Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Covid-19 Novel Corona Virus Hybrid Compartmental-Stochastic Epidemic Model

Gareth W. Peters, ...

A Covid-19 Novel Corona virus hybrid compartmental-stochastic epidemic model is developed that incorporates features such as: age, gender, assortative social-activity groups with social mixing network structure, quarantine and isolation components.
A Bayesian statistical model and estimation methodology based on forward projection adaptive Markov chain Monte Carlo is developed in order to perform the calibration of a high-dimensional nonlinear system of Ordinary Differential Equations representing an epidemic model for Covid-19.
The model is compartmental and involves stratification by age, gender and social activity-groups. The concept of social activity groups reflects different types of agents who interact in a community transmission environment including those in critical roles who must commute, those in the service sectors, those in more isolated settings. There is also two additional core features included which are of relevance to the emergence of the covid-19 epidemic which includes interventions and self/cross-excitation components.
We extend the development of a social mixing via a matrix based specific to core population groups. In particular we consider a stochastic mixing matrix framework which allows us to jointly estimate unknown attributes and parameters of the mixing matrix along with the parameters involved in the calibration of the covid-19 epidemic model. This matrix describes the social interactions between members of the population under study and relies on several quantities which are a priori unknown.
The Bayesian model developed allows one to estimate jointly the Covid-19 epidemic model parameters and well as unknown social mixing matrix parameters related to assortativity. This allows us to incorporate networks affects into the spatial-temporal spread of the epidemic.

Keywords: Covid-19

*Correspondence to: Gareth W. Peters, Email: garethpeters78@gmail.com

Prepared using simauth.cls [Version: 2010/03/10 v3.00]

0.1. Contributions

This paper formulates a model for Covid-19, which is minimalistic yet adequate for covering all aspects of the disease transmission, and incorporates seropositivity as a state associated with an individual's natural immunity after recovery.

The core components of the model incorporate relevant features that include:

- The studied population is divided into the following non-overlapping compartments:
  - S is individuals who are at risk of Covid-19 infection,
  - I is infected individuals including asymptomatic and mild cases,
  - G is infected who developed more severe cases of symptoms and required hospitalization,
  - P is those who recovered, and are seropositive and immune,
  - N is individuals who are recovered, immune but seronegative.

  In an extended model we will also consider additional components of the model that may incorporate relevant features for the Covid-19 epidemic that include:
  - Qs is quarantine of individuals at high risk that are susceptible to Covid-19
  - E is individuals that may have been exposed to Covid-19
  - Qe is quarantine of individuals that were possibly exposed
  - Qi is isolation of infected individuals that are known about after testing.
- a - age, g - gender and s - social activity-groups which indicates the level of social activity.
- Movements between compartments are occurring at per capita rates specified by the following parameters: is the force of infection dependent on the proportion of individuals in I, PSC is the probability of becoming seropositive, WIP is the Covid-19 incubation period, DAI is the duration of asymptomatic (i.e. without Covid-19) infection, DWT is the duration of treatment for Covid-19, and DI is the duration of immunity. NOTE: Subscripts denote stratification of parameters: for example, DWTg means that in our model this parameter is gender-dependent.
- Network associativity structure for propensity to interact in different social groups as well as self-excitation features and cross-excitation features.

1. Covid-19 Transmission Models

In developing a model for Covid-19 transmission dynamics one could consider one of two widely accepted modelling approaches that are adopted in the epidemiological literature for addressing epidemic modelling in a population. Fundamentally, these involve consideration of the dynamics of the spread of the disease through the population either specified according to a deterministic dynamical system of coupled ordinary differential equations (ODE) or as a possible alternative, via a system of coupled stochastic differential equations (SDE) based on individuals in the population, which combine to specify the 'state' of the system at a given time.

Typically the approach adopted will depend on both the expected epidemiological outcomes as well as the model specification which will be based on the state of knowledge of the system and the available survey data for calibration. In addition practical modelling features such as the temporal resolution and the dimension of the 'state vector' in the system model will often be of significance in such decisions regarding the use of deterministic equations versus their stochastic alternative frameworks.

As discussed in [1, 2] what differentiates models developed for epidemics in a deterministic framework versus a stochastic individual based model, is three primary considerations: 1) the size of the population under consideration; 2) the proportion of the population which is involved in the epidemic over time (see [3]); and 3) relevance of individual level effects to the research goals (see book length discussions in [4]). Hence, we note that the development of ODE models is justified typically under the assumptions that one is considering large populations in which the mean equilibrium dynamics become of interest to the epidemic modellers. This is a reasonable assumption to make when modelling the sexually active Australian population and it has the added advantage that it also lets us avoid the need to perform stochastic simulations of SDE models over time in thousands of dimensions.

