Epidemics 13 (2015) 1–9

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Estimating dynamic transmission model parameters for seasonal influenza by fitting to age and season-specific influenza-like illness incidence

Nele Goeyvaerts[a,b,*,1], Lander Willem[a,b,c,1], Kim Van Kerckhove[a,b], Yannick Vandendijck[a], Germaine Hanquet[d], Philippe Beutels[b], Niel Hens[a,b]

[a] Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Agoralaan Gebouw D, B3590 Diepenbeek, Belgium
[b] Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Universiteitsplein 1, B2610 Wilrijk, Belgium
[c] Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B2020 Antwerp, Belgium
[d] KCE – Belgian Health Care Knowledge Centre, Boulevard du Jardin Botanique 55, B1000 Brussels, Belgium

ARTICLE INFO

Article history:
Received 3 June 2014
Received in revised form 10 April 2015
Accepted 24 April 2015
Available online 2 May 2015

Keywords:
Influenza
Mathematical model
Parameter estimation
Reproduction number
Seasonal variability

ABSTRACT

Dynamic transmission models are essential to design and evaluate control strategies for airborne infections. Our objective was to develop a dynamic transmission model for seasonal influenza allowing to evaluate the impact of vaccinating specific age groups on the incidence of infection, disease and mortality. Projections based on such models heavily rely on assumed 'input' parameter values. In previous seasonal influenza models, these parameter values were commonly chosen ad hoc, ignoring between-season variability and without formal model validation or sensitivity analyses. We propose to directly estimate the parameters by fitting the model to age-specific influenza-like illness (ILI) incidence data over multiple influenza seasons. We used a weighted least squares (WLS) criterion to assess model fit and applied our method to Belgian ILI data over six influenza seasons. After exploring parameter importance using symbolic regression, we evaluated a set of candidate models of differing complexity according to the number of season-specific parameters. The transmission parameters (average R_0, seasonal amplitude and timing of the seasonal peak), waning rates and the scale factor used for WLS optimization, influenced the fit to the observed ILI incidence the most. Our results demonstrate the importance of between-season variability in influenza transmission and our estimates are in line with the classification of influenza seasons according to intensity and vaccine matching.

1. Introduction

Influenza presents as a mild disease in most healthy adults but is responsible for significant morbidity and mortality among vulnerable groups such as the elderly, patients with underlying health conditions and children. Recently, several countries have introduced routine vaccination of children for two main reasons. The first is that the rate of influenza hospitalizations in young children is as high as in elderly. The second is that children play an important role in influenza virus transmission, so that childhood vaccination would provide indirect protective effects for the community.

We developed a dynamic transmission model for seasonal influenza with the aim to enable projecting the effectiveness and cost-effectiveness of mass vaccination strategies (Beutels et al., 2013). In this paper, we focus on the dynamic transmission model and estimate key parameters by fitting the model to an age-specific time series of influenza-like illness (ILI) incidence. The model relies on data from Belgium, such as ILI incidence, vaccination coverage and demographic data, though the concepts and methods are generally applicable. In Belgium, influenza vaccines are currently recommended for people over 50 years (with priority for those over 65 years), people with underlying chronic illness, pregnant women and health care workers (Beutels et al., 2013). In order to gain insights into the existing modelling approaches, we conducted a thorough literature review (details in Supplementary material), which yielded 25 articles presenting dynamic transmission

* Corresponding author at: Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Agoralaan Gebouw D, B3590 Diepenbeek, Belgium. Tel.: +32 11 268294.
E-mail address: nele.goeyvaerts@uhasselt.be (N. Goeyvaerts).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.epidem.2015.04.002

models for seasonal influenza. We classified these articles in four main groups based on the modelling approach used.

First, there are standard mathematical models such as SIR (susceptible–infectious–recovered) and compartmental extensions thereof that are mainly designed to capture single epidemics (e.g. Glasser et al., 2010). Second, another group of models extends the previous class by including adaptive parameters for seasonality such as seasonally forced transmission rates (e.g. Finkenstadt et al., 2005; Vynnycky et al., 2008). Influenza occurs in annual epidemics during the winter period, which has been related to many factors e.g. temperature and humidity, viral production, and contact patterns (Fuhrmann, 2010; Shaman and Kohn, 2009; Willem et al., 2012). Third, some narrative reviews focused on the comparison of various dynamic models (e.g. Ballesteros et al., 2009). Fourth, a final group focused on multi-strain models to evaluate the impact of cross-immunity between different influenza strains by means of theoretical derivations or simulations (e.g. Andreasen, 2003; Prosper et al., 2011).

We used the dynamic model of Vynnycky et al. (2008) as a basis for our model because it is an age-stratified model with seasonally forced transmission rates, including annual vaccination. As influenza is mainly spread from person to person through respiratory droplets, transmission depends directly on age-specific rates of making social contact. Over the last decade, important advances were made in the collection of social contact data to parameterize infectious disease transmission models, such as the large population-based survey conducted in eight European countries as part of the POLYMOD project (Mossong et al., 2008). The use of empirical observations to inform the 'who acquires infection from whom' matrix has been successfully applied to model the transmission of different airborne infections (Goeyvaerts et al., 2010; Kretzschmar et al., 2010; Ogunjimi et al., 2009; Wallinga et al., 2006). Vynnycky et al. (2008) were the first to use POLYMOD contact data to parameterize transmission rates for seasonal influenza.

In this paper, we propose to directly estimate the dynamic model parameters by fitting the model to multi-season ILI incidence. In many countries, ILI incidence is monitored via surveillance systems such as sentinel networks or online surveys (Vandendijck et al., 2013). We capture between-season heterogeneity by allowing for season-specific parameters, such that the associated uncertainty can be propagated in predictions for future epidemics and the evaluation of vaccination strategies. By contrasting model predictions against relevant incidence data, our approach improves upon the practice of imputing pre-specified values to uncertain parameters, such as transmission or waning rates, without formal model validation or sensitivity analyses. Parameter values for dynamic models of seasonal influenza were commonly chosen ad hoc or based on inadequate data, e.g. related to historical observations or pandemic influenza.

There are few examples of dynamic transmission models for seasonal influenza that were actually fitted to incidence data. Hsieh (2010) estimated age-specific transmission probabilities by fitting to cumulative pneumonia and influenza mortality data from a single low-intensity season, however, the fit to the crude mortality data seems dubious. Finkenstadt et al. (2005) developed a stochastic model to estimate the rate of antigenic drift from multi-season ILI incidence data. Their results supported the presence of immunity loss, which we will account for by assuming continuous waning after infection or vaccination. Their dynamic model as such was less relevant to our setting as it ignored annual vaccination and age-dependent transmission. The same was true for the deterministic model by Poletti et al. (2011), which was fitted to ILI incidence data from a single season using a least-squares approach and designed to estimate the impact of behavioural changes during the 2009 H1N1 pandemic. We elaborate on their least-squares method when fitting our model to multi-season ILI incidence data, including age-specific post-stratification weights.

