Causal analysis of H1N1pdm09 influenza infection risk in a household cohort (Supplementary File)

Causal modeling with structural equations

Structural equation modeling (SEM) is a multivariate statistical method frequently used in social sciences [1,2]. This hypothesis-driven approach can be used to test a series of structural equations (*i.e.* relationships between variables) representing a theoretical causal process. Two types of variables, observed and latent, are included in SEM models. Latent variables are unobserved variables that cannot be measured directly [3]. They are estimated from several observed variables, called indicators, presumed to represent an underlying unobserved phenomenon [4]. Latent and observed variables can be classified as exogenous or endogenous. The former are independent variables not influenced by the other variables of the model, while the latter are dependent variables [5,6]. The set of hypotheses of a given SEM can be summarized with a path diagram (Figure 1). The SEM model is composed of two compartments: the measurement model and the structural model [7]. The first shows the relations between the latent variables and their indicators; arrows going from the latent variable to its respective indicators are called factor loadings and represent correlation coefficients [8]. The structural model represents the relations between the latent variables. The SEM model is a combination of the measurement model, the structural model, and potential endogenous or exogenous observed variables.

Additional information on covariates used in the SEM model

A summary of the covariates included in the SEM model is provided on table 1. Table 2 shows the correlations between the covariates included in the measurement model.

HAI titers of the subjects included in the study

Pre-epidemic HAI titers depending on subjects' infection status and age class are shown on Figure 2. For each age class, we observed that the non-infected subjects showed higher log pre-epidemic HAI titers than the infected individuals (Wilcoxon rank-sum test $P<0.001$).

Model estimation

As we included categorical dependent variables, we used the weighted least squares with mean and variance adjustment (WLSMV) estimator; standard errors for the standardized path coefficients are not provided with this estimator. Missing data were not imputed. All available information was used: a subject with a missing value for a given covariate was omitted when

estimating relationships involving this covariate, but available information for the same subject was used to estimate the other relationships.

Modifications of the hypothesized structural-equation model

Prior to the addition of the H1N1pdm09 infection outcome and of additional observed covariates, the first required step was the measurement model validation (Figure 1 in the manuscript). The measurement model had good fit indices (RMSEA=0.024, CFI=0.962). However, the factor loading of the indicator "time spent in public transport", as well as those of the average living room and bedroom temperatures, were not significant, and these indicators were therefore removed from the model (see Table 3 for the coefficients of the measurement model).

There were several differences between the final structural equation model (Figure 2 in the manuscript, RMSEA=0.023, CFI=0.943) and our hypothesized model (Figure 3, RMSEA=0.034, CFI=0.868). Using modifications indices, we added 2 regression paths: the regression of the latent variable of the compliance with preventive behaviors on the observed covariate "sex", and the regression of the 2010-11 pre-epidemic vaccination status on the 2009-10 pandemic vaccination status. We also added correlations between measurement error terms (measurement errors refer to the indicator's adequacy in measuring their underlying latent variable) for the daily duration and number of contacts for the three age classes of contacts considered (<15 years, 15-50 years and > 50 years).

Regarding the estimated coefficients, non-significant relations were removed. No association was observed between the age class and the pandemic vaccination status for 2009-10. Sex was not associated with vaccination for both the 2009-2010 and 2010-2011 seasons. Risk perception of H1N1pdm09 infection was not associated with the contact network latent variable. Latent variables describing the urban area and the neighborhood socio-economic status were not associated with 2010-2011 pandemic vaccination. Finally, four latent variables ("Indoor characteristics", "Urban environment", "Neighborhood socioeconomic status" and "Contact network") were not associated with the infection outcome, nor was the cumulative ILI incidence for 2010- 2011 season.

Additional analysis about duration and number of contacts

We further investigated the absence of a significant relationship between exposure through contact network and H1N1pdm09 infection (P=0.340). Exploration of the daily number and duration of contacts depending on subjects' age class and infection status (Figure 4) showed that subjects mainly reported contacts with individuals belonging to their age class. However, no significant differences (with Wilcoxon sum-rank tests) were observed between infected and non-infected subjects for such contacts. Non-infected subjects younger than 15 years of age had a higher daily number and duration of contacts with subjects 15 to 50 years old (P <0.05 and P <0.05) and with individuals over 50 years old (P <0.001 and P <0.001) than infected subjects. The same observation was made for the subjects aged between 15 and 50 years concerning the daily number and duration of contacts with individuals aged younger than 15 years (P <0.05 for the number of contacts and P <0.05 for the duration of contacts). We also observed that contact patterns were stable across the calendar seasons (Kruskal-Wallis rank sum test P for the daily duration of contacts=0.670 and Kruskal-Wallis rank sum test P for the daily number of contacts=0.39 – see Figure 5).

References

1 González J, De Boeck P, Tuerlinckx F. A double-structure structural equation model for three-mode data. *Psychol Methods* 2008;**13**:337–53. doi:10.1037/a0013269

2 Bollen KA, Noble MD. Structural equation models and the quantification of behavior. *Proc Natl Acad Sci U S A* 2011;**108 Suppl**:15639–46. doi:10.1073/pnas.1010661108

3 Kline RB. *Principles and practice of structural equation modeling*. Third. The Guilford Press 2010. doi:10.1038/156278a0

4 Byrne BM. *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge Academic New York 2011.

5 Geiser C. *Data Analysis with Mplus*. Guilford Press 2013.

6 Amorim LDAF, Fiaccone RL, Santos CAST, *et al.* Structural equation modeling in epidemiology. *Cad saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Esc Nac Saúde Pública* 2010;**26**:2251–62.