1.1. Covid-19 Transmission Model Version 1

The entire modelled population is viewed as being distributed between a set of non-overlapping groups ("compartments") representing the stages of disease progression. The model is intended to describe how the number of people in each compartment changes over time. For example, members of the susceptible population 'move' from S to I as they become infected, and members of the recovered seropositive population 'move' from P to S as they lose their immunity.

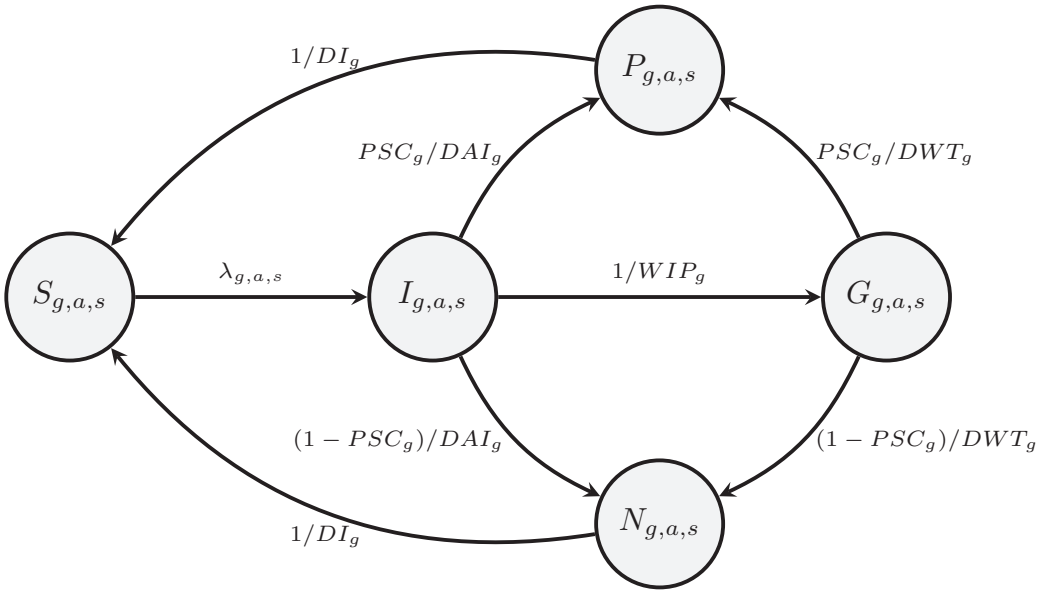

Figure 1. A compartmental Covid-19 transmission model. The studied population is divided into the following non-overlapping compartments: S is individuals who are at risk of Covid-19 infection, I is infected individuals, G is infected who developed serious symptoms requiring hospitalization, P is those who recovered, and are seropositive and immune, N is individuals who are recovered, immune but seronegative. The indices g, a and s indicate that every compartment is stratified by gender, age and level of social activity. Movements between compartments are occurring at per capita rates specified by the following parameters: λ is the force of infection dependent on the proportion of individuals in I, PSC is the probability of becoming seropositive, WIP is the Covid-19 incubation period, DAI is the duration of asymptomatic (i.e. without Covid-19) infection symptoms, DWT is the duration of treatment for Covid-19, and DI is the duration of immunity. Subscripts denote stratification of parameters: for example, DWT_g means that in our model this parameter is gender-dependent.

Underlying Epidemic Model Assumptions

The key assumptions upon which the model is based are presented. We note that the effect of these assumptions on the sensitivity of the calibration of the developed HPV models are studied and summarised in the Discussion and Technical Appendices 1 through to 5.

Population

(a) the modelled population consists of people aged 0-100 y.o. who are divided into 20 separate five-year age-groups (see Table 6). Note: five-year age-groups are commonly used for reporting results of surveys and trials;

(b) the modelled population is constant over time; consequently, immigrants and temporary visitors are not accounted for, which may be an important simplification for some countries whose population has been steadily changing throughout the pandemic due to immigration/mass exodus due to lockdown and social calendar events such as Lunar New Year Festivals; furthermore, some regions are popular destination for travellers who often maintain a high level of social activity;

(c) mortality is formally implemented, and is assumed caused by covid-19 infection and its potential associated complications. In general the propensity for mortality will take on an age and gender structure which will be explored in the model.

