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

2012•2013
FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

Human-animal interactions: Survey of contact behavior relevant for the spread of infectious diseases

Promotor :
Prof. dr. Niel HENS
dr. Nele GOEYVAERTS
dr. Kim VAN KERCKHOVE

Yimer Wasihun Kifle

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

2012•2013
FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

Human-animal interactions: Survey of contact behavior relevant for the spread of infectious diseases

Promotor :
Prof. dr. Niel HENS
dr. Nele GOEYVAERTS
dr. Kim VAN KERCKHOVE

Yimer Wasihun Kifle
Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

universiteit hasselt | Maastricht University

Certification

This is to certify that this report was written by *Yimer Wasihun Kifle* under my Supervision.

Yimer Wasihun Kifle
Student

Signature

Date

Prof. dr. Niel Hens	Dr. Nele Goeyvaerts	Mrs. Kim Van Kerckhove
Supervisor	Co-supervisor	Co-supervisor
------------------------	------------------------	------------------------
Signature	Signature	Signature
------------------------	------------------------	------------------------
Date	Date	Date

Acknowledgment

First and above all, I praise for the almighty God, for providing me this opportunity and compromise me the capability to proceed successfully. The completion of this thesis could not have been possible to be appearing in its current form without the support and guidance of several people and organizations. Therefore, I would like to offer my sincere gratitude to all of them. First, my appreciation goes to the Flemish Interuniversity Council (Vlaamse Interuniversitare Raad (VLIR)) university cooperation development program for sponsoring me financially. Your help makes me to focus on my thesis and the whole study for the last two years without any extra difficulties in life. Second, my extraordinary thanks go to my supervisor Prof. Dr. Niel Hens, Dr. Nele Goeyvaerts and Kim Van Kerckhove for your establishment of conducive environment for my project with continuous support starting from initial contacts through ongoing discussions and valuable comments up to finalizing this paper. Thirdly, I would like to acknowledge all CenStat staffs and professors for passing on your knowledge, thanks classmates for sharing ideas.

Finally, I would like to express my deepest gratitude to my parents. Mom, Dad, sisters and brothers, I know that you are always praying for my success. Thank you Mom and Dad again for your lifelong love.

Yimer Wasihun Kifle

Contents

List of Tables

List of Figures

Abstract

Background: Approximately 75 percent of recently emerging infectious diseases that affect humans are diseases of animal origin, and approximately 60 percent of all human pathogens are zoonotic. Contact between humans and pets, livestock, poultry and other animals could enable transmission of pathogens from animals to humans, which is the focus of this study. The aim of this study is to evaluate factors that could affect the chance of having and touching animals and the number of contacts between people.

Methods: Diary based design, social contact survey conducted for collecting information about participants by random digit dialing on fixed and mobile telephone lines. The relative odds of owning and touching animals between groups of a factor was analyzed using logistic regression and the number of contacts between people was analyzed using weighted negative binomial and Hierarchical zero-inflated negative binomial regression.

Results: More than, halve of participants own or touch animals. The mean and variance for the number of contacts were 13.47 and 116.78 for individual survey data, respectively and it was 15.66 and 125.89 for household survey data, implying over dispersed data. The relative odds of owning animals differ substantially through the age, household size and educational level of a participant whereas the relative odds of touching animals differ through age, educational level and owning animal status of a participant. Finally, the number of contacts between people was affected by owning animal status of a participant depending on the weekdays or weekend days.

Conclusion: Participants, who own animals, have the highest chance to touch animals and have a larger number of contacts with people as compared to those who do not own animals such that pathogens will pass and cause illness easily from animals to humans, this may enhance the spread of infectious diseases.

Keywords: *Owning animals; Touching animals; Logistic regression; Number of contacts; Negative binomial regression; Hierarchical zero-inflated negative binomial regression*

1 Introduction

Mathematical models of infectious disease transmission have become indispensable tools for understanding epidemic processes and for providing policy makers with an evidence based decisions when empirical data are limited. The achievement of mathematical models in informing important decisions to care for human and animal health has been demonstrated for many diseases including pandemic influenza, severe acute respiratory syndrome, foot and mouth disease, and new variant Creutzfeldt–Jakob disease (Grassly 2008).

Infectious diseases are illness in humans and animals caused by microbial pathogens — such as bacteria, viruses, fungi or parasites. Many pathogens live inside and on the surface of our body. They are normally harmless or even helpful, but some pathogens under certain conditions may cause disease. Transmission of pathogens can occur in various ways including physical contact, contaminated food and water, body fluids, objects, airborne inhalation, or through vector organisms (Ryan 2004). In previous studies, infections directly transmitted from person to person by the respiratory route have been of special interest for modeling because of their ability to spread quickly and affect large numbers of people. The social contact survey data were analyzed for estimating parameters in a transmission model of contact between person to person (Horby et al. 2011, Hens et al. 2009, Mossong et al. 2008 and Ogunjimi et al. 2009). However, a number of infectious diseases can be transmitted from animals to humans through human — animal interactions. The relationships between people and animals, the ability of microorganisms to cross species barriers, and many other factors that affect these interactions combine to facilitate the emergence and transmission of zoonotic pathogens (Morse 2004).

Zoonoses are type of infectious diseases that can be transmitted between humans and animals. The Centers for Disease Control and Prevention (2011-2012) report that approximately 75 percent of recently emerging infectious diseases that affect humans are diseases of animal origin, and approximately 60 percent of all human pathogens are zoonotic. Most emerging infections that affect humans are zoonotic in origin (Taylor et al. 2001) and there is evidence about organisms that are able to infect both domestic animals and wildlife species are the most likely to emerge as zoonoses (Cleaveland et al. 2001). Emerging diseases in both human beings and animals with which they interact have potential implications that may extend beyond local or national borders to a global arena (Jebara 2004.

Since the human immunodeficiency virus (HIV) commonly transmitted from person to person, acquired immune-deficiency syndrome (AIDS) is not considered as zoonosis. However, it is thought to have originated as zoonosis. Sharp et al. (2013) derived molecular phylogenies of the two distinct types of human AIDS viruses — such as HIV-1 and HIV-2. The paper indicated that HIV-1 most likely arose as a consequence of simian immunodeficiency virus (SIV) transmission from chimpanzees to humans, with the most likely chimpanzee source being West African Pan troglodytes troglodytes. HIV-2 appears to have resulted from cross-transmission of SIV between sooty mangabeys and humans in West Africa.

Contact between humans and livestock, domestic pets and poultry is also common to occur in daily life activities, which increases the likelihood of disease transmission. Yu et al. (2013) found that poultry workers had a considerable risk of infection with H9 subtypes of avian influenza virus in China. In addition, contact with pets has numerous positive benefits, including opportunities for education and entertainment. However, many pet owners and people in the process of choosing a pet have a higher risk of infection for many diseases. It is often hard to determine patterns of human-animal interactions, which makes modeling the spread of zoonoses difficult. Having access to research-based information about risk factors for contact between humans and animals can help the researcher to deal transmission probability of zoonoses. In this paper, consideration has been given to zoonotic infectious diseases from animals to humans, however, reverse zoonoses also occur in which disease is transmitted from human reservoir to animals.

This study aimed at assessing essential risk factors for owning, touching status of participants with animals and the number of contacts between people which could enable for the transmission of pathogens. In addition, the chance of touching animals given that someone owned animals will be investigated and the effect of owning animals on the number of contacts had a focus on this paper.

2 Survey Methods and Data

Survey methods used in this study are described under subsection 2.1 including method of data collection, study area and how the data was collected. There are two different data sets. The first data is collected by recruiting only one individual participant per household; it is referred as individual survey data in different sections of this paper. The second data set is referred as household data in which not only one person per household was recruited to take part but also other members of the household were participated, unlike individual survey data.

2.1 Survey Methods

A social contact survey, which was diary based design, was conducted in the Flemish geographic region of Belgium from September 2010 until February 2011. Participants were recruited by random digit dialing on fixed and mobile telephone lines and sampling was performed in order to achieve a representative geographical spread. All participants were asked to fill in a paper diary recording their contacts during one randomly assigned day without changing their usual behavior. No physical samples were collected as part of this study and the ethical committee of the Antwerp University Hospital approved the study protocol. A verbal consent was given prior to participation during the recruitment over the phone. People who agreed verbally to participate were then sent a written questionnaire and diary. The participants received and sent their questionnaires back by postal services (Willem et al. 2012).

They were able to refuse participation even after verbal agreement by not filling in the questionnaire and diary, and/or by not sending it back. The first page of the questionnaire explained that their answers would be used anonymously for scientific research purposes at the research centers. Thus, the fact that they filled in the questionnaire and diary and chose to send it in functions as a written consent. It has been obtained similar verbal consent with implicit written confirmation from the next of kin, caretakers or guardians on behalf of dependent participants (e.g. Children).

What do we mean by a contact in this social contact survey? It defines two types of contacts. The first type of contact defined when a participant has spoken in his/her presence in less than three meters with someone, however, a call or contact via the internet does not consider as a contact here. The second one is considered as contact when a participant had a physical contact with someone. If a participant has touched someone (like give a hand, give a kiss, hug, and casual contacts during sports etc.) even if it was without communication, it is also considered as a contact. These contacts were meant to be relevant for sensitive major contact types enabling pathogen transmission by close contact (Beutels et al. 2006).

