diagnostic plots. The residuals were approximately normally distributed. No observation was identified as outlier with Cook's distance. A test of multiple collinearity failed to reject any covariates in the final model since all variance inflation factors were <1.2.

4.2.1 Heath predictors of social contacts.

The health predictors of social contacts retained in the combined data model were daily activity and health grading (**Table 2**). The number of social contacts decreased in individuals who reported having problems or were unable to carry out their daily activities. A positive interaction between animal ownership and health grading was observed. **Figures 5** shows the visualization plots of the main effects of the health indicator (daily activity) retained in the model.

4.2.2 Socio-demographic predictors of social contacts.

The socio-demographic predictors of social contacts retained in the combined model are shown in **Table 2**. The number of social contacts made decreased during weekends and holiday periods and also with age. Social contacts increased with household size, education level and animal ownership. The model retained positive interaction between week period and animal ownership and between age group and week period.

Table 2: Generalised additive model with a smoother on health grading predictor

	n	**Estimate**	**S.E**	**RR**	**95% Confident interval**	
Intercept		2.687	0.268			
Daily activity						P=0.011
No problem	1529			1		
Some problems/Unable	228	-0.198	0.078	0.820	**0.704**	**0.955**
Health grading						P=0.401
Grading	1690	0.003	0.003	1.003	0.997	1.009
Age group						P<0.001
0-17 years	383			1		
18-44 years	626	-0.706	0.109	0.494	**0.399**	**0.611**
45-64 years	469	-0.754	0.110	0.471	**0.379**	**0.584**
65+ years	296	-1.163	0.152	0.312	**0.232**	**0.421**
Education level						P<0.001
Never/Primary	151			1		
Vocational	154	0.342	0.108	1.408	**1.139**	**1.739**
Lower technical	82	0.084	0.110	1.088	0.877	1.348
Lower secondary	89	0.129	0.108	1.137	0.920	1.407
Higher technical	164	0.362	0.107	1.436	**1.165**	**1.770**
Upper Secondary	217	0.300	0.101	1.350	**1.108**	**1.645**
Higher non-University	408	0.478	0.096	1.613	**1.336**	**1.948**
University	178	0.408	0.103	1.503	**1.227**	**1.842**
Household size						P<0.0001
Size =1 & 2	416			1		
Size=3	330	0.079	0.051	1.082	0.980	1.196
Size=4	441	0.157	0.052	1.170	**1.057**	**1.296**
Size=5+	217	0.275	0.062	1.317	**1.167**	**1.486**
Week period						0.003
Week day	1350			1		
Weekend	422	-0.455	0.154	0.634	**0.469**	**0.858**

Period						P<0.0001
Regular	1351			1		
Holiday	421	-0.213	0.046	0.808	0.738	0.885
Own animal						P=0.004
Yes	1055			1		
No	707	-0.863	0.295	0.422	**0.236**	**0.753**
Own animal and health grading						0.0047
No : Grading	707:1690	0.010	0.004	1.010	**1.003**	**1.017**
Age group and Week period						P<0.0001
18-44 years : Weekend	626:422	0.581	0.161	1.788	**1.305**	**2.449**
45-64 years : Weekend	469:422	0.755	0.167	2.127	**1.534**	**2.951**
65+ years : Weekend	296:422	-0.332	0.282	0.717	0.413	1.245
Week period and own animal						P<0.0001
Weekend : No	422:707	-0.412	0.092	0.662	**0.553**	**0.793**

AIC= 8306.1; R-sq.(adj) = 0.164 Deviance explained = 23.8%

4.3 Social contacts in different diaries.

The data was stratified into three different age groups based on the diary types (0 – 12 years, 13-60 years and > 60 years) and three GAM models were fitted to each stratum. In the age group 60+ years, the effect of smoking status and alcohol consumption was also investigated. **Table 3** summarises the health and socio-demographic predictors identified per diary types whereas **Table 4, 5** and **6 (Appendix 1)** show the estimates of social contacts in children, adults and elderly people.

Health indicators associated with social contacts were daily activity (in children and adults), anxiety (in adults), self-care and pains (in elderly). The number of contacts decreased with decreased ability to carry out daily activity in adults and children and with increase anxiety in adults. Elderly people who reported having pains had less social contacts whereas those who reported having problems taking care of themselves had more contacts. **Figure 5** shows the plots of main effects of health indicators retained in the final models.

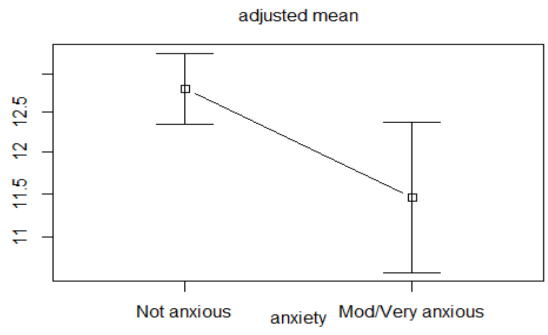
With the exception of elderly people who had increase social contact during weekend, the number of contacts decreased during weekends and holiday periods in children and adults. The social contacts also decreased significantly with age in adults. A positive association between social contacts and animal ownership was observed in children and adults, and with age and week period in adults. In adults, those who did not have animals made fewer contacts at weekend. There was a positive interaction between contacts made during weekend and holiday periods in children. Adults without animals and who were anxious made more contacts. In elderly people, there was a negative interaction between self-care and smoking status as well as self-care and province. A positive interaction between province and smoking status was also retained. **Figure 6** shows the plots of main effects of socio-demographic indicators in different diaries. **Figure 7** shows the plots of important interactions retained in the models.

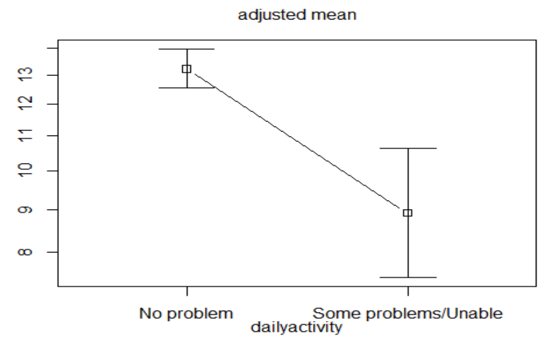
Table 3: Summary of the relationship between the number of social contacts and the health and demographic indicators in different diaries.

	Trends in the associations			
Indicators (baseline)	**Children**	**Adults**	**Elderly**	**Overall**
Health indicators				
Daily activity (No problem)	↓	↓	-	↓
Self-care (No problem)	-	-	↑	-
Pains (No problem)	-	-	↓	-
Anxiety (No problems)	-	↓	-	-
Health grading	-	-	-	-
Socio-demographic indicators				
Age group (0-12 years)	/	↓	/	↓
Education level (Never/primary)	/	↑	-	↑
Household size (1 & 2)	-	↑	/	↑
Smoking status (No)	/	/	-	/
Province (Flamant Brabant)	-	-	-	-
Temporality indicators				
Week period (week day)	↓	↓	↑	↓
Period (Regular)	↓	↓	-	↓
Animal ownership				
Own animal (Yes)	↑	↑	-	↑
Interactions				
Own animal and health grading	-	↑	-	↑
Age group and Week period	-	↑	-	↑↓
Week period and own animal	-	↑	-	↑
Week period and Period	↑		-	-
Anxiety and Own animal	-	↓	-	-
Self-care and smoking status	/	/	↓	/
Province an smoking status	/	/	↑↓	/
Self-care and Province	-	-	↓	-

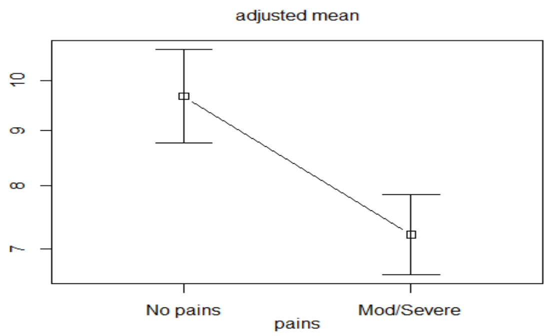
↓ : Negative association; ↑: Positive association; **/**: Not enough data; **-** : Not significant;
↑↓: Association changes from one level to another.