Pitman et al. (2013, 2012) used a model similar to Vynnycky et al. (2008) to evaluate childhood vaccination in England and Wales. In both studies (Vynnycky et al., 2008; Pitman et al., 2012), parameter values were imputed rather than estimated by fitting the model to epidemiological data. However, during preparation of the current paper, Baguelin et al. (2013) presented another model with the same purpose, using an MCMC-based method to estimate model parameters from ILI incidence data. We briefly discuss the main differences between our approaches at the end of this paper.

2. Materials and methods

In this section, we first describe the ILI incidence data after which the dynamic transmission model is introduced, outlining its general structure, the interpretation of the parameters and the model assumptions. At the end of this section, we introduce the statistical methods used to estimate the parameters in the dynamic transmission model.

2.1. ILI incidence data

The ILI incidence data were collected from a sentinel network of general practitioners (GPs) in Belgium coordinated by the Scientific Institute of Public health. Each week, the GPs report the total number of ILI consultations by four age groups, 0–4, 5–14, 15–64 and ≥65 years, using the following case definition: sudden onset of symptoms, high fever, respiratory (i.e. cough, sore throat) and systemic symptoms (i.e. headache and muscular pain). Weekly estimates for the denominators are available as well, reflecting the population covered by the sentinel network. We analyze data from the last six pre-pandemic influenza seasons, from October 2003 to August 2009, as displayed in Fig. 3. The ILI incidence data are described in more detail by Hanquet et al. (2011) and Bollaerts et al. (2013).

Additionally, a fraction of ILI patients were swabbed by the GPs in the sentinel network during the influenza activity period and tested for influenza A and B. The swabbing of ILI cases was carried out 'ad hoc' using quota for each sentinel GP, but without any objective or consistent criteria, and was therefore likely age biased. Due to the non-systematic nature of the swabbing, the weekly laboratory-confirmed influenza case data were incomplete, especially in young children and individuals over 65 years of age. Further, there were few cases of influenza B in Belgium during the study period. We therefore fitted the dynamic model to ILI incidence data rather than to influenza-confirmed ILI incidence data, and calibrated model-based outcomes for the economic evaluation (Beutels et al., 2013).

We thus assume one generic influenza virus, which should be interpreted as an average of past influenza A and B strains. Therefore in our model, waning of immunity may capture actual waning of acquired immunity as well as lack of cross protection. The model partly captures season dominance of A and B strains and transmission heterogeneity by incorporating season-specific parameters.

2.2. Dynamic transmission model

2.2.1. General structure of the model

We elaborate on the model Vynnycky et al. (2008) and use the same notation. This is an age-stratified SEIRS model with vaccination, classifying the population into compartments of susceptible ($S_a(t)$), exposed ($E_a(t)$; infected but not yet infectious), infectious ($I_a(t)$), recovered ($R_a(t)$) and vaccinated ($V_a(t)$) individuals, as displayed in Fig. 1. Both recovered and vaccinated individuals are

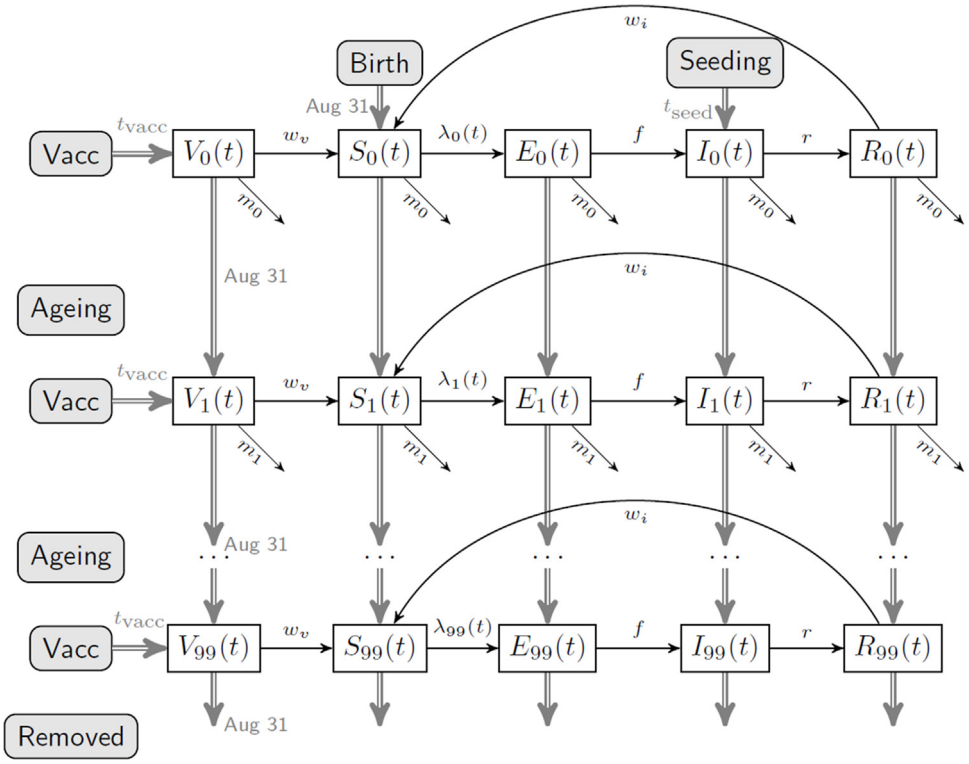

Fig. 1. Age-stratified SEIRS model with vaccination: single black arrows indicate time-continuous transitions, while double grey arrows indicate instantaneous transitions. Arrows indicating influx compartments for vaccination and seeding are suppressed from display.

assumed fully protected after infection and vaccination, respectively, until their immunity wanes. The population is stratified into age classes a of length 1 year: [0, 1 [, …, [99, 100 [years, and we assume demographic equilibrium. Belgian demographic data from 2009 obtained from Eurostat (2011) are used to determine the initial age-specific population distribution N_a and to estimate an age-specific daily mortality rate m_a. We use time steps of 1 day to ensure high precision while maintaining computational feasibility. Short definitions of the parameters in Fig. 1 are provided in Table 1 and further explained below. The corresponding set of ordinary differential equations is given in the Supplementary material.

Each year, there are three time points of transition. We assume a realistic age-structured (RAS) model in which all individuals move to the next age group on August 31 of each year (Vynnycky and White, 2010). Individuals in the final age group (99 years of age) are removed from the population, and as many newborns as there were deaths in the preceding year, are introduced. Therefore, the total population size remains constant. Further, at time point t_{vacc}, any proportion of any age group may receive influenza vaccination, irrespective of their disease or vaccination history. Finally, each year at time point t_{seed}, a fraction p_{seed} of the susceptible population within age group $[a_{1,\text{seed}}; a_{2,\text{seed}}]$ are seeded into the population as newly infectious individuals, to establish a new influenza epidemic. Note that Vynnycky et al. (2008) assume the above transition points in relation to ageing, vaccination and seeding, all occur at the same time on August 31.