Three types of diaries were used, adapted to the ages of the participants such as for children up to 12 years, peoples in between 12 and 60 years old and people older than 60 years. For example, the diaries for children (0–12 years) designed to be filled by proxy, and included school contacts, which included separate instructions for schoolteachers, whereas the proxy could also fill those for elderly. The diaries sent and collected by mail. Participants reminded by phone to fill in the diary one day in advance and followed up the day after (Willem et al. 2012). Data were single entered in a computer database and checked manually. Based on the type of diary, participants requested to fill the assigned date, their personal information (like age, gender, educational level, province, household size) and owning and touching status of participants with pets, livestock, poultry and animals in general.

2.2 Data

Both individual and household survey data summarized under table 1 by the number and proportion of participants under each category of all variables included in the study. All the variables mentioned under table 1 are included in both data sets. Depending on stages of education in Belgium, which is the same in all communities the maximum education level of a participant categorized as no education, primary school, secondary education, higher education and missing. In addition, educational level of a child with (0-12) years of age was replaced by the educational level of his/her mother. Participant's household size also categorized as 1, 2, 3, 4, 5 and more, and missing for further analysis. Depending on the assigned day, weekday indicator variable is called weekday if a day was Monday, Tuesday, Wednesday, Thursday or Friday, if not it is called weekend day.

The public holidays or event days in Flanders, Belgium were on 27 September 2010; 31 October 2010; (01, 11 and 28) November 2010; (05, 12, 19, 25 and 31) December 2010 and 01 January 2011 during the survey period. Based on these dates, the assigned date categorized as holiday (a date from the list of public holidays) or regular day (a date not from the list of public holidays). The dot on table 1 indicates the corresponding category of a variable was not included by the study design for the aforementioned data sets. For example, Brussels was not included in the individual survey data whereas it was included in household survey data. The owning and touching status are implying whether a participant own and touch animal or not, respectively. Finally, the number of contacts per participant was included as a variable in both data sets. Under individual survey data, there were 1768 individual participants and 23,838 contact people in total. Under household survey data, there were 1344 participants in 342 households and 21052 people make contact with participants.

Table 1: The number and proportion of participants under different categories of variables from individual and household survey data

			Individual-Survey Data		Household-Survey Data	
Variables	Categories	Label	Number	(Proportion)	Number	(Proportion)
Age	(0-5) years	1	174	(9.84%)	200	(14.88%)
	(6-11) years	2	128	(7.24%)	299	(22.25%)
	(12-17) years	3	80	(4.52%)	170	(12.65%)
	(18-44) years	4	626	(35.41%)	557	(41.44%)
	(45-64) years	5	468	(26.47%)	118	(8.78%)
	65+ years	6	292	(16.52%)	.	
Gender	Female	1	947	(53.56%)	676	(50.30%)
	Male	2	821	(46.44%)	668	(49.70%)
Educational level	Higher education	1	791	(44.74%)	750	(55.80%)
	Secondary education	2	809	(45.76%)	513	(38.17%)
	Primary school	3	141	(7.98%)	62	(4.61%)
	No education	4	10	(0.57%)	5	(0.37%)
	Missing*	5	17	(0.96%)	14	(1.04%)
Province	Antwerp	1	492	(27.83%)	353	(26.26%)
	Limburg	2	265	(14.99%)	162	(12.05%)
	East-Vlaanderen	3	410	(23.19%)	266	(19.79%)
	Flemish-Brabant	4	259	(14.65%)	187	(13.91%)
	West-Vlaanderen	5	328	(18.55%)	224	(16.67%)
	Missing	6	14	(0.79%)	.	
	Brussels	7	.		152	(11.31%)
Household size	1	1	98	(5.54 %)	.	
	2	2	312	(17.65%)	25	(1.86%)
	3	3	328	(18.55%)	224	(16.67%)
	4	4	439	(24.83%)	683	(50.82%)
	5+	5	218	(12.33%)	412	(30.65%)
	Missing	6	373	(21.10%)	.	
Owning status	Yes	1	1050	(59.39%)	916	(68.15%)
	No	2	706	(39.93%)	420	(31.25%)
	Missing	3	12	(0.68%)	8	(0.60%)
Touching status	Yes	1	874	(49.43%)	712	(52.98%)
	No	2	850	(48.08%)	607	(45.16%)
	Missing	3	44	(2.49%)	25	(1.86%)
Weekday indicator	Weekend days	0	424	(23.98%)	387	(28.79%)
	Weekdays	1	1344	(76.02%)	957	(71.21%)
Holiday indicator	Regular days	0	1656	(93.67%)	1330	(98.96%)
	Holidays	1	112	(6.33%)	14	(1.04%)

- The educational level of a child from 0 up to 12 years old filled by the educational level of his/her mother.

2 Statistical Methodology

Referring to table 1, 0.68 and 2.49 percent of total participants were not replying about their owning and touching status, respectively from individual survey data. Since owning and touching status of participants with animals are binary responses, which can be yes/no after excluding missing category, there will be a descriptive analysis using mosaic plot for contingency tables, chi-square test of independence and logistic regression under this section. Keep in mind that deleting missing values is not a solution for missing data problem, however, the effect of deleting few missing values here assumed to be minimal. Since the number of contacts per participant from the individual survey data is count data type and independent then Poisson, negative binomial and zero-inflated negative binomial regression will be fitted and compared (Mossong et al. 2008, Hens et al. 2009). Finally, the number of contacts from household survey data is count data type and correlated since it is taken from the same household, the extension of Poisson and negative binomial regression by including random effects called hierarchical model will be fitted and compared.

3.1 Modeling Touching and Owning Status

From social contact survey data, participants were asked to fill their status of owning and touching animals during the assigned day. Since participants who owned and touched animals expected to have a higher chance to receive microbial pathogens causing different infectious disease, owning and touching status of participants were modeled using two different logistic regressions including different risk factors like number of contacts, age, gender, educational level, household size, province of participants and etc. for both models.The essential step under modeling was the decision of which risk factor has to be kept in the logistic regression model has always been challenging. It is helpful first to study the effect on the outcome variable of each predictor by itself using graphical presentation and contingency tables. Many model selection procedures exist, no one of which is always best (Agresti 2002). Other criteria besides significance tests can help select a good model in terms of estimating quantities of interest. The best known is Akaike information criterion (AIC) which judges a model of how close it fitted values tend to be true values, in terms of a certain expected value. Akaike showed that a model that minimizes AIC is the optimal model that tends to provide better estimates of certain characteristics of the true model, such as cell probabilities. Therefore, AIC will be used for variable selection procedure in this paper.

AIC= -2(maximized log likelihood) + 2* (number of parameters in the model)

Let as assume here, owning status (Y_i) is 1 if a participant owns an animal, otherwise 0 and touching statues (Y_i) is 1 if a participant touches an animal, otherwise 0. In logistic regression the linear predictor and probability of success are related through $logit$ link function. Given that π (X_i) = P (Y_i = 1|X = x_i) is the probability of owning or touching animals with respective to the type of response used given that predictor X=x_i in the model.

$$\mathrm{logit}(\pi(X_i)) = \beta_0 * X_0 + \beta_1 * X_1 + \beta_2 * X_2 + \cdots + \beta_k * X_k \quad (1)$$

Where X_0 = 1, Xi = (X_1, X_2, ...,X_k) are the explanatory variables and β_0=intercept and β_1 , β_2, ..., β_k are regression parameters for the effect of possible risk factors like age, gender, education level etc.

3.2 Modeling Number of Contacts from Individual Survey Data

The response of interest under this section, i.e. the participant's number of contacts within a day, is a count data type and a Poisson distribution seems a plausible assumption. However, the Poisson distribution assumes equality of mean and variance that was a severe limitation of Poisson models, a property that is rarely fulfilled in practice. Therefore, it was better to consider the negative binomial distribution, which explicitly models over-dispersion, i.e. the variance is allowed to be larger than the mean (Hens et al. 2009).

Often, over-dispersion is caused by an excess variation between response probabilities or counts, possibly originating from omitting important explanatory predictors (*Hilbe 2007,* Agresti, 2002). A mixture model is a flexible way to account for over dispersion. At a fixed setting of the predictors used, given the mean (λ), the distribution of number of contacts (Y) is Poisson, but the mean itself varies according to some distribution. Suppose that given λ, Y has a Poisson distribution with mean λ, and λ has a gamma distribution, Gamma (k, μ). The gamma probability density function for λ is

$$f(\lambda; \kappa, \mu) = \frac{(\kappa/\mu)^{\kappa}}{\Gamma(\kappa)} \exp\left(-\frac{\kappa\,\lambda}{\mu}\right) \lambda^{\kappa-1}, \ \text{for } \lambda \geq 0 \quad (2)$$

This gamma distribution has $E(\lambda) = \mu,\ var(\lambda) = \mu\text{^}2\ /\ \kappa$. The parameter k>0 describes the shape. The density is skewed to the right, but degrees of skewness decreases as k increases. Marginally, the gamma mixture of the Poisson distributions yields the negative binomial distribution for Y. Its probability mass function is given below (3)

$$P(Y = y_i|\kappa, \mu) = \frac{\Gamma(yi+\kappa)}{\Gamma(\kappa)\Gamma(y_i+1)} \left(\frac{\kappa}{\mu+\kappa}\right)^{\kappa} \left(1 - \frac{\kappa}{\mu+\kappa}\right)^{y_i}, \text{where } y_i = 0,1,2, \ldots \quad (3)$$