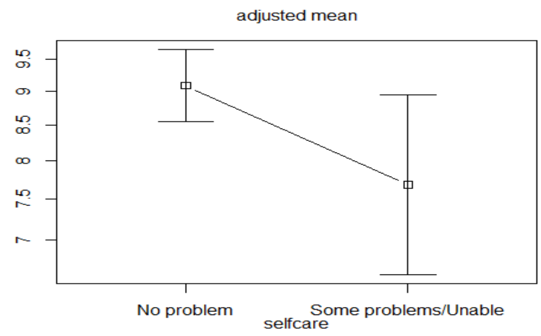
Adults



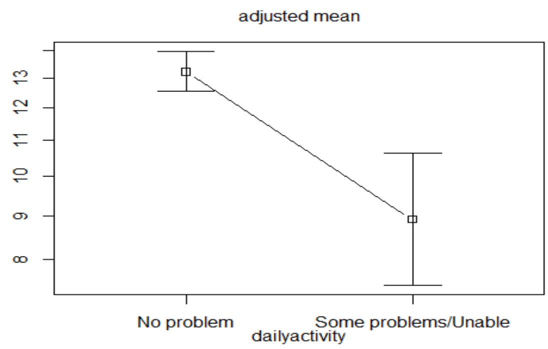


Elderly





Children



Combined

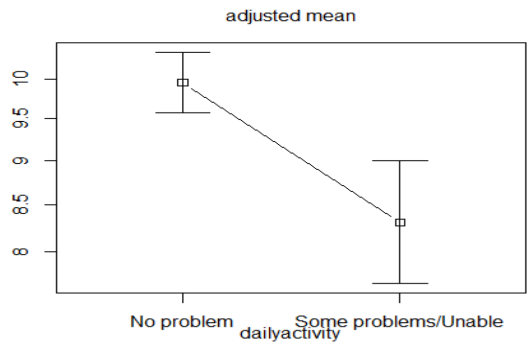


Figure 5: Plots of main effect of health indicators retained in the combined and stratified model

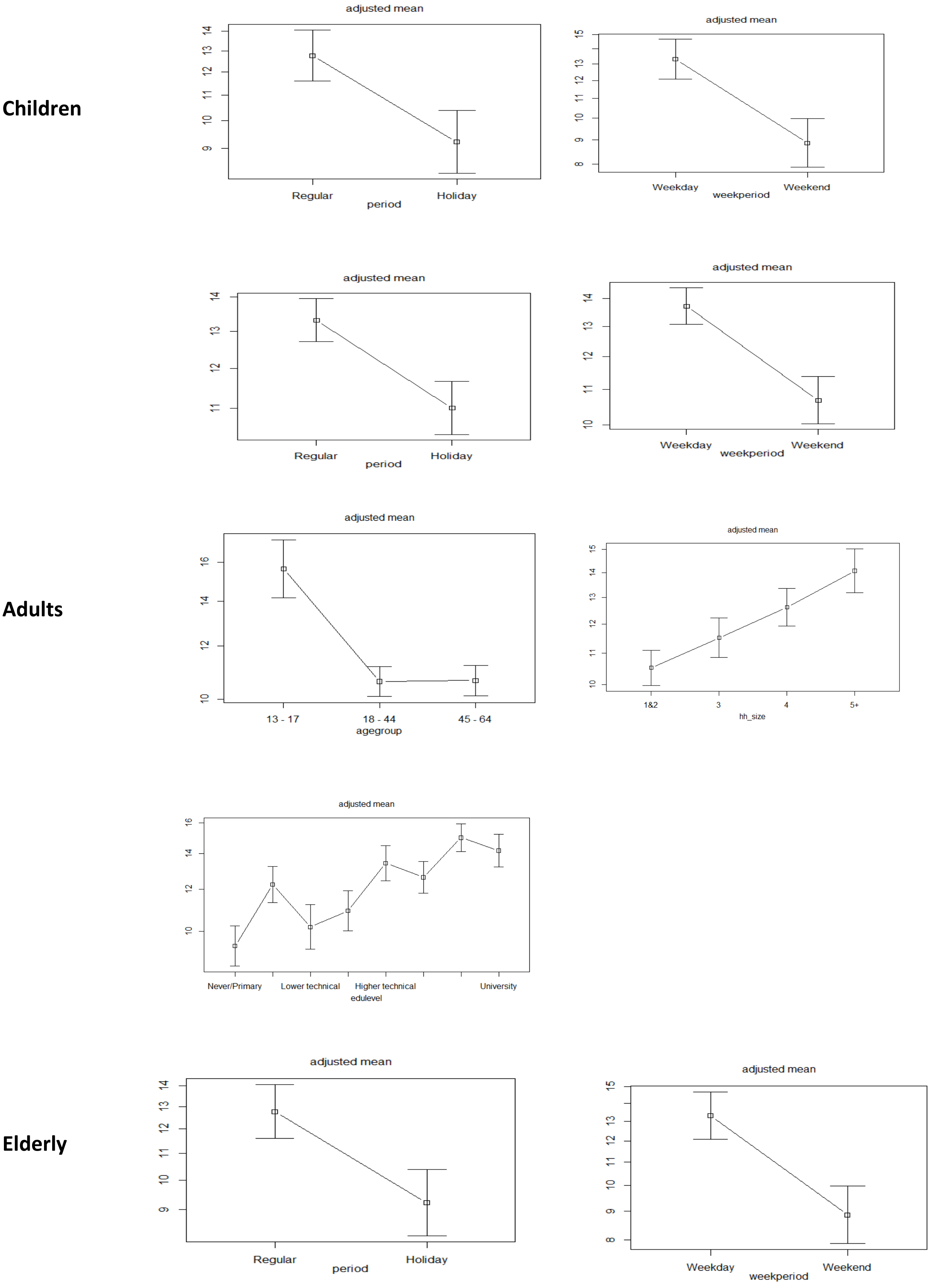
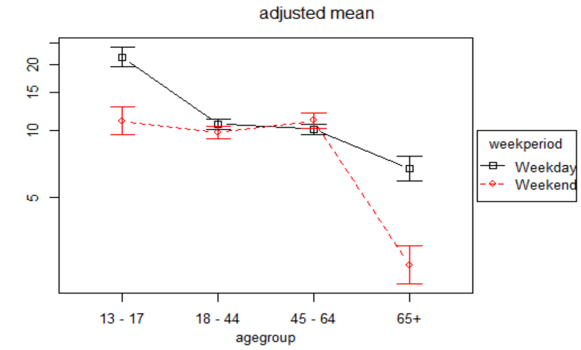
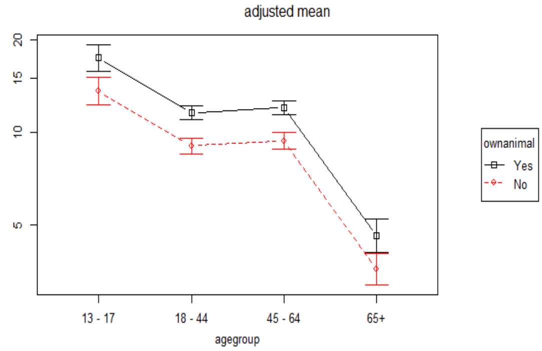


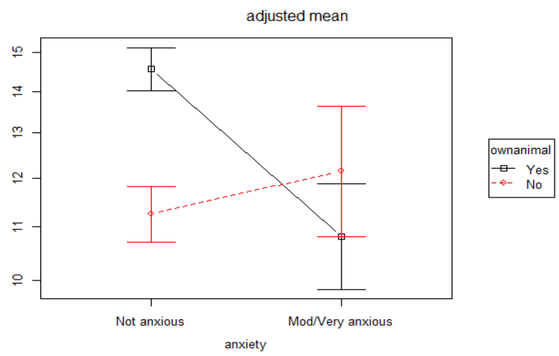
Figure 6: Plots of main effects of socio-demographic predictors retained in the model

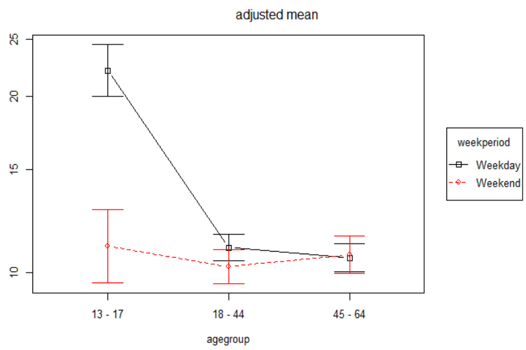
Combined



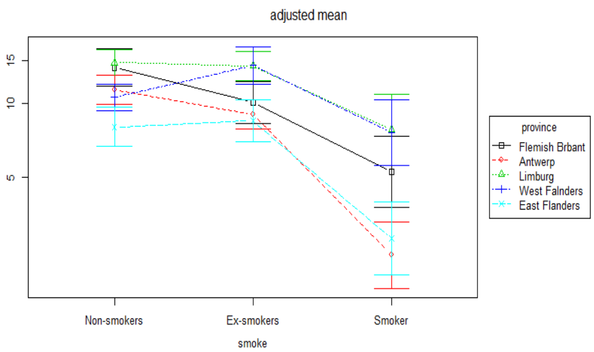


Adults





Elderly



Children

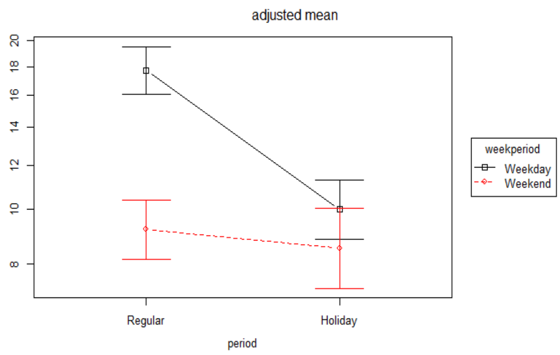


Figure 7: Plots of important interactions retained in the combined and stratified analysis.