The mass action principle relates the force of infection, $\lambda_a(t)$, to seasonally forced age-specific transmission rates:

$$\lambda_a(t) = z(t) \sum_{a'} \beta_{a,a'} I_{a'}(t), \tag{1}$$

where $\beta_{a,a'}$ denotes the average daily per capita rate at which an individual of age a' makes effective contact with a person of age a, and $z(t)$ denotes a sinusoidal seasonality function (Vynnycky et al., 2008):

$$z(t) = 1 + \delta \sin\left(\frac{2\pi(t - t_0)}{365}\right), \tag{2}$$

reflecting the relative change of the basic reproduction number at time t, $R_0(t)$, from the average basic reproduction number, $\bar{R}_0$, measured at reference time t_0. The seasonal peak of

Table 1
Short definitions of the parameters in the age-stratified SEIRS model with vaccination.

Parameter	Definition
$\bar{R}_0$	Average basic reproduction number measured at reference time t_0
δ	Amplitude of the seasonality function $z(t)$ defined in (2). δ determines the peak value of the basic reproduction number ($0 \le \delta \le 1$)
t_0	Reference time for the seasonality function $z(t)$ defined in (2), at which the basic reproduction number equals $\bar{R}_0$. The seasonal peak of transmission is 3 months later
t_{vacc}	Time point of vaccination
t_{seed}	Time point at which infectious individuals are seeded into the population
$a_{1,\text{seed}} - a_{2,\text{seed}}$	Age group targeted for seeding infectious individuals into the population
p_{seed}	Proportion of susceptibles seeded into the population as infectious individuals
f	Daily rate at which exposed individuals become infectious
r	Daily rate at which infectious individuals recover and become immune
w_i	Yearly rate at which naturally infected individuals lose immunity
w_v	Yearly rate at which vaccinated individuals lose immunity (assumed equal to w_i)
α	Scale factor to calibrate model-based infection incidence to observed ILI incidence

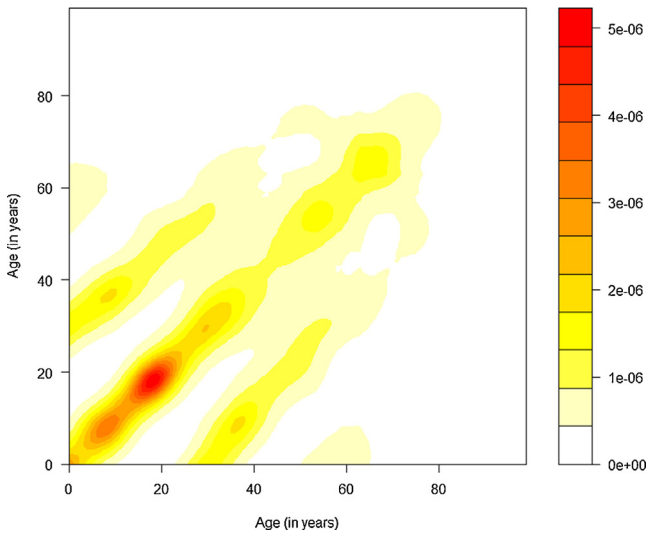

Fig. 2. Age-specific per capita rates of physical contact > 15 min, estimated from the Belgian POLYMOD data using a smooth-then-constrain approach.

transmission occurs three months after the reference time t_0. The amplitude parameter δ is bounded $0 \leq \delta \leq 1$ to ensure that $z(t) \geq 0, \forall t$. $\bar{R}_0$ is calculated as the dominant eigenvalue of the next generation matrix with elements $N_a \beta_{a,a'}/r$, and $R_0(t) = \bar{R}_0 z(t)$.

2.2.2. Social contact hypothesis

We assume that age-specific transmission rates are directly proportional to age-specific rates of making social contact, $\beta_{a,a'} = qc_{a,a'}$, where q is a proportionality constant directly related to the value of the average basic reproduction number, $\bar{R}_0$. This is the so-called "social contact hypothesis" introduced by Wallinga et al. (2006), who were the first to augment seroprevalence data with data on conversational contacts to estimate transmission rates for airborne infections. We estimate the age-specific contact rates, $c_{a,a'}$, from the Belgian POLYMOD contact survey conducted in 2006 (Mossong et al., 2008), assuming that contacts involving physical skin-to-skin touching and taking longer than 15 min are a good proxy for those events through which influenza transmission may occur. Previous modelling work revealed that this type of contact fits well the observed seroprevalence profiles for airborne infections such as varicella zoster virus and parvovirus B19 (Goeyvaerts et al., 2010, 2011; Ogunjimi et al., 2009; Melegaro et al., 2011).

The age-specific per capita contact rates, $c_{a,a'}$, are estimated using the smooth-then-constrain approach described in Goeyvaerts et al. (2010). We fit a negative binomial generalized additive model to the number of reported contacts with age classes of length 1 year, using thin plate regression splines to model the mean as a two-dimensional flexible function of age. Note that this entails a continuous contact surface and that the spline basis dimension (taken to be 11) does not correspond to the contact matrix dimension. Subsequently, using age-specific population sizes obtained from demographic data, the estimated contact surface is constrained to account for reciprocity (Wallinga et al., 2006). To reduce boundary effects due to contact data sparseness for the elderly, the contact rates for individuals of age ≥86 years are based on those of age 85 years. The estimated contact rates are displayed in Fig. 2, revealing a highly assortative mixing pattern.

2.2.3. Waning immunity

Genetic variation of the virus produces antigenic novel strains at such a high rate that most people who have had influenza or were vaccinated, are susceptible to a new circulating strain within a few years or even months after infection or vaccination (Earn and Levin, 2002; Kissling et al., 2013). In our model, this process is partially captured by allowing for waning immunity after natural infection or vaccination. The two waning rates, w_i and w_v, are assumed to be equal and age-independent, in the absence of consistent empirical evidence. For example, the literature review by Skowronski et al. (2008) does not support the historic concern that vaccine-induced antibodies wane more quickly in the elderly compared to the young.

2.2.4. Vaccination coverage and vaccine efficacy

In Belgium, seasonal influenza vaccines are currently recommended for people over 50 years and for high risk groups (Beutels et al., 2013). We use the following age-stratified vaccination coverage estimates obtained from the Belgian Health Interview Survey of 2008 conducted by the Scientific Institute of Public Health: 0.066% for 6 months – 17 years (arising from a 1% coverage in children with co-morbidities), 11% for 18–49 years, 28% for 50–64 years, 50% for 65–74 years, and 71% for ≥75 years (Beutels et al., 2013; Hanquet et al., 2011).

Two main types of influenza vaccines are currently registered in Europe: the trivalent inactivated influenza vaccine (TIV), which is injectable, and the live attenuated influenza vaccine (LAIV), which is given as a nasal spray. LAIV vaccine was not authorized in the European Union until 2011 and was not on the Belgian market during the study time period (European Commission, 2010). Therefore, for the purpose of the current paper, only TIV vaccine efficacy estimates are relevant (see below).