The negative binomial distribution has $E(y_i|x_i) = \mu$, $Var(Y) = \mu + \mu^2/k$. The mean of y_i can be formulated using linear predictors (x_i) with log link function, given as follows:

$$\log\left(E(y_i|x_i)\right) = \beta_0 + \beta_k x_i \qquad (4)$$

Where, β_0 is intercept; k=1, 2 ... number of predictors and β_k's are regression parameters. The index k^{-1} is called the dispersion parameter. As k^{-1} close to zero, gamma distribution has $Var(\lambda)$ close to zero and it converges to a distribution that has Var (Y) close to μ and it converges to the Poisson distribution with mean μ. In order to handle post-stratification with respect to age and household size, they are known factors to influence contact behavior, individual observations weighted based on 2000, Belgium census data**.**

Hens et al. (2009) mentioned that empirical count data frequently not only characterized by over-dispersion but also excess zeros. Zero-inflated count models provide a parsimonious yet powerful way to model this type of situation. Such models assume that the data are a mixture of two separate data generation processes: one generates only zeros, and the other is either a Poisson or a negative binomial data-generating process. The result of a Bernoulli trial is used to determine which of the two processes generates an observation. A standard negative binomial model would not distinguish between these two processes, but a zero-inflated model allows for this complication. The weighted negative binomial regression in (3) would be contrasted with its zero-inflated (ZINB) version. The latter is found by replacing (3) by

$$P_{ZI}(Y = y_i|\kappa, \mu) = \begin{cases} \pi(z_i) + \left(1 - \pi(z_i)\right)P(Y = 0|x_i) & \text{if } y_i = 0 \\ \left(1 - \pi(z_i)\right) P(Y = y_i|x_i) & \text{if } y_i > 0 \end{cases} \qquad (5)$$

The mean for y_i with log link function and probability π_i with logit link function given as follows, respectively:

$$\log\left(E(y_i|x_i)\right) = \beta_0 + \beta_k x_i \text{ and } \operatorname{logit}(\pi(z_i)) = \alpha_0 + \alpha_k z_i \qquad (6)$$

Where π denotes the probability of the zeros-governing process and the P ($Y=y_i|x_i$) denotes the negative binomial density function in (3). Where, β_0 and α_0 are intercepts; k=1, 2 ... number of predictors and β_k's and α_k's are regression parameters. Note that the covariate vector z_i used to allow this probability to depend on covariates, which may differ from x_i. If π=0, the zero-inflated negative binomial model simplifies to the negative binomial model. Comparing different models and variable selection were done using AIC.

3.3 Modeling Number of Contacts from Household Survey Data

Let y_{ij} be the number of contacts of the j^{th} participant from the i^{th} household where j=1,2, ..., 1344 and i=1,2,..., 342, and let x_{ij} denote a corresponding covariate mentioned under table 1 like age, gender, household size and etc. The number of contacts from household survey data is correlated count data. As mentioned in section 3.2, count data follow Poisson distribution, which is the starting model, mostly encountered over dispersion problem. In order to capture different sources of heterogeneity in the data like over dispersion, zero-inflation and correlation among observations of the same household, the Poisson model extended to different mixture models. Models for correlated over dispersed count data have been studied well in the last decade (Dobbie and Welsh, 2001; Yau and Lee, 2001; Molenberghs et al., 2007). On the one hand, marginal models accommodate correlated over dispersed data by either ignoring the data dependency during estimation and, correcting afterwards through robust sandwich estimates or by means of generalized estimating equations using the so-called working correlation matrices into the model-fitting algorithm (Hall and Zhang, 2004; Lee et al., 2006; Rose et al., 2006). On the other hand, it is intuitively also appealing to extend the zero-inflated models with cluster/subject specific random effects to account simultaneously for excess zeros and intra-cluster correlation among measurements.

In this study, negative binomial (see equation 4) and zero-inflated negative binomial (see equation 6) regression models, respectively will be extended to hierarchical negative binomial (HNB) and hierarchical zero-inflated negative binomial (HZINB) regression models by including random effects (only random intercepts here), finally a model with minimum AIC will be selected.

As discussed in Molenberghs et al. (2007, 2010), random effects can be applied to models for over dispersed count data, for example, a random intercept can be introduced in the model for the mean structure of the negative binomial model, leading to HNB.

Let b_i random intercept assumed normally distributed with mean zero and variance ${\sigma_b}^2$ then using log link function the expected of y_{ij} given that x_{ij} and b_i in the model for HNB is given using equation (7):

$$\log\left(E\left(y_{ij} \middle| x_{ij}, b_i\right)\right) = (\beta_0 + b_i) + \beta_k x_{ij} \qquad (7)$$

ZINB extended to HZINB for considering hierarchical/correlated data with the inclusion of subject specific random effects in both components of equation 6. The mean for y_{ij} with log link function and mixture probability π_{ij} with logit link function given under equation 8, respectively:

$$\log\left(\mathrm{E}\left(\mathrm{y_{ij}}\middle|\mathrm{x_{ij}},\mathrm{b_i}\right)\right) = (\beta_0 + \mathrm{b_i}) + \beta_k \mathrm{x_{ij}} \ \text{ and } \ \mathrm{logit}\left(\pi_{ij}(\mathrm{z_{ij}},\mathrm{a_i})\right) = (\alpha_0 + \mathrm{a_i}) + \alpha_k \mathrm{z_{ij}} \qquad (8)$$

Where, $b_i \sim N(0,{\sigma_b}^2)$ *and* $a_i \sim N(0,{\sigma_a}^2)$; z_{ij}'s are covariates, which may or may not be the same with x_{ij}; β_0 and α_0 are intercepts; k=1,2,..., number of covariates and β_k's and α_k's are regression parameters. Lee et al. (2006) and Yau et al. (2003) use independent random effects whereas in this study shared random effects used by assuming $a_i = c * b_i$ for some proportionality constant c, implying that ${\sigma_a}^2 = c^2 * {\sigma_b}^2$.

Notice: Household survey data was used only for fitting HNB and HZINB models.

4 Results

Under this section, first, exploratory data analysis are presented using summary measures and plots of proportions that were obtained from a cross - tabulation of different variables versus owning (see figure 1) and touching (see figure 2) status of participants with animals. Second, two fitted logistic regressions for owning (see table 2) and touching (see table 3) status were presented from individual survey data. Third, the number of contacts from individual survey data is modeled and compared using Poisson, weighted negative binomial and zero-inflated negative binomial regression. Finally, the number of contacts from household survey data is analyzed using HNB and HZINB models, and compared. R-software used for analysis of section 4.1, 4.2, 4.3 and 4.4 and SAS has been used for section 4.5.

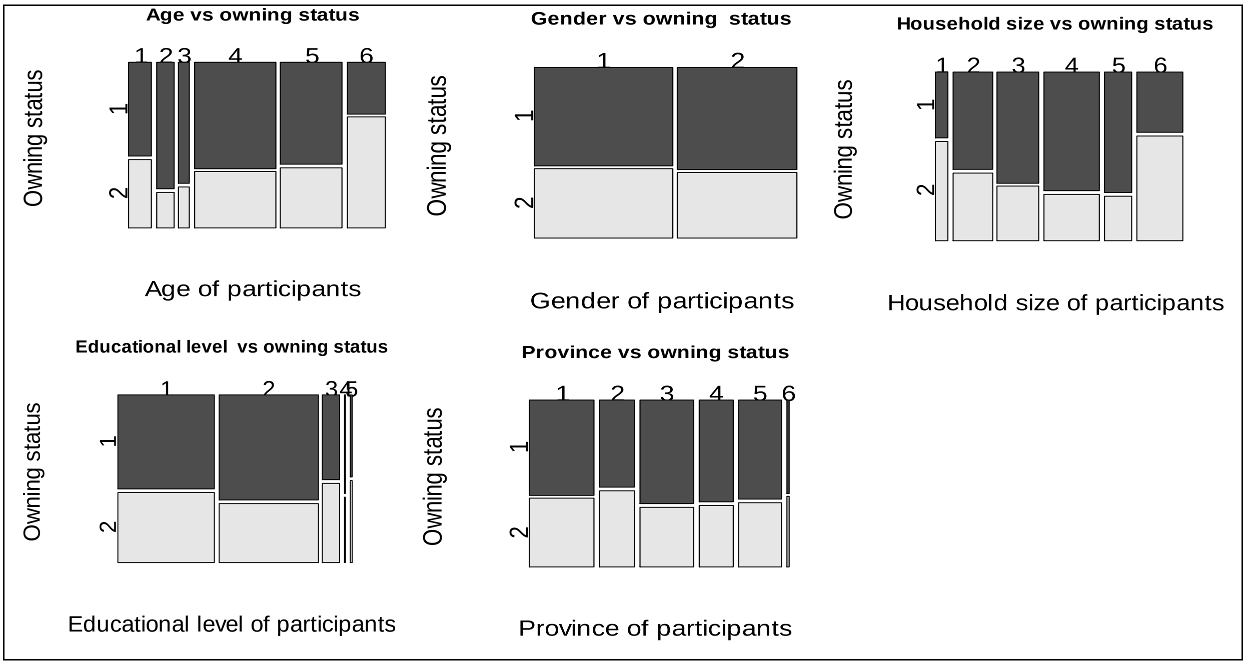
4.1 Exploratory Data Analysis

In order to visualize the degree of touching and owning animals, the types of animals were grouped into four categories — such as pets (including cat, dog and fish), livestock (including horse, sheep, cow and pig), poultry (including chicken, turkey and pigeon) and other animals (including animals which were not specified in the list). From 1768 participants, 503 participants close to 28.45 percent own cat, 437 participants close to 24.72 percent own dog and 267 participants close to 15.05 percent own fish. As a result, the number of participants who own pets is 905 almost 51.19 percent of the total participants. Additionally, 552 participants own only pets which is 31.22 percent of all participants (see table 6, 7, 8, 9 and 10 in appendix for detail). Besides owning status, participants asked to fill the questionnaire about their touching status of animal. Based on this, 450 participants around 25.45 percent touch cat; 491 participants around 27.77 percent touch dog and 6 participants around 0.34 percent touch fish. 812 participants touch pets around 45.93 percent of total participants and 693 participants touch only pets, who are almost 39.20 percent of the total participants.