(d) it is assumed that initially the number of males in the whole population and every social activity or age-group is equal to the number of females;

(e) no transition between genders is allowed (i.e. males can not become females and vice versa);

Social Mixing Behaviour

(a) the modelled population has all people belonging to one of four social activity (risk) groups (group one being the least active and group four the most active); the proportions of the population in each risk-group 1-4 are π_1, π_2, π_3 and π_4, respectively, [Census data] s.t. $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$ and $\pi_i > 0$;

(b) people are "born" into a particular risk-group and can never leave this group but their activity level is a function of their age; this restriction implies, for example, that even when someone from the most active group gets older, his or her activity level declines but does not drop to the level of representatives of a less active group of the same age; the manner in which an individual of a given age-, gender- and risk-group "interacts" with individuals of a given gender-, age- and risk-group is described by means of a social mixing matrix (discussed in section 1.2);

(c) the social mixing matrix can be based on the adjacency matrix of a spatial network that may reprent how individuals interact around key central infrastructures such as hospitals, shopping centers and major transportation hubs (for instance).

Covid-19 transmission and seropositivity

(a) we assume people seek treatment immediately upon becoming aware that they have symptomatic Covid-19;

(b) we associate seropositivity exclusively with the recovered state such that only those in the recovered (immune) state can be seropositive. Seropositive status is lost upon removal from the recovered state to the susceptible state (i.e., loss of immunity). In the context of our model the recovered/seropositive state corresponds to those people who have developed detectable antibodies to Covid-10 through an immunogenic response. Conversely, the recovered/seronegative state corresponds to those who have recovered and are immune to reinfection but have not developed detectable antibodies. In general, seropositivity can serve as a long-lasting marker of ongoing or prior infection, though it is unclear how reliable it is in the case of Covid-19 as only a proportion of those exposed to infection have been tested for detectable antibodies .... Furthermore, there is some evidence suggesting that seropositivity may be simply a marker of previous infection and an individual who is seropositive may not necessarily be immune XXX (expert knowledge required). Such a perspective would lead to a more complicated model structure and we therefore do not focus on it in this paper. However, the methodology we develop can be extended to this context.

Formulation of the model as a system of ODEs

Our model is formulated as the following system of ordinary differential equations (ODEs):

$$\begin{aligned}
\dot{S}_{g,s,a} &= -\lambda_{g,s,a}(t)S_{g,s,a} + (P_{g,s,a} + N_{g,s,a})/DI_g + \frac{1}{r}S_{g,s,(a-1)} - \frac{1}{r}S_{g,s,a} + \\
&\quad \frac{1}{R}\sum_{g,s}(S_{g,s,20} + I_{g,s,20} + G_{g,s,20} + P_{g,s,20} + N_{g,s,20}) \times \qquad (1)\\
&\quad \delta_1(a)(\pi_1\delta_1(s) + \pi_2\delta_2(s) + \pi_3\delta_3(s) + \pi_4\delta_4(s)), \qquad (2)\\
\dot{I}_{g,s,a} &= \lambda_{g,s,a}(t)S_{g,s,a} - (1/WIP_g + 1/DAI_g)I_{g,s,a} + \frac{1}{r}I_{g,s,(a-1)} - \frac{1}{r}I_{g,s,a}, \qquad (3)\\
\dot{G}_{g,s,a} &= I_{g,s,a}/WIP_g - G_{g,s,a}/DWT_g + \frac{1}{r}G_{g,s,(a-1)} - \frac{1}{r}G_{g,s,a}, \qquad (4)\\
\dot{P}_{g,s,a} &= PSC_g(I_{g,s,a}/DAI_g + G_{g,s,a}/DWT_g) - P_{g,s,a}/DI_g + \frac{1}{r}P_{g,s,(a-1)} - \frac{1}{r}P_{g,s,a}, \qquad (5)\\
\dot{N}_{g,s,a} &= (1 - PSC_g)(I_{g,s,a}/DAI_g + G_{g,s,a}/DWT_g) - N_{g,s,a}/DI_g + \frac{1}{r}N_{g,s,(a-1)} - \frac{1}{r}N_{g,s,a}. \qquad (6)
\end{aligned}$$

Notation:

- Capital letters denote the number of people in a compartment
- the subscripts denote gender (g; for males $g = 1$, for females $g = 2$), one of the four social activity-groups mentioned previously ($s \in \{1, \ldots, 4\}$), and an age-group ($a \in \{1, \ldots, 20\}$);
- the dot denotes a derivative with respect to time;
- $\delta_i(a)$, $i = 1, 2, 3, 4$ is the Kronecker delta function, equal to 1 if $a = i$ or 0 otherwise, and the system coefficients are as in Figure ??.

It should be emphasised that all coefficients in the model formulation are gender specific, i.e., they can take different values for males and females. System (2)-(6) contains a number of terms describing the process of ageing.

Each age-group in our model comprises a five-year band with the same number of people of every age included in the band. Hence, members of the population age (i.e. move to the next age-group) at a yearly rate of r.