7 Stein CM, Morris NJ, Nock NL. Structural equation modeling. *Methods Mol Biol* 2012;**850**:495–512. doi:10.1007/978-1-61779-555-8_27

8 Beran TN, Violato C. Structural equation modeling in medical research: a primer. *BMC Res Notes* 2010;**3**:267. doi:10.1186/1756-0500-3-267

Tables

Table 1 Description of the observed covariates included in the structural equation model. All covariates are categorical except quantitative (q) and log-transformed quantitative covariates (l). (i) refers to covariates describing subjects' IRIS zone. * : for categorical covariates the number of classes is shown, for quantitative covariates mean and standard deviation range are shown.

Observed variable	Number of classes / Mean(SD)*	Missing values : N (%)
Prevention highly depends on behaviors	2 (agree/disagree)	168 (13%)
Some preventive measures can reduce infection risk	2 (agree/disagree)	157 (12%)
We can reduce infection risk by taking personal measures	2 (agree/disagree)	126 (9%)
H1N1pdm09 influenza is most often fatal	2 (agree/disagree)	165 (12%)
Mechanisms of the disease are not easily understood	2 (agree/disagree)	222 (17%)
H1N1pdm09 influenza has a severe impact on bodily functions	2 (agree/disagree)	288 (22%)
Pre-epidemic HAI titer (q) (l)	3.65 (0.66)	30 (2%)
Duration of contacts with individuals aged <15 years (q) (l) (min)	3.24 (4.70)	0 (0%)
Duration of contacts with individuals aged 15-50 years (q) (l) (min)	6.07 (1.61)	0 (0%)
Duration of contacts with individuals aged >50 years (q) (l) (min)	4.74 (2.89)	0 (0%)
Daily time spent in public transport (q) (l) (min)	-2.39 (4.08)	0 (0%)
Number of contacts with individuals aged <15 years (q) (l)	-0.46 (2.52)	0 (0%)
Number of contacts with individuals aged 15-50 years (q) (l)	1.38 (1.02)	0 (0%)

Number of contacts with individuals aged >50 years (q) (l)	0.39 (1.59)	0 (0%)
Always/often washes hands after coughing/sneezing	2 (yes/no)	23 (2%)
Daily frequency of hand washing ≧5	2 (yes/no)	37 (3%)
Always/often covers mouth while coughing/sneezing	2 (yes/no)	16 (1%)
Employment rate (%) among 15-65 years (q) (i)	63.63 (7.50)	0 (0%)
% of inhabitants >15 years without a diploma (q) (i)	12.87 (5.90)	0 (0%)
Mean annual income (q) (l) (i) (k€)	10.06 (0.28)	0 (0%)
Living room type of heat	3 (electric/gas/other)	145 (11%)
Bedroom type of heat	3 (electric/gas/other)	185 (14%)
Presence of air humidifier (bedroom)	2 (yes/no)	0 (0%)
Presence of air humidifier (living room)	2 (yes/no)	1 (1%)
Average bedroom temperature >19°C	2 (yes/no)	120 (9%)
Average living room temperature >20°C	2 (yes/no)	117 (9%)
Age class	3 (<15 years, 15-50 years, >50 years)	0 (0%)
Sex (male)	2 (yes/no)	0 (0%)
2009-10 pandemic vaccination	2 (yes/no)	2 (1%)
2010-11 pre-epidemic vaccination	2 (yes/no)	15 (1%)
History of ILI (2009-10)	2 (yes/no)	48 (4%)
Cumulative ILI incidence in subjects region (2009-10) (q)	5165 (1531)	0 (0%)
Cumulative ILI incidence in subjects region (2010-11) (q)	3391 (1152)	0 (0%)
Infection with H1N1pdm09 influenza	2 (yes/no)	197 (15%)

Table 2. Correlations between the covariates included in the measurement model. All covariates are categorical except quantitative (q) and log-transformed quantitative covariates (l). (i) refers to covariates describing subjects' IRIS zone. Matrix diagonal shows variances of quantitative covariates and proportions of ‘‘ones’’ observed for dichotomous categorical covariates. For categorical covariates with more than 2 modalities (covariates 24-25), the proportion of the first modality is shown.