Livestock are domesticated animals raised in agricultural site to produce commodities such as food, fiber and labor including horse, sheep, cow and pig. Based on the social contact survey data in this study, 264 participants close to 14.93 percent own horse, 21 participants close to 1.19 percent own sheep, 18 participants close to 1.02 percent own cow and 9 participants close to 0.51 percent own pig. Among all participants, 82 participants close to 4.64 percent own livestock and 5 participants close to 0.28 percent had only livestock. In addition, 32 participants close to 1.81 percent touch horse, 6 participants close to 0.34 percent touch sheep, 10 participants close to 0.56 percent touch cow and 1 participant close to 0.06 percent touch pig from total participants. In total, 45 participants touch livestock, which is around 2.55 percent of total participants. There are 5 participants who touch only livestock almost 0.28 percent of total participants (see table 6, 7, 8, 9 and 10 in appendix for detail). There is a reduction in the number of participants, who touch livestock as compared to who own livestock.

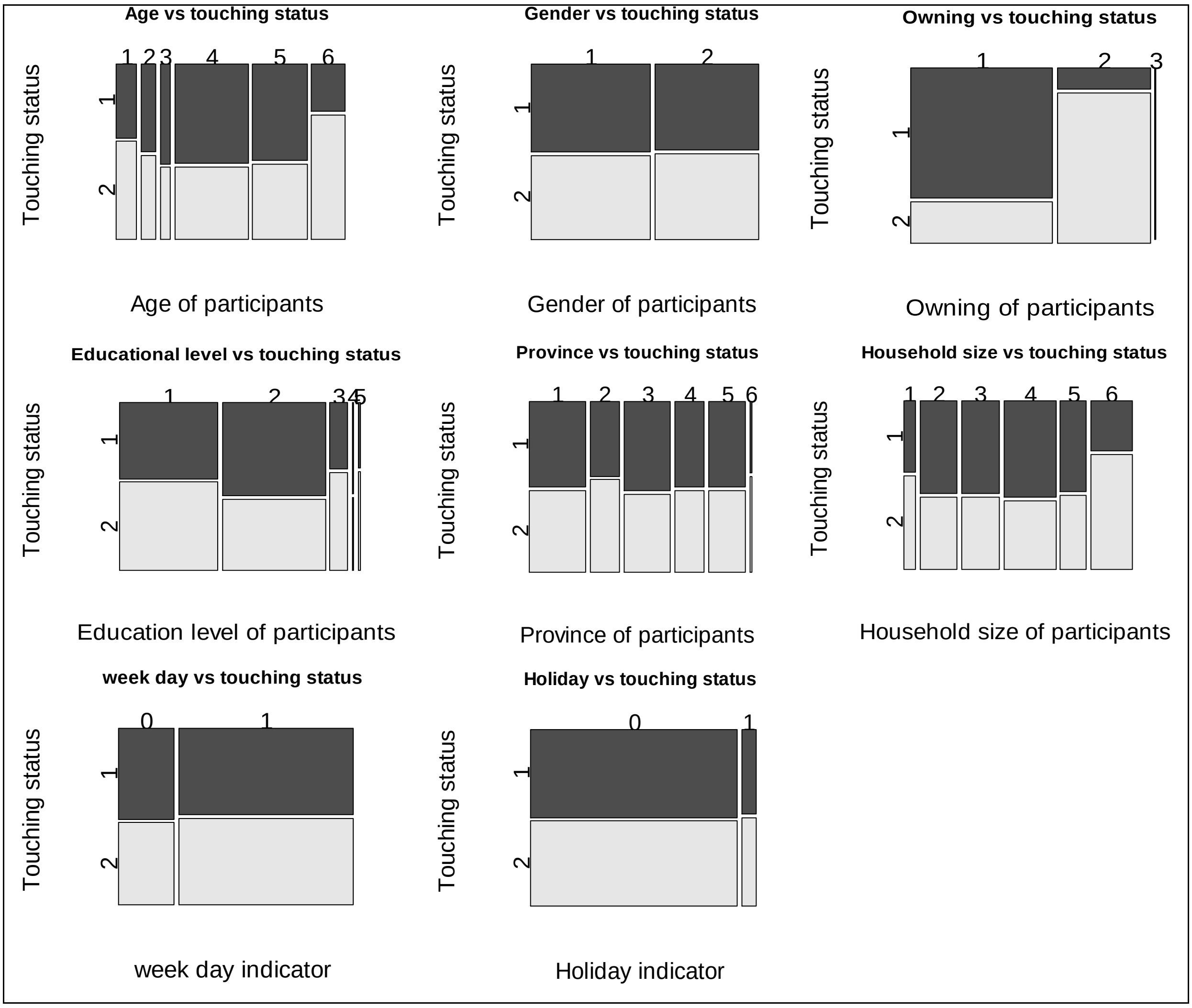
Furthermore, 53 participants close to 3.00 percent own chicken, 5 participants close to 0.28 percent own turkey, 23 participants close to 1.30 percent own pigeon and 326 participants close to 18.44 percent own other animals. Along with owning status of poultry and other animals, 29 participants close to 1.64 percent touch chicken, no participant touch turkey, 10 participants close to 0.56 percent touch pigeon, and 117 participants close to 6.62 percent touch other animals. 706 participants have not owned and 852 participants have not touched animals. Detail information about the number and proportion of participants who own and touch different animals is presented (see table 6, 7, 8, 9 and 10 in appendix).

Figure 1: The proportion of participants who own animals over their age, gender, household size, educational level and province



Under figure 1, the name of the contingency table given at the top of each plot, a vertical bar in each plot correspond to a particular category of a variable in the x-axis. The labels of the variables have mentioned on the top of the corresponding vertical bar (see table 1 from section 2.2 for labels of variables). In the y-axis, the owning status given with black color for yes (1) and gray color for no (2).

Figure 2: The proportion of participants who touch animals versus age, gender, province, educational status, owning status, weekday indicator and holiday variables



Under figure 2, on the y-axis the touching status given with black color for yes (1) and gray color for no (2), on the x-axis independent variables are given with labels of their category. The labels of the variables are mentioned on the top of the corresponding vertical bar (see table 1 from section 2.2 for labels of variables).

As we can see from figure 1, the proportion of participants who own animals is highly pronounced under 6-11 years old participants. The proportion of participants who own animals of 65+ years old participants is the lowest. The insight of getting non-significant result between male and female participants on the owning status is very high. The proportion of having animals is increasing as household size increases. The educational level and province of participants are not showing clear insight about their significant effect on the owning status.

Based on figure 2, age and owning status have shown clear insight about their effect on touching status. The effect of gender, weekday indicator and holiday variables gives the impression of non-significant result on touching status. However, the effect of household size and educational level on touching status has no clear insight.

Chi-squared test for independence used to get an insight about the association of owning status with age, gender, educational level, province and household size, and also touching status with age, gender, educational level, owning status, province, household size, weekday and holiday indicator variables. As expected from figure 1, the chi-square test (p-value<0.05 for all) gives the impression that age, household size and educational level of participant will be important variable for modeling owning status. Nevertheless, gender and province of a participant have no significant effect on owning status based on the chi-square test. Furthermore, age, household size, educational level and owning status of participants have shown significant effect on touching status (p-value<0.05 for all) whereas, gender, province, weekday and holiday indicator variables are not significant. Like any significance test, chi-squared tests of independence have limited usefulness. A small P-value indicates strong evidence of association but provides little information about the nature or strength of the association (Agresti, 2002). Therefore, two multiple logistic regressions had fitted in the next subsections in order to overcome some of the gaps, which are not answered by chi-square test and figure 1 and 2.

4.2 Modeling Owning Status

The owning status of participants taken as dichotomized outcome variable (i.e. Yes/no) after excluding 12 (0.68%) observations with missing category of owning status. Given this outcome, a multiple logistic regression was fitted with covariates — such as the number of contacts, age, gender, educational level, province and household size. The model selection was done based on AIC providing that possible two-way interactions in the model. A model with the minimum AIC of 2200 was chosen as final model and summarized under table 2 including age, household size and educational level. Given the final model, the residual deviance and Pearson residuals divided by residual degree of freedom are close to one indicating that the model had a good fit. Additionally, Hosmer Lemshow goodness of fit test gives p-value=0.55 suggesting that there is no evidence to reject the model.