To maintain a constant population size we assume that there is an inflow of people into the susceptible compartment of the youngest age-group (group 1) as defined by:

$$\frac{1}{R}\sum_{g,s}(S_{g,s,20} + I_{g,s,20} + G_{g,s,20} + P_{g,s,20} + N_{g,s,20})\delta_1(a)(\pi_1\delta_1(s) + \pi_2\delta_2(s) + \pi_3\delta_3(s) + \pi_4\delta_4(s)).$$

This term is obtained by dividing the total number of individuals leaving the oldest age-group (group 20) each year on reaching age 100, $(S_{g,s,20} + I_{g,s,20} + G_{g,s,20} + P_{g,s,20} + N_{g,s,20})/r$, evenly between 2 genders and 4 social activity-groups. To every g and s is added $S_{g,s,1}$ according to the previously defined distribution of the population across risk-groups (see Table 7).

The implementation of ageing is a mechanism for people to enter and leave the socially active population continuously and is necessary to propagate the effect of interventions that may be taken such as staggered lockdowns and protective measures for elderly that can be of increasing severity: we must ensure that isolated individuals in a particular age-group will later contribute to the number of isolated in older age-groups.

Each of the equations (2)-(6) describes the change in the number of individuals that occurs during a small time period as the sum of the number of individuals entering this compartment from other compartments and those leaving the compartment.

Discussions on construction of compartmental disease transmission models are presented in [5] and [6]. Consider, for example, compartment G (Figure ??): we can calculate the change in the number of individuals in this compartment during a small interval of time by adding up the individuals entering the compartment during this time interval (the infected who developed Covid-19, $I_{g,s,a}/WIP_g$ and the ageing members of G from age-group $a-1$, $G_{g,s,a}/20$) and subtracting the number that leave the compartment (recovered who go either to P or N, $G_{g,s,a}/DWT_g$ and the ageing members of G moving to age-group $a+1$, $G_{g,s,a}/20$). In so doing, we will obtain $I_{g,s,a}/WIP_g + G_{g,s,(a-1)}/20 - G_{g,s,a}/DWT_g - G_{g,s,a}/20$ which is the right-hand side of equation (4).

- It is necessary to point out the crucial role of the force of infection (7) in our model. This is the only non-constant coefficient we have to deal with, which introduces nonlinearity into the system (2)-(6). Its specification utilises a matrix we term the 'social mixing' matrix, which describes age- and risk-group specific patterns of social mixing behaviour in a population.

1.2. Social Network Mixing Matrix

In this section the construction and extensions developed for statistical modelling of the social interaction of members of the population, as defined by the social mixing matrix, are presented.
We consider a Human-to-Human social transmission model, which describes patterns of mixing between age and social activity-groups with respect to central city communities. Like these other models, our approach relies on certain assumptions about the way individuals form their social partnerships. This partnership formation process is commonly referred to as "social mixing".
The social mixing matrix is a $(2 \times 4 \times 4 \times 20 \times 20)$ dimensional matrix comprised of the terms:

- $c^*_{g,s,s',a,a'}$ which are the mean per capita annual rates at which an individual of gender g from a risk or social activity-group s and age-group a acquires new social partners g' from a risk-group s' and age-group a' (these can be genearl socio-demographic or location based - to be decided)...;
- $\rho_{g,s,s',a,a'}$ is the conditional probability that an individual of gender g from social activity-group s and age-group a acquires a social partner g' from social activity-group s' and age-group a'.

It is clear that estimation of all of these parameters is an almost insurmountable statistical challenge, which is one of the reasons why these parameters are often taken as fixed in any given calibration study. There are two broad approaches one could pursue. Given that we are working in a Bayesian modelling framework, the first approach would involve prior elicitation for these population parameters based on expert opinions of annual interaction rates that would be reasonably understood by social-science practitioners, or based on social-media connectedness.
The other alternative involves re-parameterising aspects of this matrix, simplifying it significantly, allowing one to account for the uncertainty associated with specification of this matrix in an appropriate simplified stochastic model.
This would involve finding suitable factors common to aspects of this matrix that could instead be taken as stochastic and estimated in the model calibration, which in turn allow one to derive each element of the social mixing matrix.
In this model, we then utilise this matrix to specify the force of infection (see equations (2) and (3)) for any individual from any subgroup (g, s, a). This term is effectively a annualised rate at which an

individual becomes infected (we could change to daily...), and it is defined as

$$\lambda_{g,s,a} = \beta_g \sum_{s',a'} \left\{ c^*_{g,s,s',a,a'} \frac{I_{g',s',a'}}{S_{g',s',a'} + I_{g',s',a'} + G_{g',s',a'} + P_{g',s',a'} + N_{g',s',a'}} \right\}, \tag{7}$$

where

- β_g is the transmission probability per social interaction, i.e. the probability that a susceptible person of gender g will become infected from his or her infectious social partner of gender g';
- $c^*_{g,s,s',a,a'}$ is the mean per capita rate per year at which an individual from g, s, a acquires new social partners from g', s', a', and
- $I_{g',s',a'}/(S_{g',s',a'} + I_{g',s',a'} + G_{g',s',a'} + P_{g',s',a'} + N_{g',s',a'})$ is a proportion of infected individuals in g, s, a which defines the probability that a new social partner from g, s, a is infected.