Covariates	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Prevention highly depends on behaviors 1	83.68																												
Some preventive measures can reduce infection risk 2	0.41	94.61																											
We can reduce the infection risk by taking personal measures 3	0.48	0.71	95.68																										
H1N1pdm09 influenza is most often fatal 4	-0.05	0.08	0.1	10.27																									
Mechanisms of the disease are not easily understood 5	0.06	-0.09	-0.09	0.22	67.12																								
H1N1pdm09 influenza has a severe impact on bodily functions 6	0.1	0.22	0.19	0.56	0.24	53.41																							
Pre-epidemic HAI titer (q) (l) 7	0.01	0	0.03	-0.03	0.06	0	0.45																						
Duration of contacts with indiv. aged <15 years (q) (l) 8	0.05	0.08	-0.02	0.03	-0.14	-0.05	0.04	22.09																					
Duration of contacts with indiv. aged 15-50 years (q) (l) 9	0.06	-0.13	-0.08	-0.08	-0.15	-0.08	-0.04	0.3	2.62																				
Duration of contacts with indiv. aged >50 years (q) (l) 10	0.01	0.08	-0.06	-0.02	0.04	-0.04	0.01	-0.11	-0.09	8.36																			
Daily time spent in public transport (q) (l) 11	-0.03	0	-0.16	0	0.05	-0.08	0.1	0.05	0.07	0.03	16.67																		
Number of contacts with indiv. aged <15 years (q) (l) 12	0.08	0.08	0	0.02	-0.14	-0.05	0.06	0.98	0.29	-0.11	0.05	6.4																	
Number of contacts with indiv. aged 15-50 years (q) (l) 13	0.08	-0.06	-0.01	-0.1	-0.16	-0.08	-0.03	0.19	0.85	-0.05	0.09	0.19	1.05																
Number of contacts with indiv. aged >50 years (q) (l) 14	0.02	0.09	-0.01	-0.04	0.03	-0.03	-0.01	-0.12	-0.11	0.95	0.03	-0.12	-0.02	2.54															
Always/often washes hands after coughing/sneezing 15	0.07	0.13	0.15	0.22	0.05	0.26	-0.07	-0.11	0.01	0.02	-0.02	-0.14	0	0.04	32.2														
Daily frequency of hand washing >=5 16	0.03	0.2	0.1	0.02	0.15	0.07	-0.08	-0.19	-0.14	0.12	-0.01	-0.21	-0.04	0.15	0.42	63.15													
Always/often covers mouth while coughing/sneezing 17	0.03	0.11	0.05	0.08	0.08	0.02	-0.07	-0.3	-0.1	0.16	0.12	-0.32	0.05	0.2	0.4	0.41	84.18												
Lives in a urban area 18	-0.01	0.13	-0.13	0.03	0.02	0.01	0.12	-0.05	0	-0.02	0.01	-0.05	0.01	-0.07	-0.06	0.06	0	78.22											
Presence of an agricultural land near habitation 19	0.08	0.04	0.07	-0.22	-0.06	0.03	-0.06	0.07	0.02	0.06	-0.03	0.08	0.04	0.1	0.04	0.03	0.03	-0.62	50.99										
Presence of livestock near habitation 20	0.06	-0.2	0.19	0.13	-0.04	0.18	-0.05	0.07	-0.02	0.04	-0.02	0.07	-0.02	0.05	0.02	-0.05	0.03	-0.59	0.62	5.24									
% of employment among 15-65 years (q) (i) 21	0.06	-0.02	0.07	-0.02	-0.03	-0.08	-0.06	0.03	0.02	0.14	0.03	0.03	0.04	0.14	0.02	0.01	0.07	-0.48	0.31	0.37	56.25								
% of inhabitants > 15 years without a diploma (q) (i) 22	-0.12	0.07	0	0.13	0.12	0.05	0	0.01	-0.03	-0.16	-0.06	0.01	-0.07	-0.17	0.02	0.06	-0.09	0.23	-0.14	-0.16	-0.55	34.91							
Mean annual income (q)(l)(i) (k€) 23	0.16	0.22	0.19	0	-0.06	-0.05	-0.03	0.03	0.01	0.07	0.04	0.04	0.05	0.07	0	0	0.11	-0.04	-0.02	0.09	0.39	-0.43	0.08						
Living room type of heat 24	-0.14	-0.06	0.05	-0.11	0.04	0	-0.04	0.01	-0.03	0.07	0.01	0.01	-0.01	0.07	0.01	0.01	0.13	-0.14	0.13	0.07	0.02	0.04	-0.05	30.26					
Bedroom type of heat 25	-0.1	0.08	0	-0.09	0.04	0.01	0.02	0.01	-0.02	0.03	-0.01	0.02	0.01	0.05	-0.02	0.02	0.08	-0.11	0.16	0.14	-0.01	0.04	-0.07	0.9	32.74				
Presence of air humidifier (bedroom) 26	0.15	-0.08	0.04	-0.17	0.09	0.13	0.05	-0.07	-0.07	-0.08	-0.15	-0.07	-0.09	-0.07	0.1	0.03	-0.02	0.03	-0.07	-0.3	0.06	0.03	-0.05	0.1	0.16	6.83			
Presence of air humidifier (living room) 27	0.31	0.07	0.15	-0.2	0.21	0.34	-0.05	0.06	-0.01	-0.2	-0.14	0.07	0.04	-0.18	0.12	-0.04	-0.01	0.07	-0.09	-0.02	-0.07	0.19	-0.24	-0.01	0.11	0.76	4.7		
Average bedroom temperature >19°C 28	-0.07	0.04	-0.07	0.18	-0.03	0.09	-0.02	0.01	-0.04	-0.19	-0.03	0.01	-0.09	-0.2	-0.01	-0.08	-0.11	0.29	-0.35	-0.11	-0.14	0.13	0.12	-0.09	-0.08	0.08	-0.03	27.73	
Average living room temperature >20°C 29	-0.17	0.04	0.1	0.12	0.04	0.18	-0.01	0	-0.12	-0.11	-0.03	-0.03	-0.17	-0.12	0.04	0.08	0.04	-0.04	-0.08	-0.05	-0.03	0.13	-0.03	-0.03	-0.11	0.12	-0.09	0.48	25.63

Table 3. Coefficients of the SEM measurement model.