Our model assumes an all-or-none vaccine effect, which means that the vaccine effectively protects a fixed proportion of vaccinated persons, i.e. providing complete immunity against infection, while it completely fails in the remaining part. The 'effective vaccination coverage' is consequently the product of the vaccination coverage and the vaccine efficacy for susceptibility (VE_S) and determines the fraction of the population that moves to the vaccinated state each year. Vaccine efficacy for infectiousness (VE_I) for TIV was found to be non-significant based on experimental challenge studies in seronegative adults (Basta et al., 2008). Thus, there is no evidence that vaccinated infected individuals are less infectious compared with unvaccinated infected individuals. In view of these findings, we prefer applying an all-or-none vaccine instead of a leaky vaccine model.

VE_S estimates for TIV are only available from experimental challenge studies in seronegative adults (Basta et al., 2008). Therefore, we use estimates of vaccine efficacy for influenza-confirmed ILI (VE for susceptibility to disease, denoted VE_{SP} (Halloran et al., 2010)) obtained from randomized controlled trials and observational studies, as a proxy for VE_S. These studies cover a broader age range and allow us to stratify VE estimates by both age and type of season, the latter according to influenza intensity (high-medium versus low intensity) and degree of matching between the vaccine and circulating viral strains (good-relative versus poor match). We use the VE_{SP} values for TIV shown in Table 2, that are literature-based estimates obtained from influenza-confirmed ILI cases, i.e. based on culture and/or PCR. The literature review and computational details are described in the Supplementary Material.

2.3. Parameter estimation

2.3.1. Weighted least squares

Vynnycky et al. (2008) impute pre-specified values, either literature-based or chosen "ad hoc", to all model parameters presented in Table 1 (note that they do not consider α, in the absence of model fitting). For example, their base-case value of $\bar{R}_0$ is 1.8, the best-fitting estimate for the 1957 influenza pandemic in the UK (Vynnycky and Edmunds, 2008). The parameters in Table 1 are uncertain and are likely to vary between populations and over time. Therefore, we propose to estimate these parameters from reported

Table 2
Literature-based VE_{SP} estimates for TIV per age group, season intensity and vaccine match. Classification of the influenza seasons is based on data from Belgium (Hanquet et al., 2011).

Age group	Intensity: high-medium Match: good-relative Seasons: 04–05, 06–07, 08–09	Intensity: high-medium Match: poor Seasons: 03–04
6 months–17 years	65%	48%
18–64 years	65%	60%
≥65 years	60%	55%

Age group	Intensity: low Match: good-relative Seasons: 07–08	Intensity: low Match: poor Seasons: 05–06
6 months–17 years	30%	16%
18–64 years	45%	22%
≥65 years	42%	20%

numbers of ILI cases. Other parameters that are well recorded, such as demographic and vaccination parameters, are included in the model as fixed values.

We estimate the model parameters using a weighted least squares (WLS) approach. Let $C_a(w_k)$ denote the number of reported ILI cases of age a in calendar week k, w_k, and let $P_a(w_k)$ denote the corresponding denominator, i.e. the number of individuals of age a covered by the sentinel network in calendar week k. The observed age-specific ILI incidence rate in calendar week k is then calculated as follows: $Y_a(w_k) = C_a(w_k)/P_a(w_k)$. To simplify notation, we suppress the dependency of the model outcome on the input parameters. Let $I_a^*(t)$ denote the number of newly infectious individuals of age a at time t, and let $N_a(t)$ denote the total number of individuals of age a at time t, as predicted by the model. The model-based age-specific incidence rate in calendar week k then equals:

$$Z_a(w_k) = \frac{\sum_{t\in w_k} I_a^*(t)}{\frac{1}{7}\sum_{t\in w_k} N_a(t)}.$$

We estimate the model parameters by minimizing the weighted sum of squared differences between the observed ILI incidence rate and the scaled model-based incidence rate:

$$\sum_{i=1}^{4}\sum_{k} v_{a_i}(w_k)\left[Y_{a_i}(w_k) - \alpha Z_{a_i}(w_k)\right]^2, \tag{3}$$

where the weighted sum is taken over all weekly ILI observations, from week 40 in 2003 to week 35 in 2009, per age group a_i: 0–4, 5–14, 15–64 and ≥65 years. The weights $v_{a_i}(w_k)$ are proportional to the corresponding denominator $P_{a_i}(w_k)$ and account for the unequal population sizes represented by the different age groups. The scale factor α calibrates the model-based incidence rate to the observed ILI incidence rate. This factor may absorb several effects, such as the probability for an infected individual to show symptoms, the GP consultation rate, i.e. the probability for a symptomatic infected to consult a GP, and the ILI reporting rate, i.e. the probability for a GP to report a symptomatic influenza case as ILI.

The WLS (3) is a direct measure of goodness-of-fit, with smaller values indicating a better fit to the ILI incidence data. The score is penalized such that models predicting more than 10 new cases per week during at least 10 weeks outside the influenza season from calendar week 40 to 20 are discarded.

2.3.2. Variable selection and optimization

We implemented the dynamic transmission model in Matlab. Starting from a completely susceptible population, the model is pre-run over a burn-in period of five influenza seasons to generate background immunity due to historical infection or vaccination. Our analysis consists of two stages. First, we explore the parameter space and identify influential parameters by sampling from a Latin hypercube design and by performing feature selection using Pareto-aware symbolic regression (SR) (Willem et al., 2014). Second, based on the results from the feature selection, we define a set of nested 'candidate' models with decreasing parameter complexity and obtain parameter estimates using an optimization algorithm.

Using a Latin hypercube design, we sample parameter combinations from a 12-dimensional parameter space, assuming all parameters are constant across seasons (parameters in Table 1 and parameter ranges in Supplementary material). Each parameter combination is then used to run the dynamic model. Using Pareto-aware SR, we analyze the relationship between the parameter values and the WLS obtained for the observed ILI incidence data.

SR aims to capture input-response behaviour with algebraic expressions, without a priori assumptions of model structure (Smits and Kotanchek, 2005; Vladislavleva, 2008). First, input variables are randomly combined into expressions using a predefined set of mathematical operators. Second, the resulting expressions are scored by their complexity (i.e. number of parameters and type of operators involved) and prediction error $1 - R^2$, with R the Pearson correlation coefficient between the observed and predicted responses (here: WLS). Both objectives are minimized and only the best scoring expressions are retained. Third, the remaining expressions are combined or adapted randomly. Next, the expressions are again scored, selected and so on. This evolutionary process is repeated over many generations to obtain an ensemble of expressions with low prediction error. The complexity objective avoids excessive growth of the expressions and the presence of a parameter in a sufficiently evolved population indicates the variable importance (Stijven et al., 2011). We use the SR algorithm from the DataModeler package in Mathematica (DataModeler, 2011).

Based on the feature selection, we define a set of nested models with decreasing parameter complexity. For each 'candidate' model, we use the GlobalSearch algorithm from the Matlab Optimization Toolbox to estimate the parameters. GlobalSearch initiates a gradient-based local solver (fmincon) from multiple starting points. Parameter constraints are taken into account when generating initial points and running the optimization algorithm (Supplementary material). We repeat this optimization process for 1000 different random number streams and select the 'best' set of parameter estimates i.e. corresponding to the lowest WLS value.