In this paper, we will treat the degrees of assortativity by age and social activity-groups, observed to have a noticeable effect on the model calibration, as uncertain, while the rest of the social mixing matrix parameters will be fixed. We do not assume that all of the parameters specifying this matrix are unknown since we want to keep the total number of parameters in our model as low as possible.
While a matrix with these features cannot encompass all the complexities of human social interaction and community mixing, it certainly enables us to explore various relatively plausible mixing scenarios.

1.3. Covid-19 Transmission Model Extended

This extended version of the model is based on compartmental extensions proposed in [7] which have been modified for this Covid-19 framework presented. This includes the features of quarantine and isolation.

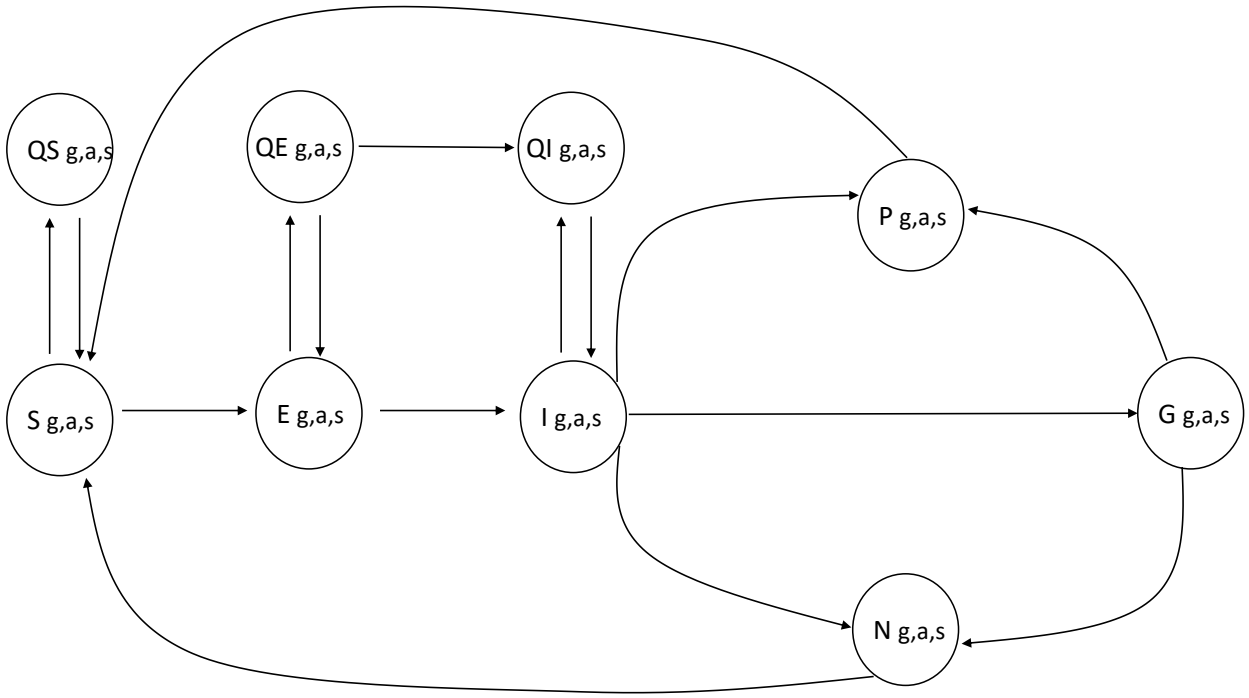

Figure 2. A compartmental Covid-19 extended transmission model. The studied population is divided into the following non-overlapping compartments: S is individuals who are at risk of Covid-19 infection, QS is the individuals who have been quarantined for safety purposes due to their higher risk factors, E are individuals who are known to have been exposed, QE are individuals who were exposed and are now quarantined, I is infected individuals, QI are infected individuals who are now quarantined, G is infected who developed serious symptoms requiring hospitalization, P is those who recovered, and are seropositive and immune, N is individuals who are recovered, immune but seronegative. The indices g, a and s indicate that every compartment is stratified by gender, age and level of social activity.