Latent variable	Indicator	Unstandardized coefficient	Standard error	pvalue	Standardized coefficient
Perception of preventive measures	Prevention highly depends on behaviors	0.606	0.115	<0.001	0.545
	Some preventive measures can reduce infection risk	0.878	0.203	<0.001	0.79
	We can reduce the infection risk by taking personal measures	1	NA	NA	0.899
Risk perception of H1N1pdm09 infection	H1N1pdm09 influenza is most often fatal	1	NA	NA	0.598
	Mechanisms of the disease are not easily understood	0.436	0.129	0.001	0.261
	H1N1pdm09 influenza has a severe impact on bodily functions	1.591	0.545	0.003	0.951
Host susceptibility	Pre-epidemic HAI titer	-1	NA	NA	-1
Contact network	Duration of contacts with indiv. aged <15 years	1	NA	NA	0.481
	Duration of contacts with indiv. aged 15-50 years	0.417	0.049	<0.001	0.547
	Duration of contacts with indiv. aged >50 years	-0.336	0.099	0.001	-0.259
	Number of contacts with indiv. aged <15 years	0.532	0.026	<0.001	0.484
	Number of contacts with indiv. aged 15-50 years	0.201	0.026	<0.001	0.427
	Number of contacts with indiv. aged >50 years	-0.224	0.05	<0.001	-0.314
Compliance with preventive	Always/often washes hands after	0.822	0.097	<0.001	0.583

behaviors	coughing/sneezing				
	Daily frequency of hand washing >=5	0.962	0.122	<0.001	0.663
	Always/often covers mouth while coughing/sneezing	1	NA	NA	0.683
Urban environment	Lives in a urban area	1	NA	NA	0.889
	Presence of an agricultural land near habitation	-0.789	0.114	<0.001	-0.701
	Presence of livestock near habitation	-0.848	0.134	<0.001	-0.754
Neighborhood socioeconomic status	% of employment among 15-65 years	1	NA	NA	0.786
	% of inhabitants > 15 years without a diploma	-0.753	0.057	<0.001	-0.742
	Mean annual income	0.024	0.003	<0.001	0.489
Indoor characteristics	Bedroom type of heat	1	NA	NA	0.98
	Living room type of heat	0.938	0.087	<0.001	0.919
	Presence of air humidifier (bedroom)	0.369	0.09	<0.001	0.362
	Presence of air humidifier (living room)	0.415	0.103	<0.001	0.406

Figures

Figure 1 Example of a path diagram. Ellipses: latent variables; boxes: observed variables.