3. Results

3.1. Symbolic regression and variable selection

We generated 50,000 parameter combinations using a Latin hypercube design and selected the subset corresponding to the 30% lowest values of WLS (a trade-off between information and noise) for further analysis and feature selection using symbolic regression. The relative parameter presence in the model ensemble obtained from ten independent evolutions, indicated the following main parameters driving goodness-of-fit to the ILI incidence data (table in the Supplementary material): the average basic reproduction number R_0, the scale factor α and the waning rate $w_i = w_v$ (present in >80% of expressions) followed by the amplitude δ and the reference time point t_0 (presence ≈40%). The following parameters were found less important in explaining the WLS: t_{vacc}, p_{seed}, $a_{1,seed} - a_{2,seed}$, r and f (presence ≤30%) and their values were fixed in the model as outlined below. The seeding time point t_{seed} also did not substantially influence the WLS (presence of 12%), however, because the timing of the epidemic peak substantially differs between seasons (Fig. 3) we retained both the reference and seeding time points (t_0 and t_{seed}) in the model as potential season-specific parameters.

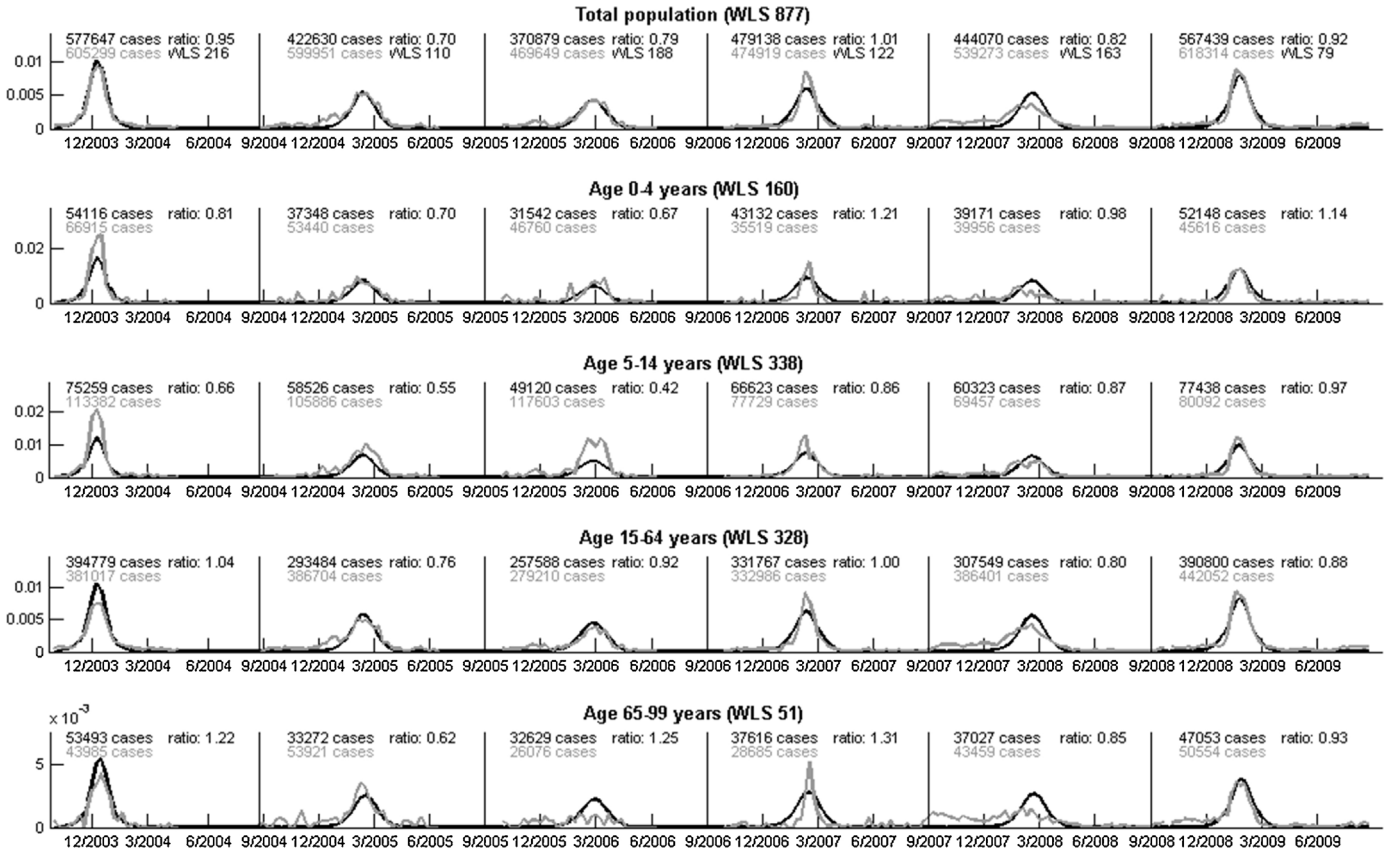

Fig. 3. Observed ILI incidence rates in Belgium 2003–2009 (grey) and corresponding model-based estimates (black) for Model 3: total population (upper panel) and stratified by age group (four panels below). For each influenza season (week 40–20), the total number of reported ILI cases and the corresponding model-based estimate are displayed together with the ratio model/observed.

We assumed that vaccination took place on October 10 each year, as influenza vaccines in Belgium are generally administered in the second or third week of October. We seeded 200 individuals in each age class of 5–50 years at time point t_{seed}, following Vynnycky et al. (2008). Finally, values for the average latent and infectious period were chosen based on the source references of the studies selected in our literature review of dynamic transmission models (Supplementary material and Beutels et al., 2013). There were two eligible source references: a review of experimental challenge studies measuring viral shedding as a proxy for infectiousness (Carrat et al., 2008) and a longitudinal study in households (Cauchemez et al., 2004). These studies suggested an average latent period of $1/f = 1.0$ days (Glasser et al., 2010) and an average infectious period of $1/r = 3.8$ days (Glasser et al., 2010; Cauchemez et al., 2004). Even though it has been suggested that children on average have a longer infectious period than adults, there is no actual data to support this assumption.

3.2. Model fit and parameter estimates

In a first model (Model 1), we allowed the parameters $\tilde{R}_0$, t_0 and t_{seed}, to be season-specific in order to capture between-season variability, e.g. due to the underlying circulation of different influenza strains. The intensity of influenza transmission, the timing of the seasonal peak of transmission and the timing of seeding were thus allowed to vary between seasons. Meanwhile we kept the amplitude δ, the waning rates and the scale factor α constant across all seasons to maintain parsimony and ensure identifiability. We then considered three submodels of decreasing complexity: a model with constant t_{seed} (Model 2), a model with constant t_{seed} and t_0 (Model 3), and a model without season-specific parameters (Model 4).