References

1. Trottier H, Philippe P. Deterministic modeling of infectious diseases: theory and methods. The Internet Journal of Infectious Diseases 2001; 1(2):3.
2. Trottier H, Philippe P. Deterministic modeling of infectious diseases: Applications to measles and other similar infections. The Internet Journal of Infectious Diseases 2002; 2(1).
3. Bartlett M. Measles periodicity and community size. Journal of the Royal Statistical Society. Series A (General) 1957; 120(1):48–70.
4. Brauer F, Castillo-Chavez C. Mathematical models in population biology and epidemiology. 2001.
5. Keeling M, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton University Press, 2007.
6. Vynnycky E, White RG. An Introduction to Infectious Disease Modelling. Oxford University Press, 2010.
7. Coburn BJ, Wagner BG, Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (h1n1). BMC medicine 2009; 7(1):30.
8. Arima Y, Winer RL, Feng Q, Hughes JP, Lee SK, Stern ME, O'Reilly SF, , Koutsky LA. Development of genital warts after incident detection of human papillomavirus infection in young men. The Journal of Infectious Diseases 2010; 202:1181–1184.
9. Winer RL, Kiviat NB, Hughes JP, Adam DE, Lee SK, Kuypers JM, Koutsky LA. Development and duration of human papillomavirus lesions, after initial infection. The Journal of Infectious Diseases 2005; 191:731–738.
10. Insinga RP, Dasbach EJ, Myers ER. The health and economic burden of genital warts in a set of private health plans in the United States. Clinical Infectious Diseases 2003; 36:1397–1403.
11. Giuliano AR, Lu B, Nielson CM, Flores R, Papenfuss MR, Lee JH, Abrahamsen M, Harris RB. Age-specific prevalence,

incidence, and duration of human papillomavirus infections in a cohort of 290 US men. The Journal of Infectious Diseases 2008; 198:827–835.

12. Trottier H, Mahmud S, Carlos M Prado J, Sobrinho JS, Costa MC, Rohan TE, Villa LL, Franco EL. Type-specific duration of human papillomavirus infection: Implications for human papillomavirus screening and vaccination. The Journal of Infectious Diseases 2008; 197:1436–1447.
13. Newall AT, Brotherton JML, Quinn HE, McIntyre PB, Backhouse J, Gilbert L, Esser MT, Erick J, Bryan J, Formica N, et al.. Population seroprevalence of human papillomavirus types 6, 11, 16, and 18 in men, women, and children in Australia. Clinical Infectious Diseases 2008; 46:1647–1655.
14. Pirotta M, Stein AN, Conway EL, Harrison C, Britt H, Garland S. Genital warts incidence and health care resource utilisation in Australia. Sexually Transmitted Infections 2009; 86:181–186.
15. Regan DG, Philp DJ, Hocking JS, Law MG. Modelling the population-level impact of vaccination on the transmission of human papillomavirus type 16 in Australia. Sex. Health 2007; 4:147–163.
16. Garnett GP, Anderson RM. Factors controlling the spread of HIV in heterosexual communities in developing countries: Patterns of mixing between different age and sexual activity classes. Philosophical Transactions of the Royal Society B: Biological Sciences 1993; 342:137–159.
17. Elbasha EH, Dasbach EJ, Insinga RP. A multi-type HPV transmission model. Bulletin of Mathematical Biology 2008; 70:2126–2176.
18. Smith AM, Rissel CE, Richters J, Grulich AE, de Visser RO. Sex in Australia: The rationale and methods of the Australian Study of Health and Relationships. Australian and New Zealand Journal of Public Health 2003; 27:106–117, doi:10.1111/j.1467-842X.2003.tb00797.x.
19. Garnett GP, Anderson RM. Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. Sexually Transmitted Diseases 1993; 4:181–191.
20. Garnett GP, Anderson RM. Balancing sexual partnerships in an age and activity stratified model of HIV transmission in heterosexual populations. IMA Journal of Mathematics Applied in Medicine & Biology 1994; 11:161–192.

Appendix 1. Sexual mixing matrix

This appendix provides details on the mathematical specification of the sexual mixing matrix and the parametrisation of the stochastic matrix that we develop. Sexual mixing matrices are widely used in modelling of sexually transmitted diseases and their main purpose is to describe how certain characteristics of an individual define his or her sexual activity (i.e. the rate of sexual partner change). These characteristics typically are the age and/or sexual activity-group an individual belongs to.
To construct a sexual mixing matrix we generally need to perform the following:

1. Derive relative sexual partner change rates from sexual behaviour data; these rates have to be stratified by age groups and sexual activity groups. Note that in our paper we skipped this step and reused the rates presented in [24];
2. Specify probabilities of sexual partnerships between individuals from different age and sexual activity groups based on assumed degrees of assortativity (see below for details). The degrees of assortativity are typically estimated from sexual behaviour survey data. However, these estimations are always very rough and speculative since the surveys are limited in their ability to quantify personal preferences in the choice of sexual partners by respondents;
3. Adjust the probabilities according to age-specific features of sexual mixing. For example, this can be an increase in probabilities that young females have much older male partners.
4. To ensure that these adjustments do not cause discrepancies in the rates of partner change between groups (i.e. to ensure that the adjusted rate at which females from some group change male partners from another group is the same as the rate at which these males change partners from the mentioned group of females), re-scale the original partner change rates. These should form a sexual mixing matrix.