Table 2: Parameter estimates, standard error and odds ratio with 95% confidence interval of logistic regression for owning status

Covariate	Category	Parameter estimate (S.E)	OR	95% CI OR [Lower : upper]
Age	0-5 years		*1.00*	
	6-11 years	0.85 (0.27)	2.34	[1.38 : 3.94]
	12-17 years	0.49 (0.32)	1.63	[0.86 : 3.07]
	18-44 years	0.54 (0.18)	1.71	[1.19 : 2.46]
	45-64 years	0.52 (0.20)	1.69	[1.13 : 2.52]
	65 + years	-0.50 (0.32)	0.61	[0.32 : 1.14]
Household size	1		*1.00*	
	2	0.76 (0.24)	2.15	[1.35 : 3.43]
	3	1.23 (0.24)	3.42	[2.12 : 5.51]
	4	1.45 (0.24)	4.25	[2.64 : 6.83]
	5 and above	1.49 (0.27)	4.45	[2.62 : 7.58]
	Missing	0.56 (0.31)	1.75	[0.95 : 3.23]
Educational level	Higher education		1.00	
	Secondary education	0.48 (0.11)	1.62	[1.30 : 2.02]
	Primary school	0.38 (0.23)	1.46	[0.94 : 2.27]
	No education	0.78 (0.70)	2.18	[0.55 : 8.63]
	Missing	0.36 (0.56)	1.43	[0.48 : 4.27]

- Intercept = -1.19 (S.E=0.28). Where, OR = odds ratio and S.E= standard error

From table 2, the relative odds of owning animals for (6-11) years of age participants was higher than other age group participants reference to (0-5) years of age participants, which is significantly different from one. For elderly participants (i.e. 65 years and above), the chance of owning animals on the logit scale decreased by 0.5 units controlling other variables in the model. However, the odds of owning animals under (12-17) years and 65 and above years of age group participants was not significantly different from (0-5) years of age participants. The relative odds of having animals is increasing when household size is also increasing in reference to participants with one household size. Controlling other covariates, the odds of having animals for participants under secondary education is 1.62 times than the odds of having animals for participants under higher education. The odds ratio of having animals for secondary education relative to higher education is significantly different from one whereas the primary education and no education levels are not significant.

4.3 Modeling Touching Status

The second multiple logistic regression was performed using touching status of participants as dichotomized outcome variable (i.e. Yes/no) after excluding 44(2.49%) observations with missing category of touching status, and including covariates — such as the number of contacts, age, gender, educational level, province, household size, owning status, week day indicator and holiday variables. The model selection was done based on AIC providing that possible two-way interactions in the model. A model with minimum AIC of 1641 chosen as final model including age, educational level and owning status, summarized under table 3. In the final model, both residual deviance and Pearson residual divided by residual degree of freedom become close to one and Hosmer Lemshow goodness of fit test give p-value=0.26 indicating that no evidence to reject the model.

Table 3: Parameter estimates, standard errors and odds ratio with 95% confidence interval of logistic regression for touching status

Covariate	Category	Parameter Estimate (S.E)	*OR*	95% Confidence Interval for OR
Age	0-5 years		*1.00*	
	6-11 years	-0.20 (0.28)	0.82	[0.47 : 1.43]
	12-17 years	0.23 (0.36)	1.26	[0.62 : 2.55]
	18-44 years	0.62 (0.22)	1.85	[1.21 : 2.85]
	45-64 years	0.52 (0.23)	1.69	[1.07 : 2.65]
	65 + years	-0.16 (0.27)	0.85	[0.50 : 1.45]
Educational level	Higher education		1.00	
	Secondary education	0.41 (0.13)	1.51	[1.16 : 1.96]
	Primary school	0.10 (0.27)	1.11	[0.65 : 1.89]
	No education	0.80 (0.91)	2.22	[0.38 : 13.15]
	Missing	0.15 (0.73)	1.16	[0.28 : 4.84]
Owning status	Yes		1.00	
	No	-3.07 (0.14)	0.05	[0.04 0.06]

- Intercept = 0.60 (S.E=0.198), where, OR=Odds ratio and S.E= standard error

Under table 3, the logistic regression with dichotomous touching status of participants is summarized using odds ratio of touching animals given three covariates — such as age, educational level and owning status of participants. The odds of touching animals for (18-44) and (45-64) years of age participants relative to the odds of touching animals for (0-5) years of age participants is higher than the other age group participants. In addition, the odds ratio of touching animals for (6-11), (12-17), and 65 and above years old participants relative to (0-5) years of age participants was not significantly different from one. The relative odds of touching animals for all educational levels of participants in reference to higher education was higher than one. However, the odds ratio of touching animals for participants under secondary education referencing participants under higher education is the only educational level that gives significant result. The probability of touching animals given a participant own animal was 0.65 after controlling the level of other covariates at their reference group in the model. This showed that a participant who own animals has higher chance to touch animals.

4.4 Modeling Number of Contacts from Individual Survey Data

Under this section, the number of contacts — that is count data type modeled using Poisson regression assuming a data without over-dispersion, negative binomial (NB) regression with over-dispersion parameter and zero-inflated negative binomial (ZINB) regression assuming a data with over-dispersion and excesses zero counts. The Poisson regression model assumes that the data are equally dispersed, which means the conditional variance equals the conditional mean. The Poisson model has been criticized for its restrictive property that the conditional variance equals the conditional mean. In real life, count data characterized by over-dispersion — that is, the variance exceeds the mean. The negative binomial regression model is a generalization of the Poisson regression model that allows for over-dispersion by introducing an unobserved heterogeneity term for each observation.

Figure 3: The observed number of contacts on histogram (see grey color) and Poisson approximation (see red color)

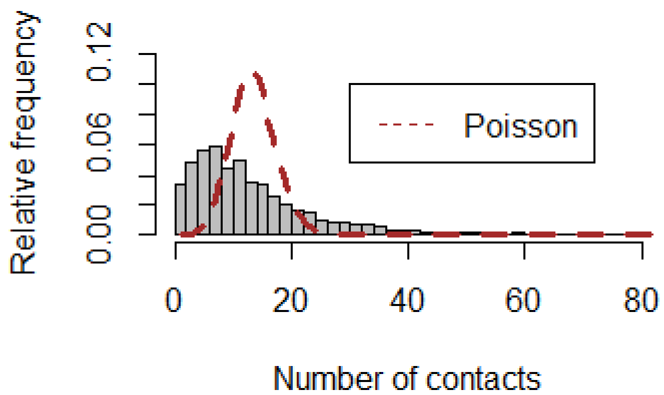


Figure 3 provides an insight about over dispersion since the Poisson approximation for number of contacts by their sample mean (see red dashed line) deviates from the observed number of contacts. The mean and variance for the number of contacts was 13.47 and 116.78, respectively— suggesting that the data had over-dispersion since the sample variance was very large as compared to the mean. Based on this insight, negative binomial regression model was used for variable selection — given that age, gender, educational level, household size, touching status, owning status, province, holiday and weekday indicator of participants, including possible two-way interactions in the model. Model selection (variable selection in regression is a special case) is a bias versus a variance tradeoff, and this is the statistical principle of parsimony. Inference under models with too few parameters or variables can be biased, while with models having too many parameters or variables, there may be poor precision or identification of effects that are, in fact, spurious. These considerations call for a balance between under and over fitted models—the so-called model selection problem (Forster 2000). The variable selection was done using AIC. A model, which contains age, household size, educational level, owning status, province, weekday indicator, holiday indicator, and interaction effect of owning status and weekday indicator, was selected with minimum AIC of 12115.

After having the selected variables, Poisson regression, NB and ZINB were fitted and compared using AIC. The over dispersion parameter is significant since its 95% CI [2.18: 2.55] is not including 0 (see table 4). Therefore, it is plausible to use the negative binomial distribution than Poisson distribution. There are 16 (0.90%) participants without contact with anyone on the assigned day, which is the number of contacts was zero for these participants. The model comparison confirmed by the minimum AIC — such that the AIC for Poisson, NB and ZINB was 18358, 12115 and 12117, respectively. The AIC values for NB and ZINB are close to each other likewise the curves for NB and ZINB predicted number of contacts are close to the actual data where as Poisson prediction deviate from the data (see figure 4 in appendix). Therefore, the most parsimonious weighted negative binomial regression fitted and summarized under table 4 to look at the effect of covariates on the number of contacts.

As we can see from table 4, participants aged from (6 -11) and (12-17) years had the highest observed mean for the number of contacts , while participants aged from 65 years and above had the lowest observed mean for the number of contacts. After building negative binomial regression, the highest relative number of contacts was under participants aged (12-17) years while the lowest was under participants aged from 65 years and above relative to (0-5) years of age participants. The relative number of contacts with 65 and above years old participants was significantly different and below 1.00 — implying that the number of contacts for 65 and above years old participants is lower than for (0-5) years old participants. Given the results under table 4, the larger average number of contacts from the observed data is belonging to the larger household size. This has been confirmed from the result of negative binomial regression using a relative number of contacts, as a result, the relative number of contacts for participants with three, four and five and above household size relative to participants with one household size is significantly different and above one.

Moreover, the average number of contacts from observed data was falling down when educational level of the participants was decreasing from higher education to no education. After taking participants with higher education as reference, the relative number of contacts for participants with secondary education, primary school and no education is significantly different and below one. The effect of owning status of participants on the number of contacts will depend on the level of weekday indicator (i.e. Week or weekend days). The relative number of contacts for participants who do not own animals relative to participants who own animals was 0.81 during the weekend days, after controlling other covariates. Limburg showed the largest average and relative number of contacts between people relative to Antwerp than other provinces. In addition, the average and relative number of contacts during the holidays was lower than regular days.