The parameter estimates obtained with the GlobalSearch algorithm are presented in Table 3, while Fig. 3 shows the fit of Model 3 to the age-stratified ILI incidence data. The fit of the other models are presented in the Supplementary Material. Allowing the average basic reproduction number $\tilde{R}_0$ to be season-specific (Model 4 versus Model 3) greatly improves the fit to the ILI incidence reported through the Belgian surveillance system: the WLS decreases from 2150 to 877. Additionally allowing t_0 and t_{seed} to vary by season seems to have a modest impact on model fit.

The model-based and observed total number of ILI cases are similar for Models 1–3 (Fig. 3 and Supplementary material). The season-specific ratios are fairly close to 1, though the dynamic models tend to underestimate the total number of reported ILI cases per season, with the largest differences observed for seasons 04–06 and 07–08. However, early case reports during these seasons are likely confounded by other pathogens causing ILI and circulating before the actual influenza epidemic. The quality of the fit does not differ substantially between age groups. The total number of ILI cases per season is best approximated for the age group of 15–64 years, whereas the models tend to underestimate the total incidence for children aged 5–14 years.

Parameter estimates can only be interpreted conditionally on the model structure and the values of the other parameters. Nevertheless, we observe some common features across different models. For example, $\tilde{R}_0$ is estimated to be highest in 03–04 and lowest in 05–06, 06–07 and 07–08 (Table 3 and Fig. 4, upper panel), which corresponds well with the seasonal classification in Table 2. The 03–04 season was dominated by the new A/Fujian/411/2002 strain, which was not matched with the influenza vaccine of that year,

Table 3
Parameter estimates for the candidate dynamic transmission models obtained by fitting to ILI incidence data from Belgium (2003–2009) using weighted least squares.

	Parameter	Constant across seasons	2003–2004	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009
Model 1	$\bar{R}_0$		3.81	2.91	2.50	2.70	2.64	3.05
WLS = 831	δ	0.14						
	t_0		01/09	30/10	22/10	24/10	20/10	03/10
	t_{seed}		11/11	22/12	01/12	28/12	29/11	29/12
	$w_i = w_\nu$	0.44						
	α	0.18						
Model 2	$\bar{R}_0$		5.06	3.83	3.29	3.71	3.71	4.09
WLS = 830	δ	0.14						
	t_0		08/09	10/10	22/11	02/11	05/09	04/12
	t_{seed}	30/12						
	$w_i = w_\nu$	0.32						
	α	0.21						
Model 3	$\bar{R}_0$		3.99	2.65	2.22	2.19	2.18	2.44
WLS = 877	δ	0.52						
	t_0	22/10						
	t_{seed}	10/10						
	$w_i = w_\nu$	0.38						
	α	0.21						
Model 4	$\bar{R}_0$	4.60						
WLS = 2150	δ	0.09						
	t_0	12/09						
	t_{seed}	12/01						
	$w_i = w_\nu$	0.25						
	α	0.24						

causing a severe epidemic (Paget et al., 2005). While values of $\bar{R}_0$ are generally higher for Models 2 and 4, this seems to be compensated by a smaller waning rate. Analyzing the 10% best results obtained with the GlobalSearch algorithm indeed reveals a strong correlation between $\bar{R}_0$, the waning rates and the scale factor α.

Models 1, 3 and 4 have reference time estimates in September–October, which means that the seasonal peak of transmission would occur in December–January. On the other hand, large variability in t_0 is observed for Model 2 with transmission peaks occurring in December–March. Estimates of the seeding time vary substantially between models, possibly indicating a lack of identifiability. Estimates of the average duration of vaccine and naturally induced immunity range from 2.3 to 4.0 years.

4. Discussion

This paper presents a dynamic transmission model for seasonal influenza, designed to evaluate age-specific vaccination strategies, with parameters directly estimated by fitting the model to relevant data of influenza-like illness. This approach greatly improves upon the existing practice of imputing pre-specified values to the parameters, without adequately validating the dynamic model outcome. Studies have already indicated the importance of parameter estimation and accounting for uncertainty when using dynamic transmission model outcomes as input for health economic evaluation (Bilcke et al., 2011; Jit and Brisson, 2011). With this work we illustrate another important dimension of uncertainty related to the parameterization of the dynamic model.

It is difficult to compare our parameter estimates directly to literature-based ones, since the interpretation of parameters depends on model parameterization and assumptions as well as the population under study, and moreover, we have found that some parameters are highly correlated. Nevertheless, we can derive an estimate of the effective reproduction number over time, $R_e(t)$, which is an important model summary statistic reflecting transmissibility as well as susceptibility (Fig. 4, lower panel). Chowell and Viboud (2008) estimated R_e from influenza-related mortality

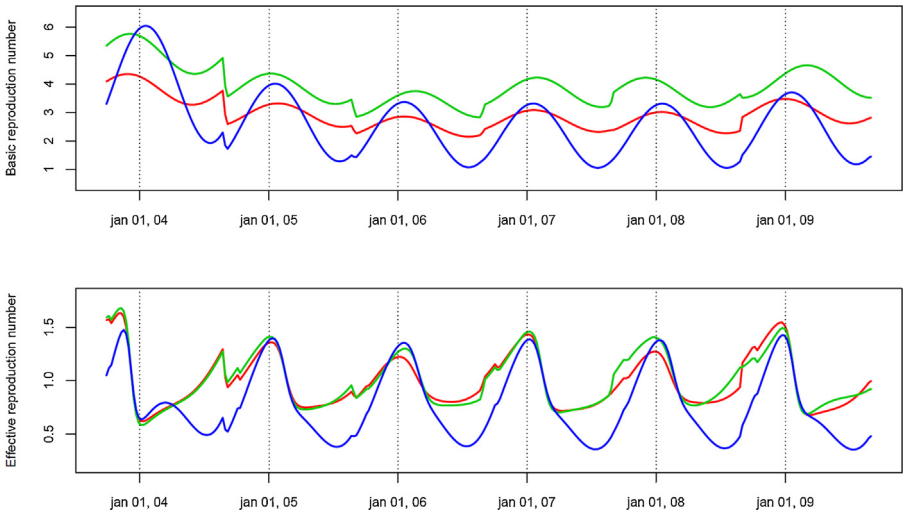

Fig. 4. Evolution of the basic reproduction number (upper panel) and effective reproduction number (lower panel) estimated by Model 1 (red), Model 2 (green) and Model 3 (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

data during one month prior to the epidemic peak across three decades of influenza in the US, France and Australia. Fig. 4 shows that our results match quite well with their estimated average of 1.3 (95% CI: 1.2, 1.4) (Chowell and Viboud, 2008). Note that small differences in model parameterization entail substantial differences between the estimated evolutions of R_0 whereas the estimates of R_e are still fairly similar (Fig. 4). Our results are also in line with other studies showing substantial between-season transmission variability for influenza, partly explained by season dominance of influenza A and B strains (Chowell and Viboud, 2008; Edlund et al., 2011).