Notations and available data From now on we will denote a gender by g, the gender opposite to g by g', a sexual activity (risk) group an by s or s' and a or a' will be an age-group. Also, when referring to an individual of gender g who belongs to a sexual activity-group s and age-group a we will simply say "an individual from g, s, a" for brevity.
In the equations describing our model we have a force of infection term for every individual from g, s, a. This term is effectively a yearly rate at which an individual becomes infected, and it is defined as

$$\lambda_{g,s,a} = \beta_g \sum_{s',a'} \left\{ c^*_{g,s,s',a,a'} \frac{I_{g',s',a'}}{S_{g',s',a'} + I_{g',s',a'} + G_{g',s',a'} + P_{g',s',a'} + N_{g',s',a'}} \right\}, \tag{8}$$

where β_g is the transmission probability per partnership, i.e. the probability that a susceptible person of gender g will get infected from his or her infectious partner of gender g'; $c^*_{g,s,s',a,a'}$ is the mean per capita rate per year at which an individual from g, s, a acquires new sexual partners from g', s', a', and $I_{g',s',a'}/(S_{g',s',a'} + I_{g',s',a'} + G_{g',s',a'} + P_{g',s',a'} + N_{g',s',a'})$ is the probability that a new sexual partner from g, s, a is infected defined as a proportion of infected individuals in g, s, a.
Now we will start constructing a sexual mixing matrix for our model according to the approach by Garnett and Anderson [25] and its version used in [26]. In the process of construction we should use the relevant data estimated from the results of Australian Study of Health and Relationships (ASHR). ASHR was a telephone survey of a random sample of about 20,000 people who were Australian residents aged from 16 to 59 y.o. Despite its limitations (see [27] and [24] for discussion) this survey provided important information on sexual behaviour of Australians and its results are currently the most representative ones

available. The data we need (as suggested in [25], [28] or [29]) are the following:

- relative sexual partner acquisition rates r_a for each age-group a (the same for males and females; see Table 6)
- relative sexual partner acquisition rates r_s for each sexual activity (risk) group s (the same for males and females; Table 7)
- overall sexual partner acquisition rate $\bar{c}$ for the whole population (both males and females):

$$\bar{c} = 0.437.$$

All these rates are per capita per year. If for all g, s, a we knew the rates at which an individual from g, s, a acquires new sexual partners from the entire pool of individuals of gender g' (denote them $c_{g,s,a}$), we could find the lowest of these rates,

$$c_{min} : \ \forall g, s, a \quad c_{min} \leq c_{g,s,a},$$

and divide by it all $c_{g,s,a}$ denoting the results of division by $r_{g,s,a}$:

$$r_{g,s,a} = \frac{c_{g,s,a}}{c_{min}}. \tag{9}$$

These would be the relative sexual partner acquisition rates for g, s, a. We can specify them as follows

$$r_{g,s,a} = r_a \times r_s \quad \forall g, s, a. \tag{10}$$

We can now calculate c_{min}. Note that $\bar{c}N_g$ is the number of partners of gender g' all individuals of gender g acquire per year. This number must be equal to the total of sexual partners of gender g' individuals of gender g from each age and sex group acquire, so

$$\bar{c}N_g = \sum_{s,a} c_{g,s,a}N_{g,s,a} = c_{min}r_{g,1,1}N_{g,1,1} + c_{min}r_{g,1,2}N_{g,1,2} + \ldots = c_{min}\sum_{s,a} r_{g,s,a}N_{g,s,a}.$$

Hence,

$$c_{min} = \frac{\bar{c}N_g}{\sum_{s,a} r_{g,s,a}N_{g,s,a}}, \tag{11}$$

also, with c_{min} at hand we can easily calculate all $c_{g,s,a}$ (see (9)).