Table 4: Observed mean and standard deviation (S.D) for the number of contacts: estimated relative number of contacts (RNC) with 95% confidence interval (CI) from weighted negative binomial regression

Covariates	Category	Mean	(S.D)	RNC	95% CI for RNC [Lower : Upper]
Age	0-5 years	13.79	(11.40)	1.00	
	6-11 years	20.41	(13.06)	1.42	[1.18 : 1.71]
	12-17 years	19.59	(15.03)	1.57	[1.28 : 1.92]
	18-44 years	13.90	(10.11)	1.08	[0.93 : 1.25]
	45-64 years	12.86	(9.85)	1.02	[0.87 : 1.19]
	65 + years	8.72	(8.10)	0.65	[0.52 : 0.81]
Household size	1	10.69	(8.61)	1.00	
	2	12.71	(9.53)	1.14	[0.98 : 1.32]
	3	13.64	(9.75)	1.24	[1.07 : 1.44]
	4	15.95	(11.99)	1.36	[1.17 : 1.58]
	5+	18.06	(12.30)	1.52	[1.29 : 1.78]
	Missing	9.04	(8.72)	1.39	[1.14 : 1.68]
Educational level	Higher education	14.73	(11.25)	1.00	
	Secondary education	12.70	(10.17)	0.87	[0.81 : 0.94]
	Primary school	11.53	(10.81)	0.78	[0.67 : 0.90]
	No education	5.50	(5.91)	0.46	[0.28 : 0.78]
	Missing	14.38	(14.40)	1.04	[0.74 : 1.49]
Owning Status	Yes	14.30	(11.19)	1.00	
	No	12.28	(10.09)	0.81	[0.70 : 0.93]
Province	Antwerp	12.76	(9.92)	1.00	
	Limburg	14.80	(11.68)	1.27	[1.14 : 1.42]
	East-Vlaanderen	13.38	(10.51)	1.02	[0.93 : 1.13]
	Flemish-Brabant	13.61	(11.28)	1.05	[0.94 : 1.17]
	West-Vlaanderen	13.92	(11.33)	1.14	[1.03 : 1.26]
	Missing	5.07	(4.75)	0.61	[0.40 : 0.96]
Weekday indicator	Weekend day	12.40	(10.69)	1.00	
	Week day	13.84	(10.82)	0.99	[0.90 : 1.10]
Holiday indicator	Regular day	13.73	(10.93)	1.00	
	Holiday	9.93	(7.93)	0.82	[0.71 : 0.95]
Owning: weekday indicator*	No : Weekday	12.77	(10.23)	1.21	[1.03 : 1.42

- Estimate of intercept is 2.34 with standard error of 0.094
- Estimate of dispersion parameter is 2.37 and standard error is 0.094 with 95% confidence interval of [2.18 : 2.55]
- * indicates interaction between two covariates

4.5 Modeling Number of Contacts from Household Survey Data

The number of contacts from household survey data are correlated each other since observations are clustered in the same household. The mean and variance for the number of contacts from household survey data was 15.66 and 125.89, respectively, in which the data showed sense of over-dispersion since the sample variance was very large as compared to the mean. There are 6 participants with zero number of contacts; their coverage is 0.45 percent from 1344 participants. Both HNB and HZINB were fitted within the maximum likelihood framework using the SAS procedure NLMIXED. Variable selection was based on AIC and using HNB model; moreover, HNB and HZINB were compared based on AIC. Initial values were obtained from ZINB for intercept and regression parameters of both component of HZINB (see equation 6) and the initial value for the variance of random intercept was obtained from HNB using SAS procedure GLIMMIX. Estimation was done by maximum likelihood framework using Adaptive Gaussian Quadrature. After imposing different combination of age, gender, educational level, province, household size, owning status, touching status, weekday indicator and holiday indicator explanatory variables with possible two-way interactions in HNB model, a model with age, household size, owning status, weekday indicator, holiday indicator and interaction effect between owning status and weekday indicator has a minimum AIC (i.e. 9334). The AIC for HZINB using the selected variables was 9332. Both models have very closed AIC but HZINB had smaller AIC than HNB, therefore, HZINB is used for further discussion about the effect of explanatory variables.

There was no a single covariate plays a role in explaining the model for mixture probability π_{ij} (see table 5). Therefore, the logit component of HZINB modeled only using averaged intercept (α_0) and random intercept (a_i). In the log scale component of HZINB, age, household size, owning status, weekday indicator, holiday indicator, and interaction between owning status and weekday indicator variables were important to model the correlated number of contacts. However, there was no significant evidence to support the effect of gender, province and educational level on the number of contacts. In order to interpret the regression parameters under HZINB, it is necessary to condition on b_i. The expected number of contacts on the log-scale will for 45-64 years old participant in reference to 0-5 years of age participant conditional on b_i is not significant. However, the expected number of contacts on the log-scale for 6-11 years, 12-17 years and 18-44 years old participant in reference to 0-5 years old participant is significant, given that b_i.

Table 5: Parameter estimates standard errors (SE) and 95% confidence interval (CI) of zero-inflated negative binomial with shared random effects (HZINB) model.

Model for $\log\left(E(y_{ij} \mid x_{ij}, b_i)\right)$		
Effect (reference)	Estimate (SE)	95% CI [Lower : Upper]
Intercept (β_0)	2.26 (0.18)	[1.90 : 2.62]
Age (0-5 years)		
6-11 years	0.19 (0.05)	[0.08 : 0.29]
12-17 years	0.17 (0.06)	[0.05 : 0.29]
18-44 years	0.11 (0.04)	[0.02 : 0.20]
45-64 years	-0.01 (0.07)	[-0.15 : 0.12]
Household size (2)		
3	0.07 (0.18)	[-0.28 : 0.42]
4	0.14 (0.17)	[-0.20 : 0.48]
5+	0.37 (0.18)	[0.02 : 0.72]
Owning-status (Yes)		
No	-0.30 (0.11)	[-0.51 : -0.08]
Missing	-0.71 (0.43)	[-1.57 : 0.14]
Weekday-Indicator (Weekend days)		
Weekday	0.15 (0.08)	[-0.001 : 0.30]
Holiday-Indicator (Regular days)		
Holiday	-0.53 (0.28)	[-1.07 : 0.01]
Owning-status x Weekday-Indicator		
No : Weekday	0.27 (0.13)	[0.02 : 0.53]
Missing : Weekday	0.92 (0.50)	[-0.06 : 1.90]
Dispersion (k^{-1})	0.16 (0.01)	
Variance of b_i (σ_b^2)	0.21 (0.02)	
Proportionality factor (c)	1.14 (2.96)	
Model for $\text{logit}\left(\pi_{ij}(z_{ij}, a_i)\right)$		
Intercept (α_0)	-6.98 (1.38)	[-9.70 : -4.26]

In the same fashion with the result obtained from individual survey data, the number of contacts increased with the increase in the household size conditional on b_i in HZINB model. Besides, the effect of owning animals on the number of contacts from household survey data depends on the level of weekday indicator, like in the individual survey data. People have less contact during holidays as compared to regular days.

5 Discussion

Infectious diseases can be considered to spread over social networks of people or animals. To mitigate their spreading we need to understand microbial pathogens, social and behavioral factors and the social structure of infection. Pathogens can be transmitted from one host to the other through airborne, direct or indirect physical contact, contaminated food or water, sexual contact or vector borne etc. Social contact survey has proved an indispensable approach for understanding epidemics of infectious disease (Masuda et al. 2013). Recruitment for social contact surveys is difficult because participants must be willing to give information about their social behavior. The contact survey conducted as part of this study as well required participants to complete a survey anonymously, without changing their usual behavior. When conducting a survey, having a representative sample of the population is of paramount importance, as a result, post-stratification adjusted weights were used to compensate for having representative sample.

Like people, animals also carry germs and viruses which can pass from animals to humans called zoonotic illnesses like HIV-AIDS and avian influenza virus. In this study, almost half of the participants own or touch pets. People derive a lot of joy and solace from their pets. Even though, pets also carry certain bacteria, viruses, parasites, and fungi that can cause illness if transmitted to humans. However, not all illnesses are transmitted to humans which are common among house pets such as: distemper, canine parvovirus, and heartworms. Humans got these animals-borne diseases when they are bitten or scratched or have contact with an animal's waste, saliva, or dander. Besides pets, people own and touch livestock, poultry and other animals which could carry different transmissible pathogens that cause illness in humans. Estimating transmission parameter and the basic reproduction number in mathematical models is the crucial step for modeling the spread of zoonotic illnesses even it is hard to determine animal-human interactions easily.

In this study, different factors assessed for owning status, touching status and number of contacts between people that could enable transmission of different pathogens from animals to humans. The chance of owning and touching animals for 65 and above years old participants is lower than participants with other age levels. This result is strange because people are connecting them with pets as they grow older in reality of getting pleasure and to boost their morale and optimism.

However, participants aged from 18 up to 64 years have higher odds of touching animals; so, they are highly exposed for zoonotic illnesses. The odds of owning animals is increasing with the increase in household size but household size is not significant for touching status. Participants who involve under higher education have the lowest chance to own and touch. As expected, there is a higher chance to touch animals given that a person has owned animals.