The fit of the models were not indicative of a systematic age bias, though underestimation of the total incidence was more pronounced in children of age 5–14 years. This could be due to age-specific heterogeneity that is not represented by our dynamic model e.g. related to inherent differences in susceptibility or infectivity, potential transmission events not captured by the social contact survey, or overreporting of ILI in children by the sentinel GPs. We indeed assumed that the GP reporting rate was constant over time and age, i.e. that the reported ILI incidence are representative of the true ILI incidence. In the absence of auxiliary data, it is difficult to disentangle these effects. Furthermore, we assumed that ILI incidence correlates well with true influenza incidence whereas ILI might also reflect infections from other pathogens such as RSV. Comparing age-stratified ILI incidence to influenza-confirmed ILI incidence shows that the seasonality and the peaks coincide well (Beutels et al., 2013). We also assumed that physical contacts with a total duration of more than 15 min are a good proxy for influenza transmission and that these contact rates are constant throughout the year. Note that our model partly captures time-varying contacts rates through the sinusoidal seasonality function. Since VE_{SP} ignores asymptomatic cases, the proportion of effectively vaccinated might be overestimated. Experimental challenge studies show that the proportion of symptomatic illness in the vaccinated infected is smaller than in the unvaccinated infected, and thus $VE_S < VE_{SP}$ (Basta et al., 2008).

Estimating the parameters in the dynamic transmission model was a non-trivial computationally intensive task, even for the model without season-specific parameters (Model 4). The GlobalSearch algorithm turned out to be sensitive to the initial values and was only able to identify local optima. It was therefore impossible to assess the variance of parameter estimates and to account for the uncertainty originating from the contact data. This precluded the assessment of model identifiability and a simulation study to test the performance of our estimation approach. As future research, we aim to explore and compare likelihood-based and Bayesian approaches using Markov Chain Monte Carlo (MCMC) techniques. Recently, advances have been made in using Bayesian approaches to estimate dynamic transmission model parameters, mainly in the context of rotavirus infection (Baguelin et al., 2013; Bilcke et al., 2015; Weidemann et al., 2014). Note that our approach does not account for serial correlation in the ILI case reports, however, a methodological extension would not be straightforward.

This work differs from Baguelin et al. (2013) who used MCMC to fit their model to ILI incidence data combined with virological confirmation of ILI cases and auxiliary serological data for one season. For Belgium, serological data were not available. Baguelin et al. (2013) considered a strain-specific model and stratified the population by seven age groups. While their Bayesian estimation approach allowed quantifying uncertainty, also with respect to the contact data, the method is highly data-driven and each season and each strain circulating within that season was modelled in isolation. In contrast, we assumed one generic influenza virus and modelled influenza dynamics across multiple seasons including waning and boosting of immunity, facilitating future projections of vaccination.

Though the assumption of a generic influenza virus is a strong one, there are other seasonal influenza modelling studies that made the same assumption (Finkenstadt et al., 2005; Axelsen et al., 2014). Extending our model to account for the diversity of circulating strains which differ from year to year would require many additional assumptions and would likely be a poorer approximation of reality. The advantage of our approach is that it relies on a format of ILI incidence data, as they prevail in many countries using a variety of surveillance systems. Our method could thus prove useful to countries with smaller sized populations and can be used to model ILI incidence from participatory (syndromic) surveillance as well (e.g. Influenzanet).

Axelsen et al. (2014) very recently published a modelling approach similar to the one described in this paper. They fitted an SIRS model to multi-season ILI data from Tel Aviv, Israel, also assuming that ILI incidence is representative of true influenza incidence and modelling disease dynamics neutrally for all strains. Axelsen et al. (2014) also allowed for continuous waning of immunity to capture antigenic drift and included season-specific parameters to model large antigenic jumps. They did not consider the impact of seasonal vaccination nor age-related risk, since their goal was not to use the model to project vaccination scenarios, but to understand the relative contribution of various seasonal drivers. Axelsen et al. (2014) showed that climate variables (temperature, humidity) are important covariates for the seasonality function. Though it is very useful to explore the impact of climate-related dynamics, the use of this model for long-term predictions is not yet possible.

In order to appropriately account for herd immunity when making projections of the effectiveness and cost-effectiveness of different large-scale options for seasonal influenza vaccination, a dynamic transmission model should be incorporated with health economic evaluation (Beutels et al., 2002). The structure and parameter assumptions in dynamic transmission models are extremely influential for such projections, since they extend on the age-specific incidence of ILI and influenza infections. The method we proposed here for parameter estimation instead of imputation is a step forward to improving such projections. The dynamic transmission models resulting from this study were further used to evaluate the impact of various age-stratified vaccination scenarios in Belgium (Beutels et al., 2013).

Acknowledgements

NG is beneficiary of a postdoctoral grant from the AXA Research Fund. LW acknowledges support from an interdisciplinary doctoral grant of the University of Antwerp (Bijzonder Onderzoeksfonds, BOF). YV acknowledges support from a doctoral grant of Hasselt University (BOF11D04FAEC). NH acknowledges support from the Antwerp University Scientific Chair in Evidence-Based Vaccinology, financed in 2009–2014 by an unrestricted gift from Pfizer. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged.

This study was co-financed by the Health Care Knowledge Centre (KCE) of the Belgian Federal government and benefited from discussions held as part of the KCE's expert committee and with Joke Bilcke, Adriaan Blommaert and Pieter Neels. We thank Françoise Wuillaume, Viviane van Casteren and Isabelle Thomas (Scientific Institute for Public Health) for collecting and providing sentinel data on ILI and influenza, and Nancy Thiry for useful discussions. The computational resources and services used in this work were provided by the Hercules Foundation and the Flemish Government – Department EWI. We thank Geert Jan Bex for support with using the VSC cluster.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.epidem.2015.04.002

References

Andreasen, V., 2003. Dynamics of annual influenza A epidemics with immuno-selection. J. Math. Biol. 46 (6), 504–536.

Axelsen, J.B., et al., 2014. Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. Proc. Natl. Acad. Sci. U. S. A. 111 (26), 9538–9542.

Baguelin, M., et al., 2013. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. PLoS Med. 10 (10), pe1001527.

Ballesteros, S., Vergu, E., Cazelles, B., 2009. Influenza A gradual and epochal evolution: insights from simple models. PLoS ONE 4 (10).

Basta, N.E., et al., 2008. Estimating influenza vaccine efficacy from challenge and community-based study data. Am. J. Epidemiol. 168 (12), 1343–1352.

Beutels, P., et al., 2002. Economic evaluation of vaccination programmes – a consensus statement focusing on viral hepatitis. Pharmacoeconomics 20 (1), 1–7.

Beutels, P., et al., 2013. Seasonal Influenza Vaccination: Prioritizing Children or Other Target Groups? Part II: Cost-Effectiveness Analysis 2013, Health Technology Assessment (HTA). Belgian Health Care Knowledge Centre (KCE), Brussels (KCE Reports 204. D/2013/10.273/43).

Bilcke, J., et al., 2011. Accounting for methodological, structural, and parameter uncertainty in decision-analytic models: a practical guide. Med. Decis. Making 31 (4), 675–692.