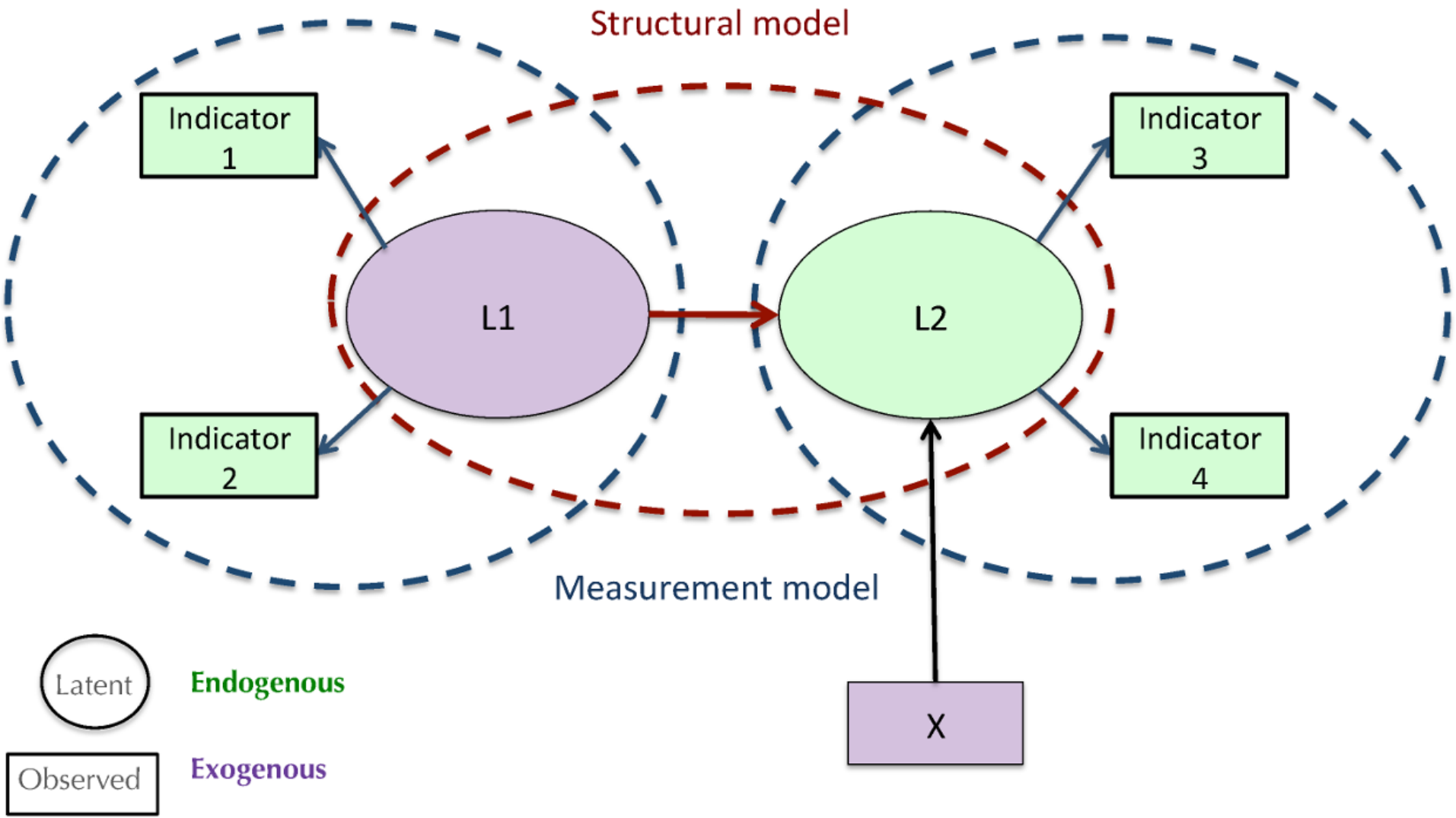

Structural model
Indicator 1
Indicator 2
L1
L2
Indicator 3
Indicator 4
Measurement model
X
Latent
Observed
Endogenous
Exogenous

Figure 2 Pre-epidemic immunity depending on subjects' infection status and age class. y-axis shows the proportion of subjects with pre-epidemic HAI titers above or equal to the HAI titers on the x-axis.

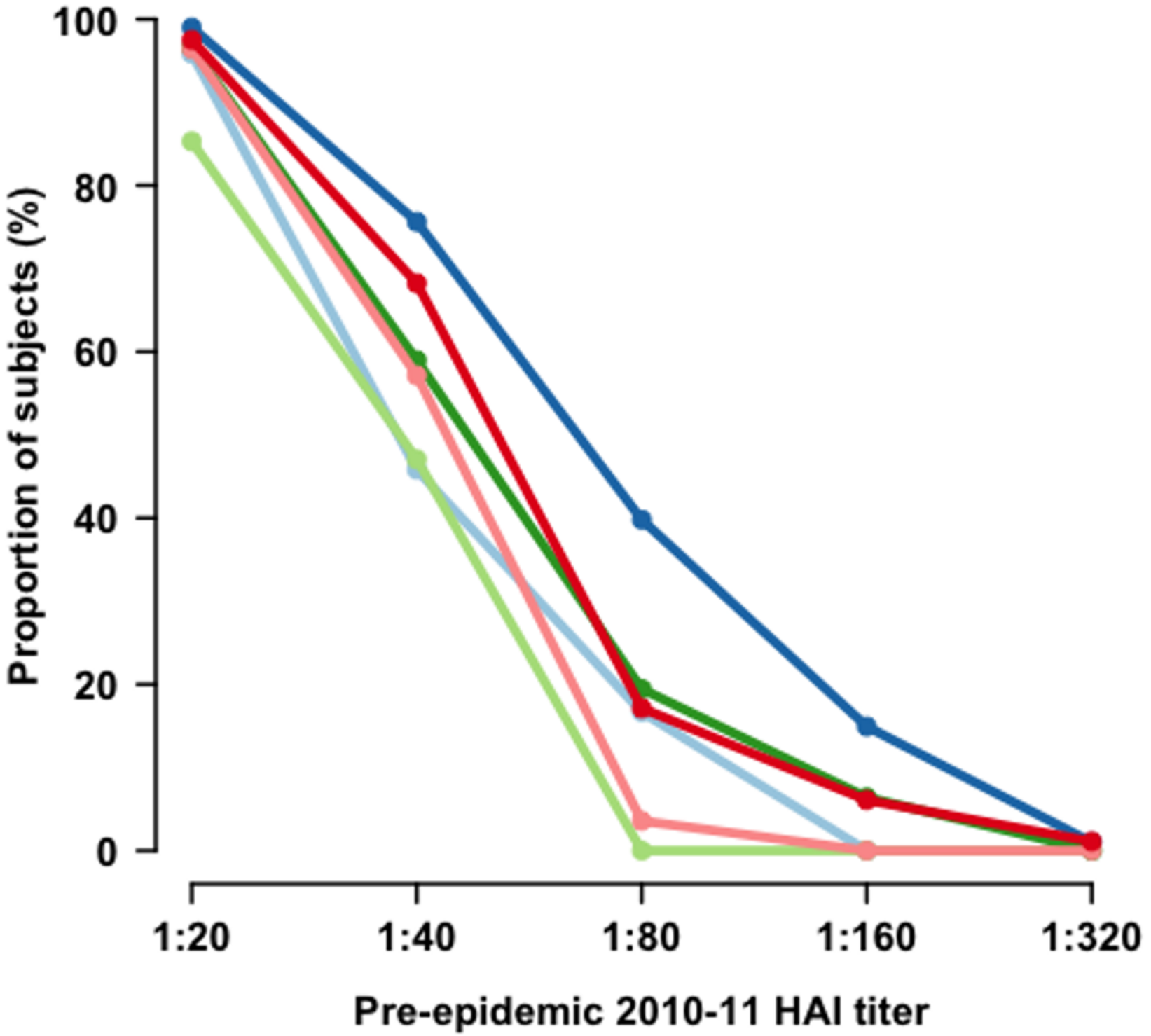

100
80
60
40
20
0
Proportion of subjects (%)
1:20
1:40
1:80
1:160
1:320
Pre-epidemic 2010-11 HAI titer
Legend
<15 years (infected)
<15 years (non-infected)
15-50 years (infected)
15-50 years (non-infected)
>50 years (infected)
>50 years (non-infected)

Figure 3 Hypothesized relationships between latent variables and exogenous/endogenous observed variables. Ellipses: latent variables; boxes: observed variables. Dotted background: latent variables related to exposure to H1N1pdm09. Striped background: latent variable related to susceptibility to H1N1pdm09 infection. For clarity, latent variables indicators are not shown.