4.4 Distribution of the frequency of social contacts in health indicators

Figure 8 shows the plots of the simulated data cumulative distribution and empirical distributions. In all the cases there was a poor fit between the simulated power law data and the empirical data. The Goodness of fit test rejected the hypothesis that the frequency of events in the health predictors followed a power law distribution.

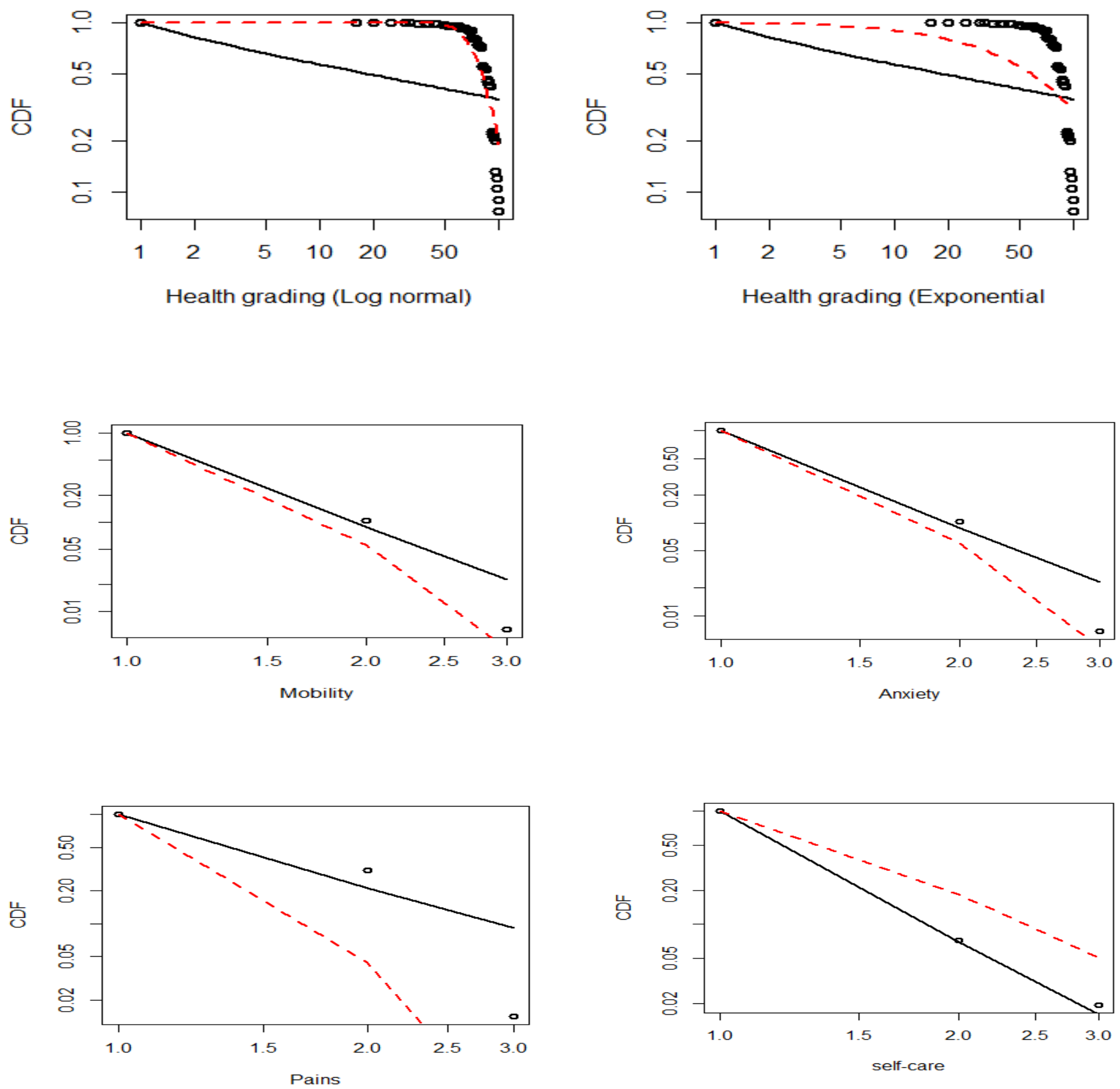


Figure 8: Plots of the simulated data CDF of power law (line) and log normal (circle) with lines of best fit (dash line).

5. Discussion

Patterns of social mixing are key determinants of epidemic spread. This studied aimed at modelling the social contacts in 1774 participants from the Flanders region in Belgium and to identify potential health and socio-demographic predictors of social contacts. The Poisson model was a poor candidate model because of over-dispersion. The negative binomial model was a potential model, but could not account for the non-linearity of some predictors in the model. The model that was suitable for the data was the Generalised additive model.

Previous studies on social contacts in Belgium modelled the number of social contacts, demographic, temporality, seasonal and animal ownership (Hens *et al.*, 2009, Willem *et al*., 2012; Kifle *et at*., (2015). In this study, in addition to socio-demographic indicators, we explored the relationship between the health status of participants and their number of social contacts. This study revealed that the number of contacts people have per day is significantly affected by their health status. These findings are in line with a recent study (Eames *et al*., 2013) that demonstrated an association between Illness and a reduction in the number of social contacts. In elderly people the number of social contact increase in those who were unable to take care of themselves. This group of people probably receive more attention because of their poor health conditions and hence increase in social contacts. Health indicators of social contacts varied across diary types indicating the importance of the population structure in the understanding and modelling of the transmission of infectious diseases and the development of control strategies.

The number of social contacts recorded at weekend and during holiday period reduced significantly. This is a clear indication that the weekends and holidays could be periods of low disease transmission in case of epidemics outbreaks. These results are in line with previous studies on social contact in Belgium (Hens *et al.,* 2009). It was reported that the 2009 influenza epidemics declined by 40% during school holidays compared to school periods (Eames *et al*., 2012, Eames *et al*., 2011). Isolation and quarantine measures from historical background have proven to be very successful in the control of major epidemics and pandemics (*Ou et al., 2003)*. Isolating or keeping people at home during epidemics can help reduce the number of social contacts and hence reduce disease transmission (Van Kerckhove *et al*., 2013). In the recent epidemics of Ebola in West Africa (Chowell *et al*., 2014), closing of schools, isolation and quarantine measures had a positive impact in the decline of transmission of the epidemics.

Transmission Models of infectious diseases have always made assumption on social contact patterns between individuals with the "Who get information from who" ("WAIFW") matrix (Anderson and May*,* 1991). These assumptions are used because of lack of information on social contact patterns among populations. Increase interest on social mixing patterns is making available valuable data that is increasingly being used in more accurate estimation of transmission parameters such as the basic reproductive number (Ro) (Hens *et al.*, 2009, Willem *et al*., 2012).

Human infectious diseases are not only transmitted from humans to humans but also from animals to humans. Humans can catch diseases from infected animals (Taylor *et al*., 2001, Woolhouse *et al.,* 2005). These infections generally called zoonotic diseases can be life threatening. In this study, the number of contacts increased significantly in individuals who own animals. These animal owners constitute a good link between zoonotic infections and humans and can play a major rule in the spread of these diseases. Recent studies by Kifle *et al.,* (2015) emphasized on the importance of animal-human contacts in the modelling of zoonotic infections. A better understanding of the animal-human contact will contribute greatly to the improvement on the modelling of the transmission of zoonotic infections and their prevention.

6. Conclusion.

This study aimed at modelling the social contacts in the sample population from the Flanders in Belgium and to identify health and socio-demographic indicators of social contacts. This study demonstrates that individual health status and socio-demographic characteristics influence the number of contact they make per day. These finding could be exploited in the modelling of transmission of epidemics and development of control strategies.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petran and F. Csaaki (Eds.), *International Symposium on Information Theory, Akadeemiai Kiadi, Budapest, Hungary*, 267–281.