Bilcke, J., et al., 2015. Quantifying Parameter and Structural Uncertainty of Dynamic Disease Transmission Models Using MCMC: an Application to Rotavirus Vaccination in England & Wales. Med. Decis. Making, pii: 0272989X14566013. [Epub ahead of print].

Bollaerts, K., et al., 2013. Contribution of respiratory pathogens to influenza-like illness consultations. Epidemiol. Infect. 141 (10), 2196–2204.

Carrat, F., et al., 2008. Time lines of infection and disease in human influenza: a review of volunteer challenge studies. Am. J. Epidemiol. 167 (7), 775–785.

Cauchemez, S., et al., 2004. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Stat. Med. 23 (22), 3469–3487.

Chowell, G.M.A.M., Viboud, C., 2008. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. Epidemiol. Infect. 136, 852–864.

DataModeler Release 8.0. 2011, Evolved Analytics LLC.

Earn, D.J.D., Levin, D.J.S.A., 2002. Ecology and evolution of the flu. Trends Ecol. Evol. 17 (7), 334–340.

Edlund, S., et al., 2011. Comparing three basic models for seasonal influenza. Epidemics 3 (3–4), 135–142.

European Commission, 2010. European Public Assessment Reports. Summary of Product Characteristics: FLUENZ Nasal Spray Suspension. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/001101/WC500103711.pdf (accessed 08.08.12).

Eurostat, 2011. Population Table for Belgium, 2009. Eurostat, Luxembourg.

Finkenstadt, B.F., Morton, A., Rand, D.A., 2005. Modelling antigenic drift in weekly flu incidence. Stat. Med. 24 (22), 3447–3461.

Fuhrmann, C., 2010. The effects of weather and climate on the seasonality of influenza: what we know and what we need to know. Geogr. Compass 4 (7), 718–730.

Glasser, J., et al., 2010. Evaluation of targeted influenza vaccination strategies via population modeling. PLoS ONE 5 (9).

Goeyvaerts, N., et al., 2010. Estimating infectious disease parameters from data on social contacts and serological status. J. R. Stat. Soc. Ser. C 59, 255–277.

Goeyvaerts, N., et al., 2011. Model structure analysis to estimate basic immunological processes and maternal risk for parvovirus B19. Biostatistics 12 (2), 283–302.

Halloran, M.E., Longini, I.M., Struchiner, C.J., 2010. Design and Analysis of Vaccine Studies. Springer, New York.

Hanquet, G., et al., 2011. Seasonal Influenza Vaccination: Priority Target Groups – Part I. Good Clinical Practice (GCP). Belgian Health Care Knowledge Centre (KCE), Brussels (KCE Reports 162A. D/2011/10.273/43).

Hsieh, Y.H., 2010. Age groups and spread of influenza: implications for vaccination strategy. BMC Infect. Dis. 10, p106.

Jit, M., Brisson, M., 2011. Modelling the epidemiology of infectious diseases for decision analysis: a primer. Pharmacoeconomics 29 (5), 371–386.

Kissling, E., et al., 2013. Low and decreasing vaccine effectiveness against influenza A(H3) in 2011/12 among vaccination target groups in Europe: results from the I-MOVE multicentre case-control study. Eur. Surveill. 18 (5).

Kretzschmar, M., Teunis, P.F., Pebody, R.G., 2010. Incidence and reproduction numbers of pertussis: estimates from serological and social contact data in five European countries. PLoS Med. 7 (6), pe1000291.

Melegaro, A., et al., 2011. What types of contacts are important for the spread of infections? Using contact survey data to explore European mixing patterns. Epidemics 3 (3–4), 143–151.

Mossong, J., et al., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5 (3), 381–391.

Ogunjimi, B., et al., 2009. Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. Math. Biosci. 218 (2), 80–87.

Paget, W.J., Meerhoff, T.J., Meijer, A., 2005. Epidemiological and virological assessment of influenza activity in Europe during the 2003–2004 season. Eur. Surveill. 10 (4), 107–111.

Pitman, R.J., White, L.J., Sculpher, M., 2012. Estimating the clinical impact of introducing paediatric influenza vaccination in England and Wales. Vaccine 30 (6), 1208–1224.

Pitman, R.J., Nagy, L.D., Sculpher, M.J., 2013. Cost-effectiveness of childhood influenza vaccination in England and Wales: results from a dynamic transmission model. Vaccine 31 (6), 927–942.

Poletti, P., Ajelli, M., Merler, S., 2011. The effect of risk perception on the 2009 H1N1 pandemic influenza dynamics. PLoS ONE 6 (2), pe16460.

Prosper, O., et al., 2011. Modeling control strategies for concurrent epidemics of seasonal and pandemic H1N1 influenza. Math. Biosci. Eng. 8 (1), 141–170.

Shaman, J., Kohn, M., 2009. Absolute humidity modulates influenza survival, transmission, and seasonality. In: Proc. Natl. Acad. Sci. U. S. A.

Skowronski, D.M., Tweed, S.A., De Serres, G., 2008. Rapid decline of influenza vaccine-induced antibody in the elderly: is it real, or is it relevant? J. Infect. Dis. 197 (4), 490–502.

Smits, G.F., Kotanchek, M., 2005. Pareto-front exploitation in symbolic regression. Genet. Program. Theor. Pract. II 8, 283–299.

Stijven, S., Minnebo, W., Vladislavleva, E., 2011. Separating the wheat from the chaff: on feature selection and feature importance in regression random forests and symbolic regression. In: 13th Annual Conference Companion on Genetic and Evolutionary Computation, ACM.

Vandendijck, Y., Faes, C., Hens, N., 2013. Eight years of the great influenza survey to monitor influenza-like illness in Flanders. PLOS ONE 8 (5), pe64156.

Vladislavleva, E., 2008. Model-based problem solving through symbolic regression via Pareto genetic programming. Open Access Publications from Tilburg University.

Vynnycky, E., Edmunds, W.J., 2008. Analyses of the 1957 (Asian) influenza pandemic in the United Kingdom and the impact of school closures. Epidemiol. Infect. 136 (2), 166–179.

Vynnycky, E., White, R.G., 2010. An Introduction to Infectious Disease Modelling. Oxford University Press, USA.

Vynnycky, E., et al., 2008. Estimating the impact of childhood influenza vaccination programmes in England and Wales. Vaccine 26 (41), 5321–5330.

Wallinga, J., Teunis, P., Kretzschmar, M., 2006. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. Am. J. Epidemiol. 164 (10), 936–944.

Weidemann, F., et al., 2014. Bayesian parameter inference for dynamic infectious disease modelling: rotavirus in Germany. Stat. Med. 33 (9), 1580–1599.

Willem, L., et al., 2012. A nice day for an infection? Weather conditions and social contact patterns relevant to influenza transmission. PLoS ONE 7 (11), pe48695.

Willem, L., et al., 2014. Active learning to understand infectious disease models and improve policy making. PLoS Comput. Biol. 10 (4), e1003563.