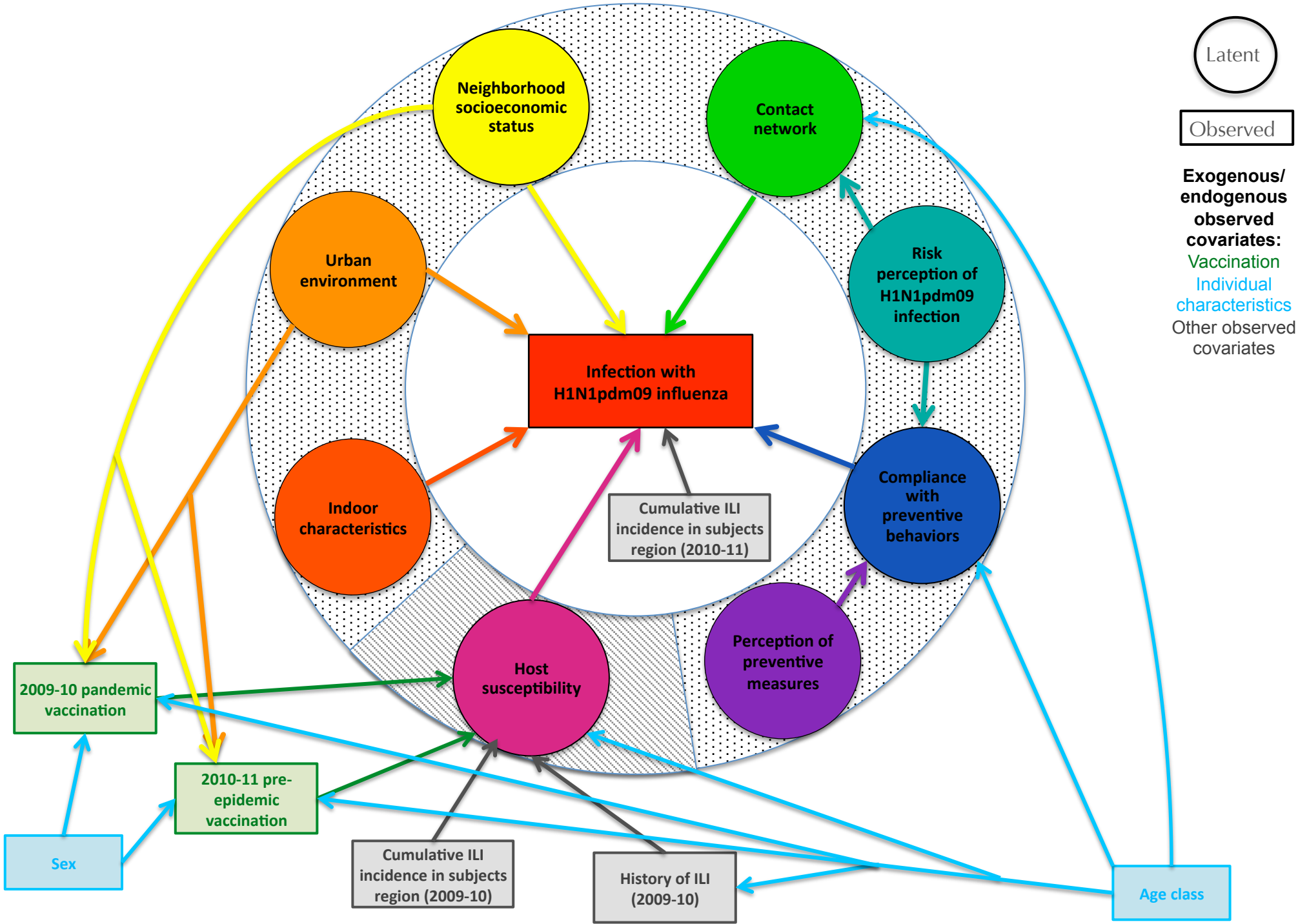

Latent
Observed
Exogenous/ endogenous observed covariates:
Vaccination
Individual characteristics
Other observed covariates
Neighborhood socioeconomic status
Contact network
Urban environment
Risk perception of H1N1pdm09 infection
Infection with H1N1pdm09 influenza
Indoor characteristics
Cumulative ILI incidence in subjects region (2010-11)
Compliance with preventive behaviors
Host susceptibility
Perception of preventive measures
2009-10 pandemic vaccination
2010-11 pre-epidemic vaccination
Sex
Cumulative ILI incidence in subjects region (2009-10)
History of ILI (2009-10)
Age class

Figure 4 Contact patterns depending on subjects' infection status and age class. y-axis shows the proportion of subjects with daily number/duration of contacts above or equal to the daily number/duration of contacts on the x-axis.