Anderson, R., & May, R. (1991). Infectious diseases of humans: dynamics and control. *Oxford University Press, Oxford.*

Chowell, G. & Hiroshi N. (2014). Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Medicine*, 12, 196.

Colin, S. G. (2015). Analysis of Heavy Tailed Distributions. Package 'poweRlaw'. Version 0.50.0. https://cran.r-project.org/web/packages/poweRlaw /poweRlaw.pdf#page.2 (accessed 28/08/2015)

Cook, R. D. (February 1977). "Detection of Influential Observations in Linear Regression". *Technometrics (American Statistical Association),* 19 (1), 15–18. doi:10.2307/1268249.

Eames K.T.D., Tilston, N. L., & Edmunds, W. J. (2011). The impact of school holidays on the social mixing patterns of school children. *Epidemics,* 3, 103–8.

Eames, K., T. D. , Natasha, L. Tilston, Ellen B. , & John, E. W. (2012). Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza. *PLoS Comput Biol*. 8(3). e1002425. doi: 10.1371/journal.pcbi.1002425.

Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P., & Beutels,P. (2010). Estimating infectious disease parameters from data on social contactsand serological status. *Appl. Stat*, 59, 255–277.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models (with discussion). *Statistical Science* 1, 297–318.

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

http://www.reuters.com/article/2015/08/07/belgium-agriculture-tuberculosis-idUSL5N10I21V20150807 (accessed August 2015).

Kifle, Y. W., Goeyvaerts, N., Kim, V. K., Willem, L. , Faes, C. , Leirs, H. , Hens, N. , & Beutels, P. (2015). Animal Ownership and Touching Enrich the Context of Social Contacts Relevant to the Spread of Human Infectious Diseases. *PLoS One*, 10(7), e0133461.

Mossong J., Hens N., Jit M., Beutels P, Auranen K., et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*. 5:e74.

Ogunjimi, B., Hens, N., Goeyvaerts, N., Aerts, M., Van Damme, P., & Beutels, P. (2009).Using empirical social contact data to model person to person infectious diseasetransmission: an illustration for varicella. *Math. Biosci,* 218, 80–87.

Ou J., Li, Q., & Zeng, G. (2003). Efficiency of quarantine during an epidemic of severe acute respiratory syndrome—Beijing, China, 2003. *MMWR Morb Mortal Wkly Rep, 52*, *1037*–40.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & 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,*
doi: 10.1073/pnas.1009094108

Statistical Consulting Group, San Diego State University: http://scg.sdsu.edu/logit_r/ (accessed 16/08/2015)

Van Kerckhove, K., Hens, N., John E.W., & Eames K. T. D. (2013). The Impact of Illness on Social Networks: Implications for Transmission and Control of Influenza. *American Journal of Epidemiology Advance Access published*. doi: 10.1093/aje/kwt196

Taylor, L. H., Latham, S. M. & Woolhouse, M. E. (2001). Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci*, 356: 983-989.

Wallinga, J., Teunis, P. & Kretzschmar, M. (2006). Using data on social contacts toestimate age-specific transmission parameters for respiratory-spread infectiousagents. Am. J. *Epidemiol*, 164, 936–944.

Willem, L., Van Kerckhove K., Chao, DL, 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. doi:10.1371/journal.pone.0048695.

Wood, S.N., (2006). Low rank scale invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4), 1025-1036.

Woolhouse, M. E., & Gowtage–Sequeria, S. (2005). Host range and emerging and reemerging pathogens. *Emerg Infect Dis*, 11,1842-1847.

Appendix 1: Estimates of GAM in different diary types

Table 6: Generalised additive model for social contact in children diary (0 – 12 years).

	n	Estimate	S.E	RR	95% C.I	
Intercept		3.173	0.058			
Daily activity						
No problem	290			1		
Some problems/Unable	17	-0.395	0.177	0.674	**0.674**	**0.476**
Week period						
Week day	232			1		
Weekend	85	-0.655	0.109	0.520	**0.520**	**0.420**
Period						
Regular	241			1		
Holiday	76	-0.572	0.109	0.565	**0.565**	**0.456**
Own animal						
Yes	213			1		
No	103	-0.205	0.085	0.815	**0.815**	**0.690**
Week period and Period						
Weekend : Holiday	85:76	0.496	0.201	1.642	**1.642**	**1.106**

AIC = 1892.204; R-sq.(adj) = 0.174 Deviance explained = 22%; Theta= 3

Table 7: Generalised additive model for adult diary (13 – 60 years)

	n	Estimate	S.E	RR	95% C.I	
Intercept		2.855	0.264	17.371		
Daily activity						P=0.0479
No problem	979			1		
Some problems/Unable	109	-0.158	0.080	0.854	**0.730**	**0.999**
Anxiety						P=0.0036
Not anxious	987			1		
Moderate/Very anxious	101	-0.275	0.095	0.759	**0.631**	**0.914**
Age group						P<0.0001
13-17 years	66			1		
18-44 years	626	-0.705	0.111	0.494	**0.397**	**0.614**
45-60 years	401	-0.741	0.113	0.477	**0.382**	**0.595**
Education level						P<0.0001
Never/primary	72			1		
Vocational	119	0.271	0.114	1.311	**1.049**	**1.637**
Lower technical	55	0.085	0.124	1.089	0.854	1.389
Lower secondary	56	0.152	0.112	1.164	0.935	1.449
Higher technical	127	0.363	0.111	1.437	**1.155**	**1.788**
Upper Secondary	167	0.300	0.106	1.350	**1.098**	**1.660**
Higher non-University	343	0.470	0.101	1.600	**1.312**	**1.952**
University	148	0.411	0.108	1.508	**1.220**	**1.865**

Household size						P<0.0001
Size= 1 & 2	409			1		
Size= 3	249	0.097	0.052	1.102	0.994	1.221
Size= 4	301	0.181	0.053	1.199	**1.080**	**1.330**
Size= 5+	124	0.288	0.062	1.333	**1.181**	**1.506**
Week period						P=0.001
Week days	831			1		
Weekend	261	-0.515	0.157	0.597	**0.439**	**0.812**
Period						P<0.0001
Regular	839			1		
Holiday	253	-0.197	0.047	0.821	**0.749**	**0.901**
Own animal						P<0.0001
Yes	705			1		
No	382	-1.213	0.309	0.297	**0.162**	**0.545**
Health grading						P=0.8334
Grading	1050	0.001	0.003	1.001	0.995	1.007
Health grading and own animal				1		P=0.0001
No : Grading	382:1050	0.014	0.004	1.014	**1.007**	**1.021**
Age group and week period						P=0.0001
13-44 years	63			1		
18-44 years : Weekend	626:261	0.625	0.163	1.867	**1.356**	**2.572**
45-60 years : Weekend	401:261	0.709	0.171	2.032	**1.454**	**2.840**
Week period and own animal						P<0.0001
Weekend :No	261382	-0.370	0.093	0.691	**0.576**	**0.829**
Anxiety and Own animal						P=0.016
Mod/Very anxious :No	101:382	0.361	0.150	1.435	**1.068**	**1.927**

AIC= 7899.87; R-sq.(adj) = 0.131; Deviance explained = 15.6%

Table 8: Generalised additive model for elderly diary (60+ years)

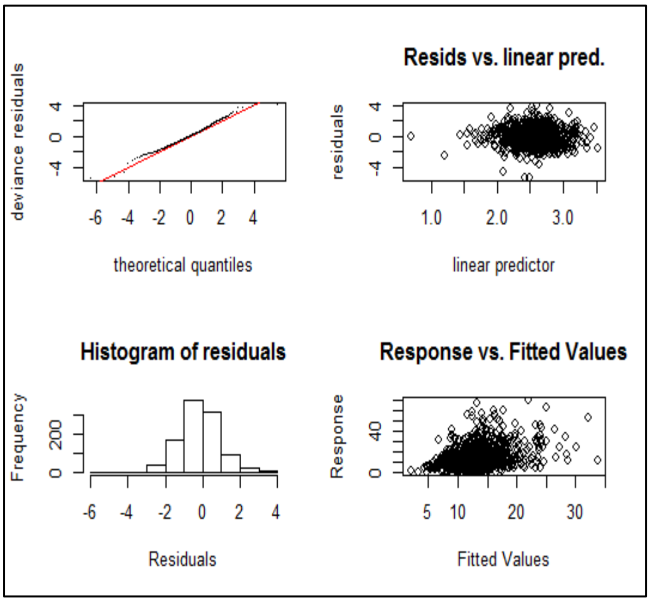
	n	Estimate	S.E	RR	95% C.I	
Intercept		2.185	0.164			
Self-care						p<0.0001
No problem	294			1		
Some problems/Unable	67	1.588	0.386	4.89	**2.297**	**10.432**
Pains						P=0.0175
No pains	154			1		
Moderate/Severe	207	-0.210	0.088	0.81	**0.682**	**0.964**
Week period						P=0.0166
Week day	287			1		
Weekend	76	0.230	0.096	1.26	**1.043**	**1.520**