Given the individual contact survey data, one person per household was recruited to take part. As mentioned under data and survey method section, contact was defined in two ways in this study: first, when a person has spoken in his/her presence in less than three meters with someone and second, when a person had physical contact with someone. Both types of contact could enable pathogen transmission by close contact. Based on this reasoning, the number of contacts for each participant got focus in this study. As a result, the average and relative number of contacts with (6 -11) and (12-17) years old participants were considerably higher than other age group participants. As expected, pathogen transmission will be higher for larger household size since the average and relative number of contacts is increasing with the increase in household size. Educated people have higher social contact than people who are illiterate. One of the most important findings of this study is that the effect of owning status on the number of contacts between people. Consequently, the effect of owning animals on the number of contacts depends on weekday indicator. The relative number of contacts for participants who have not owned animals versus who have owned animals is below one during weekend days. Considering provinces, the average and relative number of contacts in Limburg are larger than Antwerp, East-Vlaanderen, Flemish-Barabant and West-Vlaanderen. People also show a higher number of contacts during regular days than holidays.

From the household contact survey data, there is no sufficient evidence to support effect of gender, educational level and province on the correlated number of contacts. However, the effect of age, household size, owning status, holiday indicator and interaction between owning status and weekday indicator has shown important role on number of contacts in line with the result from individual survey data.

6 Conclusion

In summary, age, household size and educational level of participants had a significant effect on owning animals, however, no evidence to support the effect of the number of contacts, gender and province. In addition, touching status is explained by age, educational level and owning status of participants while the number of contacts, household size, gender, province, weekday and holiday indicator have not shown evidence to support their effect on touching status. Finally, we observe a significant effect of age, household size, owning status, holiday indicator and interaction between owning status and weekday indicator on the number of contacts from both individual and household survey data. Educational level and province of a participant plays important role on the number of contacts from individual survey data, however, there is no supporting clue about the effect of both variables on the number of contacts from household survey data. However, there is no evidence to support the effect of gender on the number of contacts on both individual and household survey data.

References

Agresti, A. (2002). Categorical Data Analysis (2nd edn.). John Wiley & Sons, Inc., Hoboken, New Jersey.

Beutels P, Shkedy Z, Aerts M, 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.

Cleaveland, S., Laurenson, M. K. & Taylor, L. H. (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. Series B, Biological Sciences 356, 991-999

Dobbie, M.J. and Welsh, A.H. (2002). Modelling correlated zero-inflated count data. Australian and New Zealand Journal of Statistics 43, 431-444.

Forster, Malcolm R. (2000). “Key Concepts in Model Selection: Performance and Generalizability.” Journal of Mathematical Psychology 44:205-31.

Grassly NC, Fraser C (2008) Mathematical models of infectious disease transmission. Nat Rev Microbiol 6: 477–487.

Hall, D.B. and Zhang, Z. (2004). Marginall models for zero-inflated clustered data. Statistical Modelling 4, 161-180.

Hens N, Ayele GM, Goeyvaerts N, Aerts M, Mossong J, et al. (2009) Estimating the impact of school closure on social mixing behavior and the transmission of close contact infections in eight European countries. BMC Infect Dis 9-187.

Hilbe, Joseph M. (2007), Negative Binomial Regression, Cambridge, UK: Cambridge University Press.

Horby P, Thai PQ, Hens N, Yen NTT, Mai LQ, et al. (2011) Social Contact Patterns in Vietnam and Implications for the Control of Infectious Diseases. PLoS ONE 6(2): e16965. doi:10.1371/journal.pone.0016965

Jebara, K. B. (2004) Surveillance, detection and response: managing emerging diseases at national and international levels. Revue Scientifique et Technique – Office International des Epizooties **23,** 709-715

Lee, A.H., Wang, K., Scott, J. A., Yau, K.K.W. and McLachlan, G.J.(2006). Multi-level zero-inflated Poisson regression modeling of correlated count data with excess zeros. Statistical Methods in Medical Research 15, 47-61.

Masuda N and Holme P.(2013) Predicting and controlling infectious disease epidemics using temporal networks: 5-6.

Molenberghs, G., Verbeke, G., Demetrio, C.G.B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. Lifetime Data Analysis 13, 513-531.

Molenberghs, G., Verbeke, G., Demetrio, C.G.B. and Vieira, A.M.C. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. Statistical Science 25, 325-347.

Morse, S. S. (2004): Factors and determinants of disease emergence. Revue Scientifique et Technique – Office International des Epizooties 23, 443-451

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 5-74.

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

Ryan KJ; Ray CG (editors) (2004). Sherris Medical Microbiology (4th Ed.). McGraw Hill. ISBN 0-8385-8529-9.

Rose, C.E., Martin, S.W., Wannemuehler, K.A. and Plikaytis, B.D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of Biopharmaceutical Statistics 16.

Sharp P.M., and Hahn B.H., (2013). Origins of HIV and the AIDS Pandemic. Cold Spring Harbor Laboratory Press.

Taylor, L. H., latham,S.M.& Woolhouse, M. E. (2001): Risk factors for human disease emergence. Philosophical transactions of the Royal Society of London. Series B, Biological Sciences 356, 983-989

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

Yau, K.K.W., and Lee, A.H. (2001). Zero-inflated Poisson regression with random effects to evaluate an eccupational injury prevention programme. Statistics in Medicine 20, 2907-2920.

Yu Q, Liu L, Pu J, Zhao J, Sun Y, Shen G, Wei H, Zhu Y, Zheng R, Xiong D, Liu X, and Liu J., (2013). Risk perceptions for avian influenza virus infection among poultry workers, China. Emerg Infect Dis.

Appendix

Table 6: Categorical frequency distribution for possible combination of owning or touching pets, livestock, poultry and other animals

Possibilities of either owning or touching pets, lives stocks, poultry and other animals	Number (proportion) of participants who owned animals	Number (proportion) of participants who touched animals
None of them	706 (39.93%)	850 (48.08%)
Pets only	552 (31.22%)	693 (39.20%)
Livestock only	5 (0.28%)	5 (0.28%)
Poultry only	40 (2.26%)	11 (0.62%)
Other only	66 (3.73%)	42 (2.38%)
Pets and livestock only	21 (1.19%)	28 (1.58%)
Pets and poultry only	84 (4.75%)	17 (0.96%)
Pets and other only	123 (6.96%)	59 (3.34%)
Livestock & poultry only	1 (0.06%)	0 (0.00%)
Livestock & other only	2 (0.11%)	1 (0.06%)
Poultry & other only	28 (1.58%)	3 (0.17%)
Pets, livestock & poultry only	21 (1.19%)	3 (0.17%)
Pets, livestock & other only	9 (0.51%)	8 (0.45%)
Pets, poultry & other only	75 (4.24%)	4 (0.23%)
Livestock, poultry & other only	3 (0.17%)	0 (0.00%)
All of them	20 (1.13%)	0 (0.00%)
Missing	12 (0.68%)	44 (2.49%)

Table 7: The number and proportion of participants who owned cat, dog, horse, chicken, turkey, sheep, pigeon, pig, cow, fish, and other animals

Animals	Number (proportion) of participants owning animals		
	No	*Yes*	*Missing*
Cat	1253 (70.87%)	503 (28.45%)	12 (0.68%)
Dog	1319 (74.60%)	437 (24.72%)	12 (0.68%)
Horse	1703 (96.32%)	53 (3.00%)	12 (0.68%)
Chicken	1492 (84.39%)	264 (14.93%)	12 (0.68%)
Turkey	1751 (99.04%)	5 (0.28%)	12 (0.68%)
Sheep	1735 (98.13%)	21 (1.19%)	12 (0.68%)
Pigeon	1733 (98.02%)	23 (1.30%)	12 (0.68%)
Pig	1747 (98.81%)	9 (0.51%)	12 (0.68%)
Cow	1738 (98.30%)	18 (1.02%)	12 (0.68%)
Fish	1490 (84.27%)	267 (15.05%)	12 (0.68%)
Other	1430 (80.88%)	326 (18.44%)	12 (0.68%)

Table 8: The number and proportion of participants who owned pets, livestock, poultry and other animals

	Number (proportion) of participants owning animals					
	No		Yes		Missing	
Pets	851	(48.13%)	905	(51.19%)	12	(0.68%)
Livestock	1674	(94.68%)	82	(4.64%)	12	(0.68%)
Poultry	1484	(83.94%)	272	(15.38%)	12	(0.68%)
Other	1430	(80.88%)	326	(18.44%)	12	(0.68%)

Table 9: The number and proportion of participants who touched cat, dog, horse, chicken, turkey, sheep, pigeon, pig, cow, fish and other animals

	Number(proportion) of participants touching animals					
	No		*Yes*		*Missing*	
Cat	1274	(72.06%)	450	(25.45%)	44	(2.49%)
Dog	1233	(69.74%)	491	(27.77%)	44	(2.49%)
Horse	1692	(95.70%)	32	(1.81%)	44	(2.49%)
Chicken	1695	(95.87%)	29	(1.64%)	44	(2.49%)
Turkey	1724	(97.51%)	0	(0.00%)	44	(2.49%)
Sheep	1718	(97.17%)	6	(0.34%)	44	(2.49%)
Pigeon	1714	(96.95%)	10	(0.56%)	44	(2.49%)
Pig	1723	(97.45%)	1	(0.06%)	44	(2.49%)
Cow	1714	(96.95%)	10	(0.56%)	44	(2.49%)
Fish	1718	(97.17%)	6	(0.34%)	44	(2.49%)
Other	1610	(90.89%)	117	(6.62%)	44	(2.49%)

Table 10: The number and proportion of participants who touched pets, livestock, poultry, and, other animals