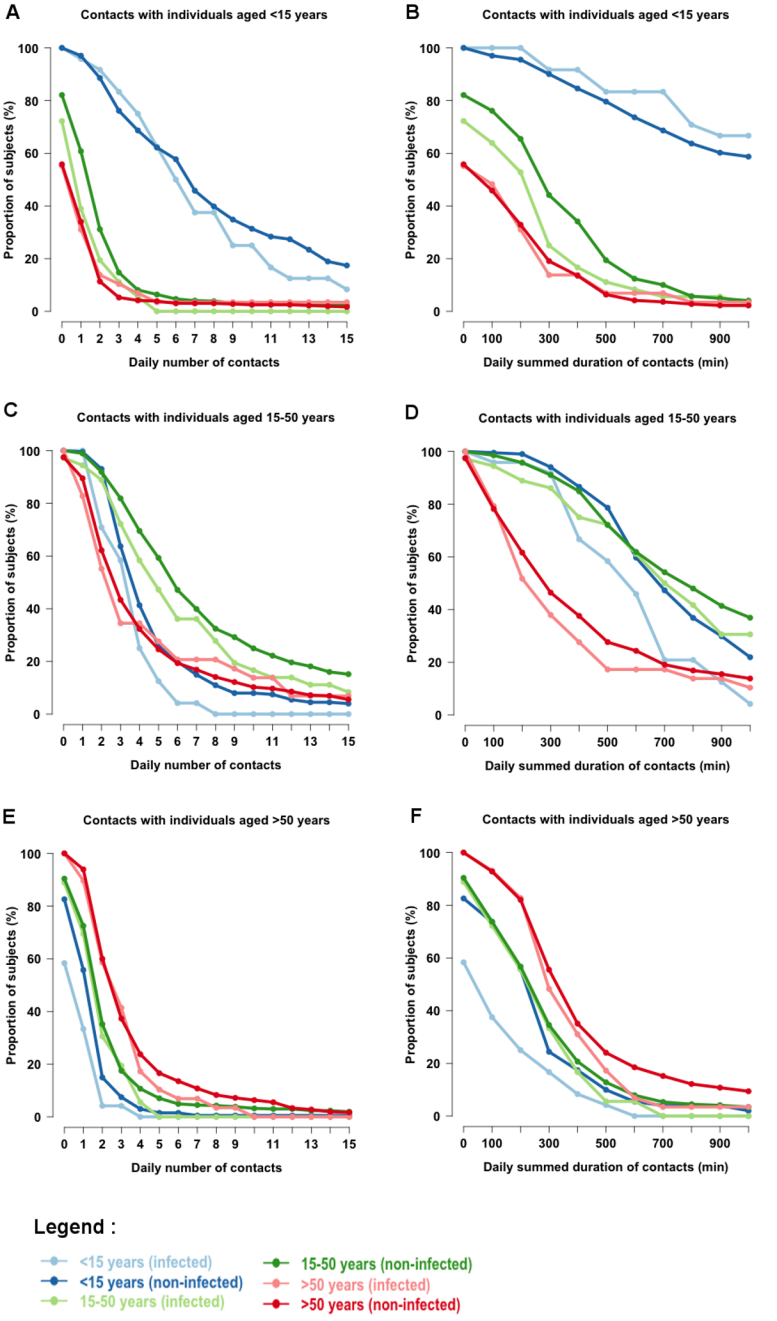
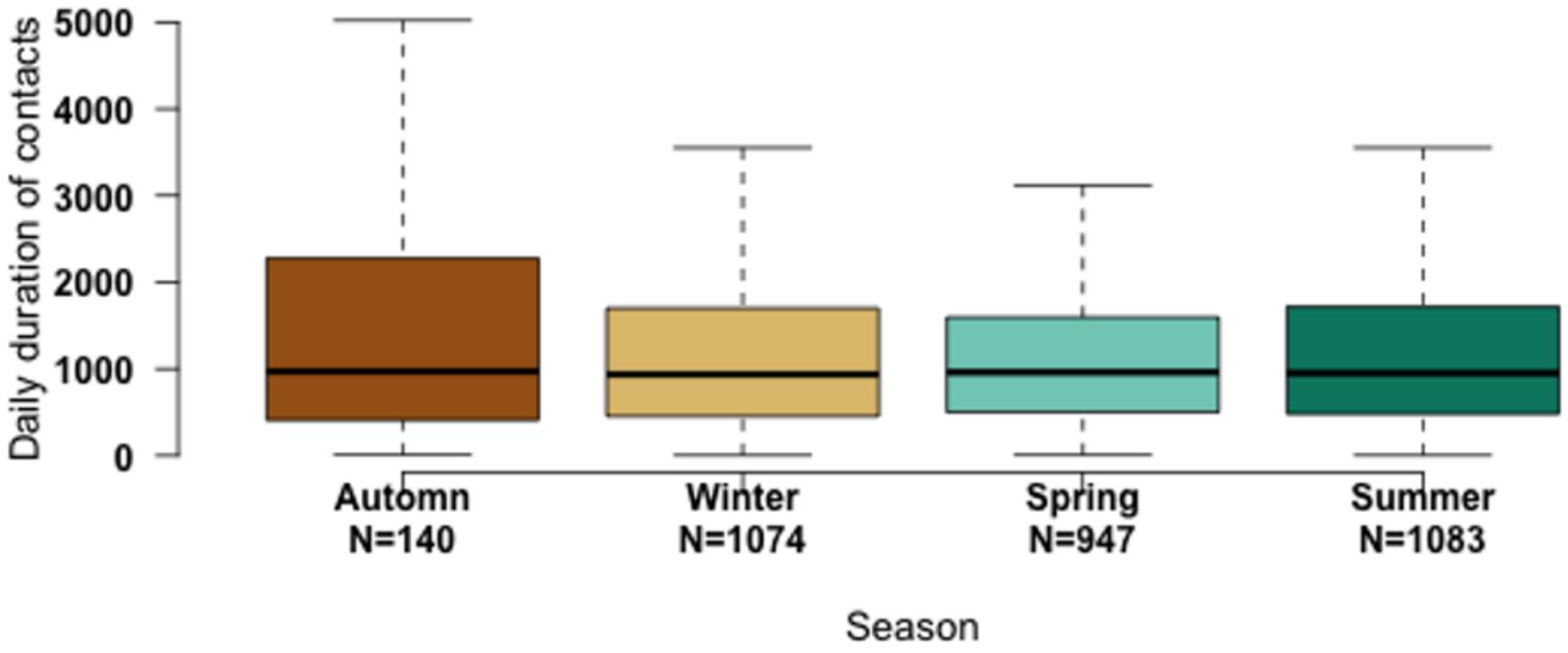
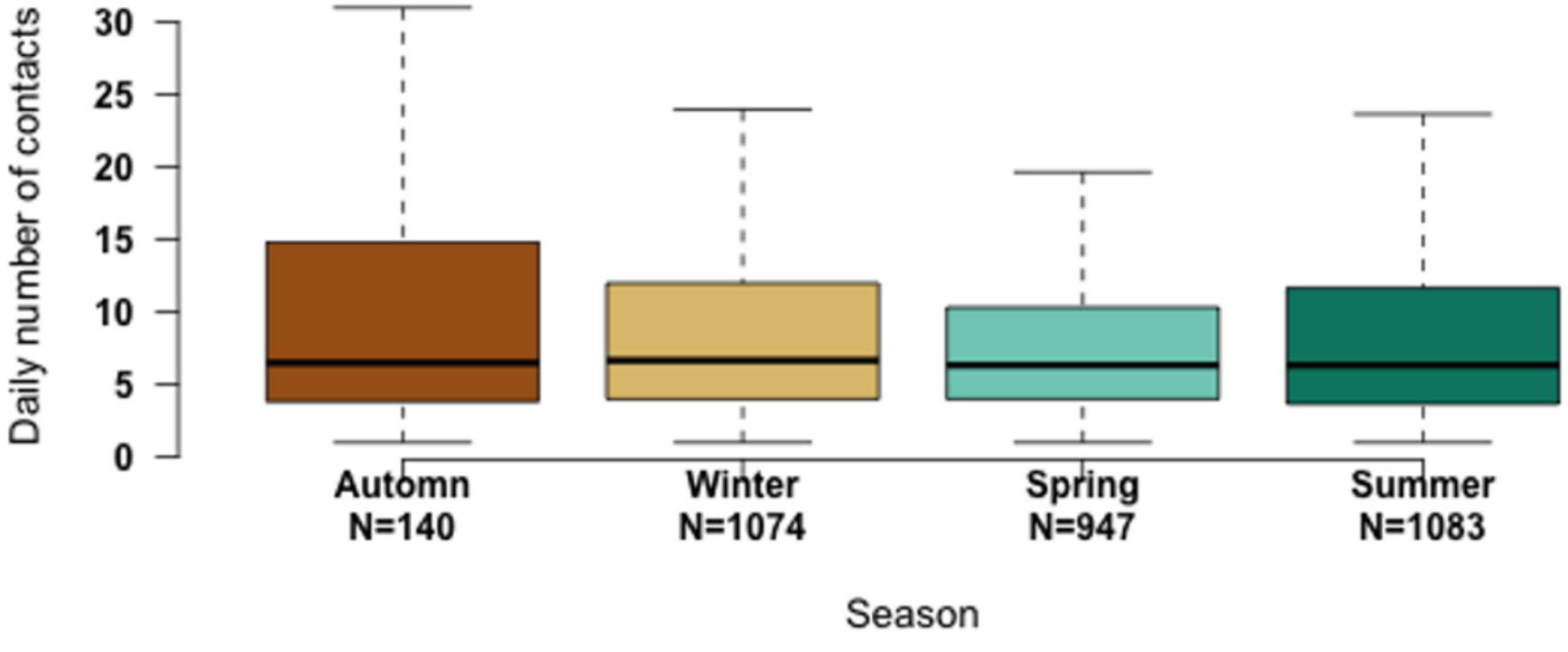

A
Contacts with individuals aged <15 years
Proportion of subjects (%)
Daily number of contacts
B
Contacts with individuals aged <15 years
Proportion of subjects (%)
Daily summed duration of contacts (min)
C
Contacts with individuals aged 15-50 years
Proportion of subjects (%)
Daily number of contacts
D
Contacts with individuals aged 15-50 years
Proportion of subjects (%)
Daily summed duration of contacts (min)
E
Contacts with individuals aged >50 years
Proportion of subjects (%)
Daily number of contacts
F
Contacts with individuals aged >50 years
Proportion of subjects (%)
Daily summed duration of contacts (min)
Legend :
<15 years (infected)
<15 years (non-infected)
15-50 years (infected)
15-50 years (non-infected)
>50 years (infected)
>50 years (non-infected)

Figure 5 Boxplots of the daily duration and number of contacts depending on calendar seasons. N corresponds to the number of subjects reporting contacts for the considered calendar season.


Daily duration of contacts
5000
4000
3000
2000
1000
0
Automn
N=140
Winter
N=1074
Spring
N=947
Summer
N=1083
Season


Daily number of contacts
30
25
20
15
10
5
0
Automn
N=140
Winter
N=1074
Spring
N=947
Summer
N=1083
Season