Province						P=0.1978
Flamant Brabant	47			1		
Antwerp	100	-0.167	0.193	0.85	0.579	1.236
Limburg	65	0.119	0.216	1.13	0.737	1.722
West Flanders	64	-0.294	0.200	0.75	0.503	1.103
East Flanders	74	0.006	0.198	1.01	0.683	1.482
Smoking status						P=0.3218
Non-smoker	171			1		
Ex-smokers	133	-0.276	0.222	0.76	0.492	1.172
Smoker	44	-0.460	0.398	0.63	0.289	1.378
Self-care and smoking status						P=0.0003
Some problems/Unable : Ex-smoker	67:133	-0.662	0.226	0.52	**0.331**	**0.803**
Some problems/Unable : Smoker	67:44	-1.566	0.467	0.21	**0.084**	**0.522**
Province an smoking status						P=0.0152
Antwerp : Ex-smokers	100:133	0.279	0.272	1.32	0.775	2.254
Limburg : Ex-smokers	65:133	0.604	0.290	1.83	**1.036**	**3.232**
West Flanders : Ex-smokers	64:133	0.855	0.296	2.35	**1.317**	**4.200**
East Flanders : Ex-smokers	74:133	0.610	0.281	1.84	**1.060**	**3.195**
Antwerp : Smoker	100:44	-0.287	0.468	0.75	0.300	1.877
Limburg : Smoker	65:44	0.679	0.505	1.97	0.732	5.309
West Flanders : Smoker	64:44	0.852	0.476	2.34	0.922	5.959
East Flanders : Smoker	74:44	0.166	0.486	1.18	0.455	3.059
Self-care and Province						P=0.0014
Some problems/Unable : Antwerp	67:100	-0.464	0.424	0.63	0.274	1.443
Some problems/Unable : Limburg	67:65	-0.922	0.391	0.40	**0.185**	**0.855**
Some problems/Unable: West Flanders	67:64	-0.529	0.409	0.59	0.264	1.313
Some problems/Unable : East Flanders	67:74	-1.762	0.477	0.17	**0.067**	**0.438**

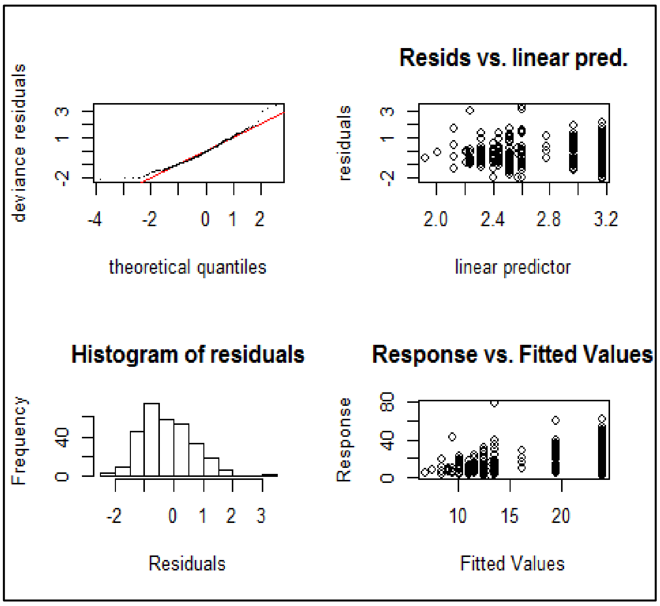
AIC=2104.71; R-sq.(adj) = 0.0945 Deviance explained = 22.3%

Appendix 2: Model diagnostic plots

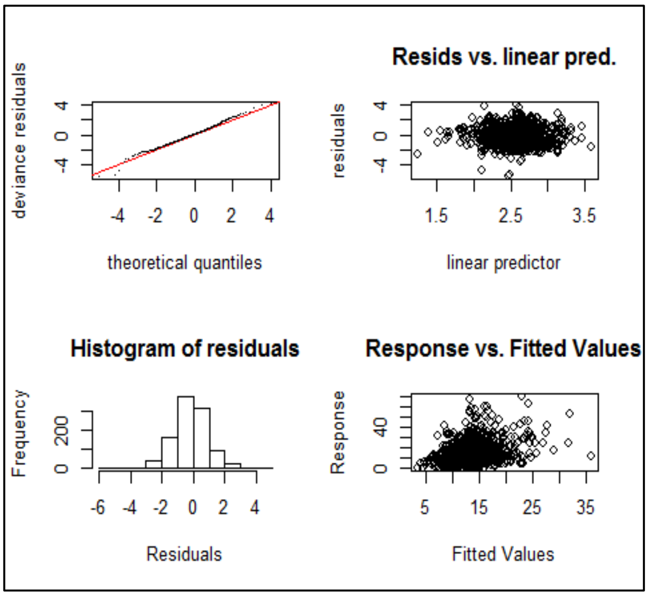
Combined



Children



Adult



Elderly

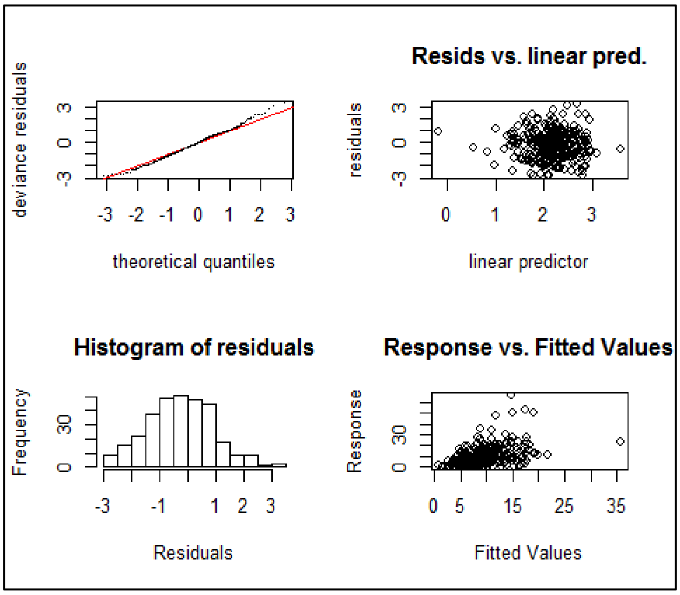


Figure 9a: Diagnostic plots for the GAM in the diary 13-60 years)