	Number of participants touching animals (proportion)					
	No		Yes		Missing	
Pets	912	(51.58%)	812	(45.93%)	44	(2.49%)
Livestock	1679	(94.97%)	45	(2.55%)	44	(2.49%)
Poultry	1686	(95.36%)	38	(2.15%)	44	(2.49%)
Other	1607	(90.89%)	117	(6.62%)	44	(2.49%)

Figure 4: The observed number of contacts on a histogram and predicted number of contacts on Poisson (see black color), negative binomial (see red color) and zero-inflated negative binomial (see green color) regression curves

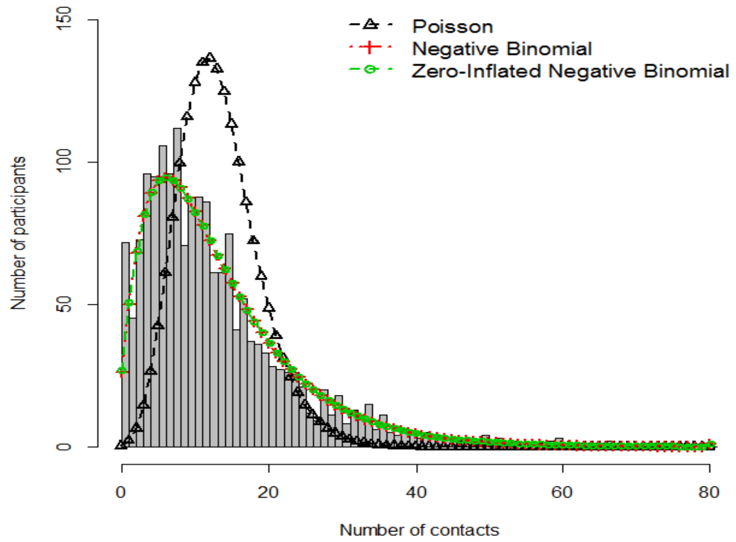


Figure 5: The observed number of contacts on a histogram and predicted number of contacts on hierarchical negative binomial (see red color) and hierarchical zero-inflated negative binomial (see green color) regression curves

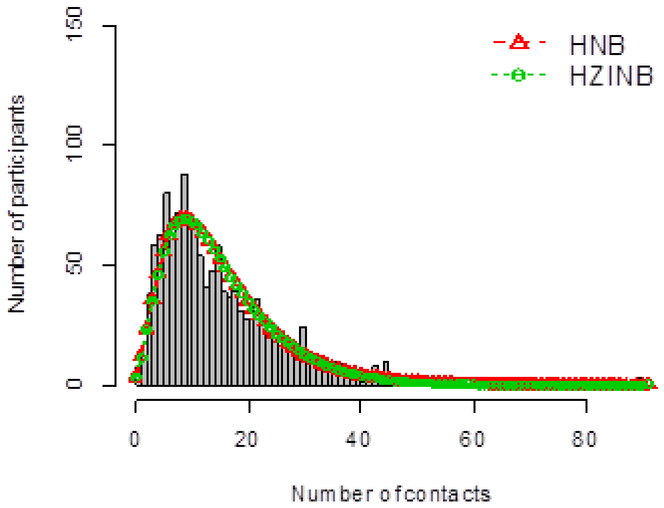


Table 11: Summary of logistic regression for owning status including all covariates and possible two-way interactions

Coefficients	Parameter Estimate	Standard Error	Z-value	P-value
Intercept	-1.52	0.36	-4.253	2.11E-05 ***
Number of contacts (Count)	0.02	0.02	1.057	0.29029
Age:6-11 years	1.18	0.48	2.467	0.01364 *
Age:12-17 years	1.09	0.53	2.072	0.03831 *
Age:18-44 years	0.67	0.30	2.223	0.02623 *
Age:45-64 years	0.43	0.32	1.339	0.18068
Age:65 + years	-0.18	0.41	-0.444	0.65685
Gender: male	0.10	0.11	0.937	0.34857
Household size :2	0.74	0.24	3.076	0.00209 **
Household size :3	1.20	0.25	4.871	1.11E-06 ***
Household size :4	1.39	0.25	5.693	1.25E-08 ***
Household size :5+	1.43	0.28	5.214	1.84E-07 ***
Household size :missing	0.58	0.32	1.831	0.06715
Education: Secondary	0.53	0.12	4.564	5.02E-06 ***
Education: Primary school	0.39	0.23	1.718	0.08579
Education: No education	0.93	0.71	1.300	0.19346
Education: Missing	0.50	0.57	0.890	0.37366
Province: Limburg	-0.09	0.17	-0.570	0.56897
Province: East-Vlaanderen	0.28	0.15	1.922	0.05463
Province: Flemish-Brabant	0.25	0.17	1.454	0.14595
Province: West-Vlaanderen	0.12	0.16	0.754	0.45068
Province: Missing	1.03	0.58	1.789	0.07354
Age (6-11 years): count	-0.02	0.02	-0.968	0.33326
Age (12-17 years): count	-0.03	0.02	-1.531	0.12565
Age (18-44 years): count	-0.01	0.02	-0.672	0.50168
Age (45-64 years): count	0.01	0.02	0.334	0.73861
Age (65 + years): count	-0.04	0.02	-1.812	0.06991

Table 12: Summary of logistic regression for touching status including all covariates and possible two-way interactions

Coefficients	Parameter Estimate	Standard Error	Z-value	P-value
Intercept	1.34	0.47	2.87	0.00410 **
Number of contacts (Count)	-0.003	0.006	-0.43	0.66681
Age:6-11 years	-0.16	0.29	-0.56	0.57743
Age:12-17 years	0.22	0.37	0.60	0.54558
Age:18-44 years	0.58	0.23	2.50	0.01235 *
Age:45-64 years	0.56	0.25	2.20	0.02755 *
Age:65 + years	0.57	0.42	1.35	0.17583
Gender: male	-0.68	0.56	-1.22	0.22144
Household size :2	-0.20	0.42	-0.48	0.63263
Household size :3	-0.46	0.42	-1.09	0.27718
Household size :4	-0.40	0.42	-0.96	0.33766
Household size :5+	-0.36	0.46	-0.78	0.43594
Household size : missing	-1.36	0.51	-2.67	0.00759 **
Education: Secondary	0.43	0.14	3.09	0.00198 **
Education: Primary school	0.13	0.29	0.437	0.66223
Education: No education	0.99	0.89	1.10	0.27013
Education: Missing	0.30	0.73	0.41	0.68010
Owning status: No	-3.11	0.15	-21.27	< 2E-16 ***
Province: Limburg	-0.22	0.21	-1.08	0.28099
Province: East-Vlaanderen	-0.13	0.18	-0.72	0.47209
Province: Flemish-Brabant	-0.06	0.20	-0.28	0.77805
Province: West-Vlaanderen	-0.11	0.19	-0.58	0.55967
Province: Missing	0.09	0.72	0.13	0.89636
Weekday indicator: weekday	-0.21	0.15	-1.38	0.16902
Holiday indicator: holiday	-0.12	0.27	-0.44	0.65967
Gender(male): Household size (2)	0.50	0.64	0.78	0.43430
Gender(male): Household size (3)	0.69	0.63	1.09	0.27401
Gender(male): Household size (4)	0.68	0.61	1.11	0.26838
Gender(male): Household size (5+)	0.35	0.65	0.54	0.59189
Gender(male):Household-size (missing)	0.99	0.64	1.55	0.12151

Table 13: Summary of weighted negative binomial regression for the number of contacts including all covariates and possible two-way interactions

Coefficients	Parameter Estimate	Standard Error	Z-value	P-value
Intercept	2.35	0.11	20.91	< 2E-16***
Age:6-11 years	0.35	0.09	3.69	0.000226***
Age:12-17 years	0.45	0.10	4.42	9.85E-06***
Age:18-44 years	0.07	0.07	0.97	0.330576
Age:45-64 years	0.01	0.08	0.17	0.863984
Age:65 + years	-0.43	0.10	-4.16	3.17E-05***
Gender: male	-0.03	0.04	-0.75	0.452345
Household size :2	0.13	0.08	1.72	0.086183
Household size :3	0.21	0.08	2.80	0.005072**
Household size :4	0.31	0.08	4.05	5.13E-05***
Household size :5+	0.42	0.08	5.12	3.09E-07***
Household size :missing	0.32	0.09	3.47	0.000520***
Education: Secondary	-0.14	0.04	-3.62	0.000292***
Education: Primary school	-0.26	0.07	-3.61	0.000301***
Education: No education	-0.77	0.26	-3.01	0.002655**
Education: Missing	0.04	0.18	0.21	0.831665
Owning status: No	-0.21	0.072	-2.94	0.003283**
Province: Limburg	0.24	0.06	4.33	1.47E-05***
Province: East-Vlaanderen	0.02	0.05	0.46	0.645889
Province: Flemish-Brabant	0.05	0.06	0.82	0.410808
Province: West-Vlaanderen	0.13	0.05	2.59	0.009739**
Province: Missing	-0.50	0.22	-2.26	0.023850 *
Weekday indicator: weekday	-0.005	0.05	-0.09	0.931204
Holiday indicator: holiday	-0.20	0.08	-2.66	0.007857**
Owning status (No): Weekday indicator (weekdays)	0.19	0.08	2.37	0.017919*

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Human-animal interactions: Survey of contact behavior relevant for the spread of infectious diseases

Richting: **Master of Statistics-Biostatistics**
Jaar: **2013**

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,

Kifle, Yimer Wasihun

Datum: **11/09/2013**