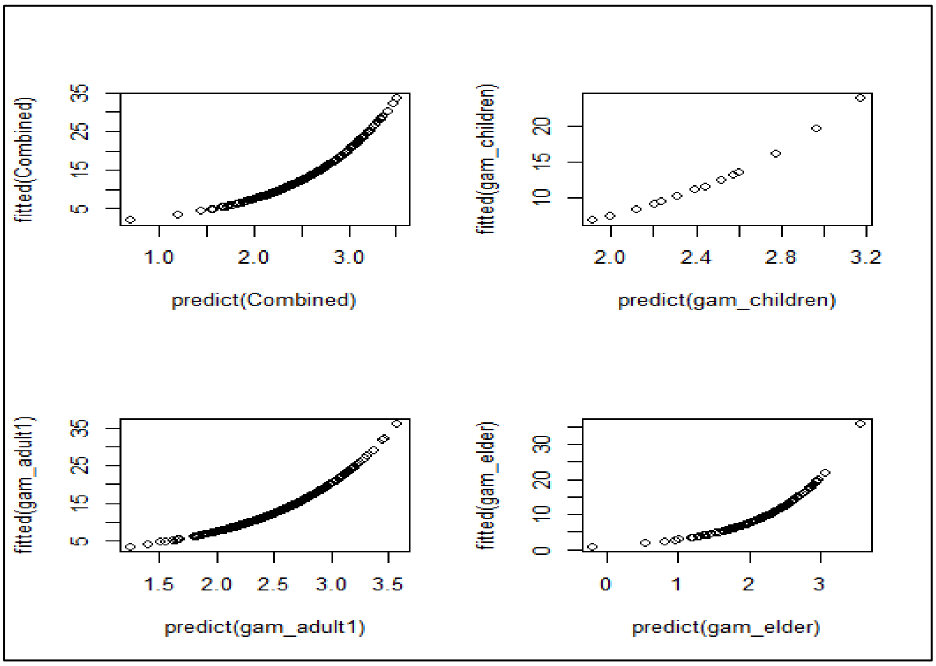


Figure 10b: Plots of fitted versus predicted

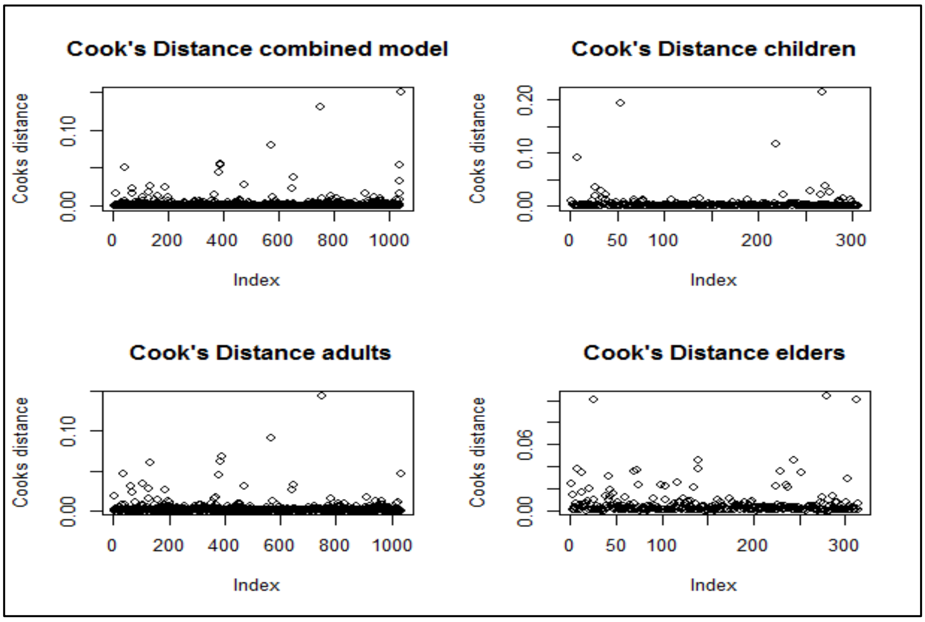


Figure 10c: Cook's distance

Appendix 3: Smoothers used in the models

Table 6: Different smoother tested in the model and their effect on the non-linear predictor (health grading).

Type of splines	Smoother	Intercept	Health grading Effect	Smoother effect	AIC
Basic splines					
Cubic spline	cr	yes	Not significant	Not significant	8282.34
Penalised cubic splines	cs	Yes	Not significant	Significant	8282.34
Thin plate regression splines					
Thin plate spline	tp	No	Significant	Significant	8291.8
	ts	Yes	Not significant	Not significant	8291.82
Duchon splines					
Duchon splines	ds	No	Significant	Significant	8292.1
P-splines					
P-splines	ps	Yes	Not significant	Significant	8291.14
	cp	yes	Not significant	Significant	8292.8
Random effects					
Random effects	re	yes	Not significant	Significant	8293.75

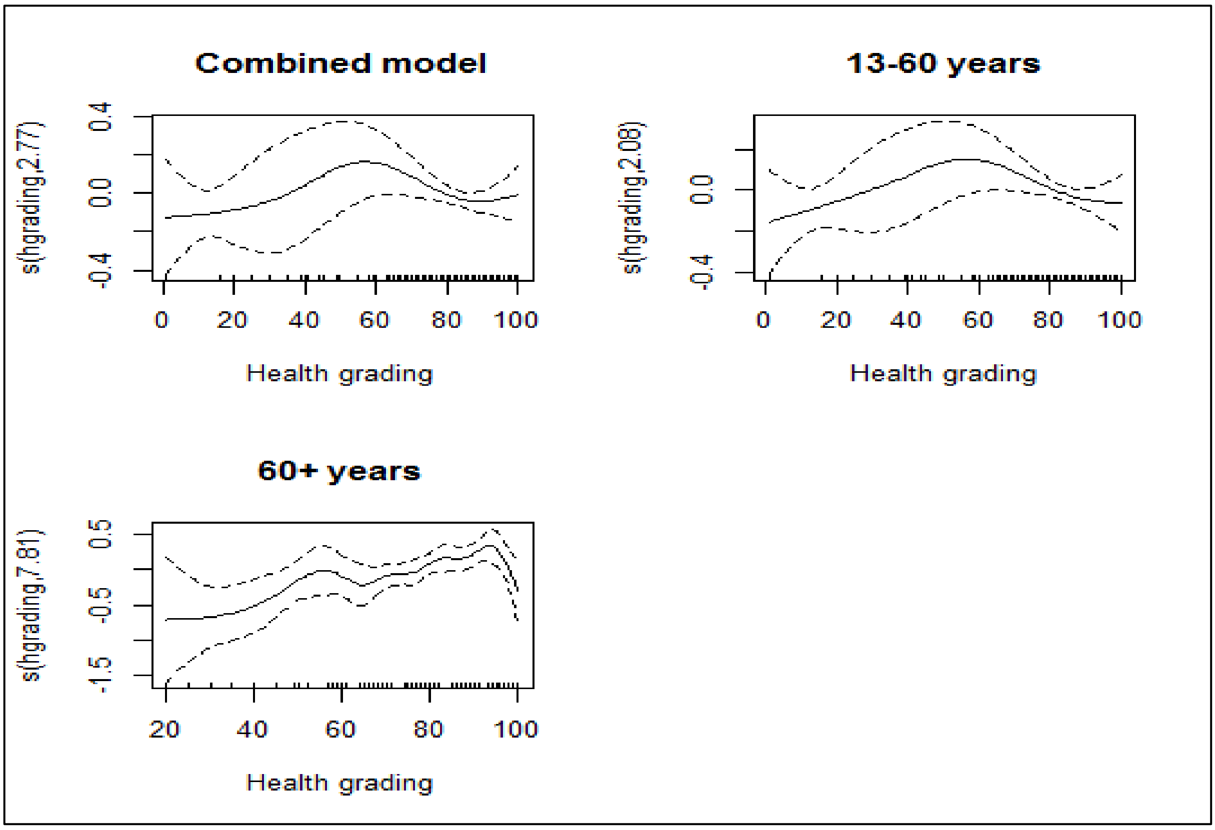


Figure 9: Plots of smoothen effect in combined and in different diary types.

Appendix 4: R code

#####Importing the data####

```
rm(list=ls(all=TRUE))
setwd("C://Users//Nicholas//Desktop//THESIS")
getwd()

contact = read.csv("contactnew1.csv", header = TRUE)
str(contact)
View(contact)
str(contact)
summary(contact)
dim(contact)
```

####R packages####

```
library(MASS)
library(car)
library(mgcv)
library(poweRlaw)
library(gsl)
library(numDeriv)
library(fitdistrplus)
library(phia)
```

######Declearing categorical variable######

```
contact[["period"]] <- factor((as.numeric(contact[["period"]])),
levels = 1 : 2 , labels = c("Regular", "Holiday"))

contact[["mobility"]] <- factor((as.numeric(contact[["mobility"]])),
levels = 1 : 3 , labels = c("No problem", "Some problems","Bedridden"))

contact[["weekperiod"]] <- factor((as.numeric(contact[["weekperiod"]])),
levels = 1 : 2 , labels = c("Weekday", "Weekend"))

contact[["selfcare"]] <- factor((as.numeric(contact[["selfcare"]])),
levels = 1 : 2 , labels = c("No problem", "Some problems/Unable"))

contact[["dailyactivity"]] <- factor((as.numeric(contact[["dailyactivity"]])),
levels = 1 : 2 , labels = c("No problem", "Some problems/Unable"))

contact[["pains"]] <- factor((as.numeric(contact[["pains"]])),
levels = 1 : 2 , labels = c("No pains", "Mod/Severe"))

contact[["anxiety"]] <- factor((as.numeric(contact[["anxiety"]])),
levels = 1 : 2 , labels = c("Not anxious", "Mod/Very anxious"))

contact[["agegroup"]] <- factor((as.numeric(contact[["agegroup"]])),
levels = 1 : 6 , labels = c("0-5", "6-12","13-17","18-44","45-64","65+"))

contact[["gender"]] <- factor((as.numeric(contact[["gender"]])),
levels = 1 : 2, labels = c("Males","Females"))

contact[["province"]] <- factor((as.numeric(contact[["province"]])), levels = 1:5,
```

```
labels = c("Flemish Brbant","Antwerp","Limburg","West Falnders","East Flanders"))

contact[["edulevel"]] <- factor((as.numeric(contact[["edulevel"]])),  levels = 2 : 9 ,
labels = c("Never/Primary","Voccational","Lower technical","Lower secondary",
"Higher technical", "Upper Secondary","Higher non-University","University"))

contact[["smoke"]] <- factor((as.numeric(contact[["smoke"]])),  levels = 1 : 3,
                        labels = c("Non-smokers","Ex-smokers","Smoker"))

contact[["Alcohol"]] <- factor((as.numeric(contact[["Alcohol"]])),
 levels = 1 : 2, labels = c("Yes","No"))

contact[["Alcohol_qty"]] <- factor((as.numeric(contact[["Alcohol_qty"]])),
levels = 1 : 5, labels = c("1-2 glases",">2 glases", "1/2 glases several times/week",
">2 glasses sevra time/week","Severat times/month"))

contact[["hh_size"]] <- factor((as.numeric(contact[["hh_size"]])),
levels = 2 : 5 , labels = c("<3","3","4","5+"))

contact[["ownanimal"]] <- factor((as.numeric(contact[["ownanimal"]])),
levels = 1 : 2, labels = c("Yes","No"))

contact[["touchanimal"]] <- factor((as.numeric(contact[["touchanimal"]])),
levels = 1 : 2, labels = c("Yes","No"))

contact[["diary"]] <- factor((as.numeric(contact[["diary"]])),  levels = 1 : 3 ,
                        labels = c("0-12 years","12-60 years","> 60 years"))

######Plotting histogram of number of social contacts######
par(mfrow=c(1,2))
hist(contact$logcnt, prob=FALSE,col="red",breaks=20,xlim=c(0,6), ylim=c(0,350),
    ylab="Frequency", xlab ="Log(number of contacts)",
    main ="Degree distribution plot for the number of contacts")

hist(contact$hgrading, ylab="Frequency", xlab ="Number of contacts",
    main ="Health grading & Number of contacts")


############ Number of contacts versus predictors ############
logcnt=log(contact$nbcnts+1)

par(mfrow=c(2,3))
plot(contact$logcnt ~ period, ylab= "Log(number of contacts)", xlab="Period",
main = "Nbr of contacts holiday/ school periods", data=contact)

plot(contact$logcnt ~ weekperiod, ylab= "Log(number of contacts)", xlab="Week period",
main = "Nbr of contacts Week days/weekend", data=contact)

plot(contact$logcnt ~ diary, ylab= "Log(number of contacts)", xlab="Diary type",
 main = "Diary types and Nbr of social contacts", data=contact)

plot(contact$logcnt ~ mobility, ylab= "Log(number of contacts)", xlab="Mobility",
    main = "Mobidity and Nbr of contacts", data=contact)
```

```
plot(contact$logcnt~ selfcare, ylab= "Log(number of contacts)", xlab="Self care",
    main = "Self care  and Nbr of contacts", data=contact)

plot(contact$logcnt ~ dailyactivity, ylab= "Log(number of contacts)", xlab="Daily activities",
    main = "Daily activities and Nbr of contacts", data=contact)

plot(contact$logcnt ~ pains, ylab= "Log(number of contacts)", xlab="Pains",
    main = "Pains and Nbr of contacts", data=contact)

plot(contact$logcnt ~ anxiety, ylab= "Log(number of contacts)", xlab="Anxiety",
    main = "Anxiety and number of contacts", data=contact)

plot(contact$logcnt ~ hgrading, ylab= "Log(number of contacts)", xlab="Health grading",
    main = "Health grading and Nbr of contacts", data=contact)

plot(contact$logcnt ~ gender, ylab= "Log(number of contacts)", xlab="Gender",
    main = "Gender and Nbr of contacts", data=contact)

plot(contact$logcnt~ agegroup, ylab= "Log(number of contacts)", xlab="Age group",
    main = "Age group and Nbrof contacts", data=contact)

plot(contact$logcnt ~ province, ylab= "Log(number of contacts)", xlab="Province",
    main = "Province and Nbr of contacts", data=contact)

plot(contact$logcnt ~ smoke, ylab= "Log(number of contacts)", xlab="Smoking status",
    main = "Smoking status and Nbr of contacts", data=contact)

plot(contact$logcnt ~ hh_size, ylab= "Log(number of contacts)", xlab="Household size",
    main = "Household size and Nbr of contacts", data=contact)

plot(contact$logcnt ~ edulevel, ylab= "Log(number of contacts)", xlab="Education level",
    main = "Education level and Nbr of contacts", data=contact)

plot(contact$logcnt ~ ownanimal, ylab= "Log(number of contacts)", xlab="Owning animal",
    main = "Owning animals and Nbr of contacts", data=contact)

plot(contact$logcnt ~ touchanimal, ylab= "Log(number of contacts)", xlab="Touching animals",
    main = "Touching animals and Nbr of contacts", data=contact)

###ASSOCIATION PLOTS
plot(contact$logcnt~Alcohol, ylab= "Log(number of contacts)", xlab="Week period",
    main = "Nbr of contacts Week days/weekend", data=contact)
```

Fitting generalized linear model for count data######
####Stratification of data into different diary types####

```
(contact_children <- contact[1:317, ])
(contact_adult <- contact[318:1410, ])
(contact_elder <- contact[1411:1774, ])
```

Negative binomial model#####

```
nb_combined <- glm.nb(nbcnts ~
dailyactivity+agegroup+edulevel+hh_size+weekperiod+period+ownanimal+agegroup*weekperiod
```

```
              +hh_size*period +hh_size*weekperiod +weekperiod*ownanimal.
              data = contact,  weights=weights)
summary(nb_combined)
confint(nb_combined, level = 0.95)
anova(nb_combined, test = "Chisq")
```

GAM combined

```
Combined= gam(nbcnts~dailyactivity+s(hgrading, bs = "cr")+agegroup +edulevel+hh_size
+weekperiod+period+ownanimal +hgrading*ownanimal+agegourp*weekperiod
+weekperiod*ownanimal, family= negbin(3),weights=weights, data=contact)
summary(Combined)
AIC(Combined)
anova(Combined, test = "Chisq")
```

####STRATIFIED ANALYSIS#####

####GAM children ####

```
gam_children = gam(nbcnts~dailyactivity+weekperiod+period+ownanimal+period
            +weekperiod*period, family=negbin(3), weights=weights,data=contact_children)
summary(gam_children)
AIC(gam_children)
anova(gam_children, test = "Chisq")
```

####GAM adult#####

```
gam_adult = gam(nbcnts~s(hgrading, bs = "cr")+dailyactivity+anxiety+agegroup+edulevel+hh_size
           +weekperiod+period+ownanimal+hgrading*ownanimal+agegroup*weekperiod
           +weekperiod*ownanimal+anxiety*ownanimal, family=negbin(3), weights=weights,
data=contact_adult)
summary(gam_adult)
AIC(gam_adult)
plot(gam_adult)
anova(gam_adult, test = "Chisq")
```

####GAM elder#####

```
gam_elder = gam(nbcnts~selfcare+s(hgrading,bs="cr")+pains+weekperiod+province+smoke
          +smoke*selfcare+smoke*province+province*selfcare, family=negbin(3),
           weights=weights, data=contact_elder)
summary(gam_elder)
AIC(gam_elder)
plot(gam_elder)
anova(gam_elder, test = "Chisq")
```

#####VISUALIZATION PLOTS FOR MAIN EFFECT AND INTERACTIONS###

###Main effect and interactions combined####

```
gam_adult.means <- interactionMeans(combined)
plot(combined.means, atx="dailyactivity")
plot(combined.means, atx="agegroup")
plot(combined.means, atx="edulevel")
plot(combined.means, atx="hh_size")
plot(gam_adult.means, atx="weekperiod")
plot(combined.means, atx="period")
plot(combined.means, atx="ownanimal")
plot(combined.means, atx="agegroup", traces="weekperiod")
plot(combined.means, atx="ownanimal", traces="weekperiod")
```

###Main effect and interactions gam_children####

```
gam_children.means <- interactionMeans(gam_children)
plot(gam_children.means, atx="dailyactivity")
plot(gam_children.means, atx="ownanimal")
plot(gam_children.means, atx="weekperiod")
plot(gam_children.means, atx="period")
plot(gamchildren.means, atx="period", traces="weekperiod")
```

###Main effect and interactions gam_adult####

```
gam_adult.means <- interactionMeans(gam_adult)
plot(gam_adult.means, atx="dailyactivity")
plot(gam_adult.means, atx="anxiety")
plot(gam_adult.means, atx="ownanimal")
plot(gam_adult.means, atx="weekperiod")
plot(gam_adult.means, atx="period")
plot(gam_adult.means, atx="agegroup")
plot(gam_adult.means, atx="edulevel")
plot(gam_adult.means, atx="hh_size")
plot(gam_adult.means, atx="agegroup", traces="weekperiod")
plot(gam_adult.means, atx="ownanimal", traces="weekperiod","anxiety")
```

###Main effect and interactions gam_elder####

```
gam_children.means <- interactionMeans(gam_elder)
par(mfrow=c(1,2))
plot(gam_elder.means, atx="selfcare")
plot(gam_elder.means, atx="pains")
plot(gam_elder.means, atx="weekperiod")
plot(gam_elder.means, atx="province")
plot(gam_elder.means, atx="smoke")
plot(gam_elder.means, atx="smoke", traces=c("selfcare","province")
plot(gam_elder.means, atx="province", traces="selfcare")
```

###Checking for multicolinearity###

```
vif(glm(nbcnts ~ dailyactivity+hgrading+agegroup +edulevel+hh_size
    +weekperiod+period+ownanimal,family= negbin(3),weights=weights, data=contact))

vif(glm(nbcnts ~ dailyactivity+weekperiod+period+ownanimal+period
      +weekperiod*period, family=negbin(3), weights=weights,data=contact_children))
vif(glm(nbcnts ~ hgrading+dailyactivity+anxiety+agegroup+edulevel+hh_size
  +weekperiod+period+ownanimal,family=negbin(3), weights=weights, data=contact_adult))
vif(glm(nbcnts ~ selfcare+hgrading+pains+weekperiod+province+smoke,
          family=negbin(3), weights=weights, data=contact_elder))
```

#####DIAGNOSTIC PLOTS #####

```
gam.check(Combined)
gam.check(gam_children)
gam.check(gam_adult)
gam.check(gam_elder)
```

###predicted vs fitted plots###

```
par(mfrow=c(2,2))
fitted(Combined)
```

```
plot(predict(Combined),fitted(Combined))

fitted(gam_children)
plot(predict(gam_children),fitted(gam_children))

fitted(gam_adul1)
plot(predict(gam_adult1),fitted(gam_adult))

fitted(gam_elder)
plot(predict(gam_elder),fitted(gam_elder))
```

######Outliers identification using Cooks distance#####

```
par(mfrow=c(2,2))
cd_combined <- cooks.distance(combined)
plot(cd_combined,ylab="Cooks distance",main="Cook's Distance combined model", abline(h=1))
contact[which(cooks.distance(gamfinal4)>1),]

cd_children <- cooks.distance(gam_children)
plot(cd_children,ylab="Cooks distance",main="Cook's Distance children", abline(h=1))
contact[which(cooks.distance(gam_children)>1),]

cd_adult <- cooks.distance(gam_adult)
plot(cd_adult,ylab="Cooks distance",main=" Cook's Distance adults", abline(h=1))
contact[which(cooks.distance(gam_adult1)>1),]

cd_elder <- cooks.distance(gam_elder)
plot(cd_elder,ylab="Cooks distance",main="Cook's Distance elders", abline(h=1))
contact[which(cooks.distance(gam_elder)>1),]
compareCoefs(gam_elder, update(gam_elder, subset=-c(1039))
```

###Ploting of smoothing effect)

```
par(mfrow=c(2,2))
plot(combined, xlab="Health grading", main="Combined model")
plot(gam_adult,xlab="Health grading",main="13-60 years" )
plot(gam_elder,xlab="Health grading",main="60+ years")
plot(gam_children, xlab="Health grading",main="0-12 years")
```

Fitting a discrete power law##

```
m  <- contact$new[which(contact$hgrading>0), ]
set.seed(1)
m = ceiling(rlnorm(1000, 1))
occur = as.vector(table(m))
occur = occur/sum(occur)
p = occur/sum(occur)
y = rev(cumsum(rev(p)))
x = as.numeric(names(table(m)))
plot(x, y, log="xy", ylab="Frequency", xlab="Health grading",
    type="l", main="Health grading", lwd=2)
```

######Verifying if the frequency of health indictors follows a power law ####
Fitting a discrete power law on health grading (same code for other health indicators###

```
m1 = displ$new(contact$hgrading)
```

```
m1$setPars(estimate_pars(m1))
m1$getXmin()
m1$getPars()

####fitting a lognormal model####
m2 = dislnorm$new(contact$hgrading)
m2$setPars(estimate_pars(m2))

plot(m2, ylab="CDF", xlab="Health grading", lwd=2) ##plotting power law and lognormal model
lines((m1), lwd=2)
lines(m2, col=2, lty=2,lwd=2)

comp = compare_distributions(m1, m2)
comp$p_two_sided
```

######code for weights calculation####

```
rm(list=ls(all=TRUE))
setwd("C://Users//Nicholas//Desktop//THESIS")
getwd()
```

###Importing data##

```
contact = read.csv("contact.csv", header = TRUE)
str(contact)
View(contact)
str(contact)
summary(contact)
dim(contact)
```

#####Calculating the weights for household size ##########

```
age1<-c(1724,  26443, 163259, 224769, 158086)
age2<-c(1178,  24742, 94847, 243304, 225630)
age3<-c(1261,  27758, 103465, 242259, 256139)
age4<-c(7430, 40866, 125694, 223525, 221667)
age5<-c(60967, 118726, 163650, 168142, 129636)
age6<-c(110168, 192794, 182301, 112135, 57707)
age7<-c(104327, 138143, 193544, 196422, 86339)
age8<-c(95167, 108320, 171765, 253006, 146232)
age9<-c(96169, 119803, 181743, 254277, 162998)
age10<-c(97330, 166079, 197750, 196205, 115096)
age11<-c(99192, 249739, 186425, 111855, 56279)
age12<-c(103068, 331525, 141144, 53175, 26915)
age13<-c(85979, 296975, 76096, 21247, 13835)
age14<-c(101268, 311708, 56808, 13337, 10373)
age15<-c(610978, 509148, 124023, 35248, 27415)

agematrix<-as.matrix(rbind(age1,age2,age3,age4, age5,age6,age7,
        age8,age9,age10,age11,age12,age13,age14,age15))
dimnames(agematrix) <- list(NULL, NULL)
Pop_size<-c(574281,589701,630882,619182,641121,655105,718775,774490,
          814990,772460,703490,655827,494132,493494,1306812)
```

 # Population density= Joint cells/total population size

```
for(j in 1:15){
```

```
  for(k in 1:5){
    contact$pfraction[contact$ageweight==j & contact$hh_size==k] <-agematrix[j,k]/sum(Pop_size)
  }
}
agehsize1<- c(0, 2, 50, 56, 39, 1)
agehsize2<- c(0, 1, 19, 64, 40, 1)
agehsize3<- c(0, 3, 14, 36, 19, 0)
agehsize4<- c(0, 5, 4, 30, 24, 1)
agehsize5<- c(2, 15, 26, 16, 14, 1)
agehsize6<- c(10, 41, 33, 25, 13, 3)
agehsize7<- c(11, 38, 30, 34, 9, 2)
agehsize8<- c(17, 22, 37, 43, 12, 4)
agehsize9<- c(15, 24, 31, 55, 15, 1)
agehsize10<- c(13, 34, 40, 45, 19, 0)
agehsize11<- c(17, 72, 32, 28, 10, 1)
agehsize12<- c(12, 44, 9, 5, 4, 0)
agehsize13<- c(1, 9, 2, 1, 0, 70)
agehsize14<- c(0, 2, 1, 0, 0, 70)
agehsize15<- c(0, 0, 0, 1, 0, 221)

agehsizematrix<-as.matrix(rbind(agehsize1,agehsize2,agehsize3,agehsize4,agehsize5,agehsize6,
  agehsize7,agehsize8,agehsize9,agehsize10,agehsize11,
agehsize12,agehsize13,agehsize14,agehsize15))

dimnames(agehsizematrix) <- list(NULL, NULL)
sample_size<-c(148,125,72,64,74,125,124,135,141,151,160,74,83,73,222)

# Sample density= Joint cells/total sample size
for(s in 1:15){
  for(t in 1:5){
contact$sfraction[contact$ageweight==s & contact$hh_size==t] <-
agehsizematrix[s,t]/sum(sample_size)
  }
}

# Assigning crude weights=pop density/sample density except Hsize=6
attach(contact)
contact$weights<-contact$pfraction/contact$sfraction
detach(contact)

# Assigning crude weights=pop density/sample density for Hsize=6
mar_weight<-(Pop_size/sum(Pop_size))/(sample_size/sum(sample_size))
for(j in 1:15){
  contact$weights[contact$ageweight==j
          & contact$Hsize==6] <-mar_weight[j]
}
contact$weightcount=contact$weight*contact$nbcnts
View(contact)
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Modelling health and socio-demographic indicators of social contacts in the Flranders, Belgium

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Tendongfor, Nicholas

Datum: **1/09/2015**