Proportionate and assortative mixing Now we would like to specify $\rho_{g,s,s',a,a'}$, which are the conditional probabilities that an individual of gender g from sexual activity-group s and age-group a gets a sexual partner of the opposite gender g' from sexual activity-group s' and age-group a'. There are three possibilities to consider:

1. so-called "proportionate" sexual mixing by age-group or sexual activity-group; this means that $\rho_{g,s,s',a,a'}$ may be assumed equal to the proportion of partnerships "generated" by individuals of gender g' from age-group a' (and/or sexual activity-group s):

$$\rho_{g,s,s',a,a'} = \frac{\sum_{s'} c_{g',s',a'}N_{g',s',a'}}{\sum_{s',a'} c_{g',s',a'}N_{g',s',a'}} \tag{12}$$

or

$$\rho_{g,s,s',a,a'} = \frac{\sum_{a'} c_{g',s',a'} N_{g',s',a'}}{\sum_{s',a'} c_{g',s',a'} N_{g',s',a'}}. \tag{13}$$

In other words, an underlying assumption here is that more active members of gender g' have higher chances to become a new sexual partner to someone of gender g. Note that in case of proportionate mixing by both age and sexual activity-group we should simply define $\rho_{g,s,s',a,a'}$ and a product of the right-hand sides of (12) and (13).

2. "assortative" mixing (also known as "with-like" mixing); again, this can be by age or sexual activity-group or both, and if it's, for example, by age-group, we define

$$\rho_{g,s,s',a,a'} = \delta_{aa'}. \tag{14}$$

That is, the probability of establishing a partnership is 1 if a potential partner is of the same age and 0 otherwise.

3. "disassortative" mixing ("with-unlike") suggests that a partnership can be formed only with someone from a different age (or sexual activity) group.

We would like to be able to adjust the sort of mixing in our model depending on our needs. Therefore, we specify the probabilities $\rho_{g,s,s',a,a'}$ as follows

$$\rho_{g,s,s',a,a'} = \left((1-EPSa)\delta_{aa'} + EPSa\frac{\sum_{s'} c_{g',s',a'} N_{g',s',a'}}{\sum_{s',a'} c_{g',s',a'} N_{g',s',a'}}\right) \times \left((1-EPSr)\delta_{ss'} + EPSr\frac{\sum_{a'} c_{g',s',a'} N_{g',s',a'}}{\sum_{s',a'} c_{g',s',a'} N_{g',s',a'}}.\right) \tag{15}$$

Parameters $EPSa$ and $EPSr$ help us vary the extent of assortativeness by age and sexual activity-group: if $EPSa = 0$ mixing is fully assortative by age-group; if $EPSa = 1$ it is fully proportionate by age-group. In a similar fashion we can vary $EPSr$ and control assortativeness by sexual activity-group.

Age related adjustments Here we introduce some adjustments to emphasise the effect of a steady popularity of the partnerships between older males and younger females. Let for now g denote males and g' females. We reduce the probabilities that males in the age-group 3 and higher have female partners of the same age:

$$\rho_{g,s,s',a,a'} \to \rho_{g,s,s',a,a'}(1-\Gamma) \quad \text{if} \begin{cases} a = a' \\ a \geq 3 \end{cases} \tag{16}$$

Similarly, we reduce the probabilities that females in the age-groups 1 to 5 form a partnership with males of the same age:

$$\rho_{g',s,s',a,a'} \to \rho_{g',s,s',a,a'}(1-\Gamma) \quad \text{if} \begin{cases} a = a' \\ a \leq 5 \end{cases} \tag{17}$$

To compensate for the effect of these adjustments we increase the probabilities for males to have a female partner one age-group younger but belonging to the age-groups not higher than 5,

$$\rho_{g,s,s',a,a'} \to \rho_{g,s,s',a,a'} + \Gamma\rho_{g,s,s',a,a} \quad \text{if} \begin{cases} a = a' + 2 \\ a' \leq 5 \end{cases} \tag{18}$$

and also we increase the probabilities that female have one age-group older males from the age-group 3 or higher as a partner:

$$\rho_{g',s,s',a,a'} \to \rho_{g',s,s',a,a'} + \Gamma\rho_{g',s,s',a,a} \quad \text{if} \begin{cases} a = a' - 2 \\ a' \geq 3 \end{cases} \tag{19}$$

So far we have used the rates $c_{g,s,a}$ which describe new sexual partner acquisitions by individuals from g, s, a from the whole population of gender g'. So to find out the rates of acquisition of new sexual partners from a certain age and sexual activity-group (a' and s') we should multiply $c_{g,s,a}$ by $\rho_{g,s,s',a,a'}$. However, now we would like to make the rates $c_{g,s,a}$ group-specific in terms of the groups the sexual partners are selected from. This will be achieved via balancing supply and demand of sexual partners.

Balancing supply and demand We want the following to hold for all g, s, a and g', s', a':

$$c_{g,s,a}\rho_{g,s,s',a,a'}N_{g,s,a} = c_{g',s',a'}\rho_{g',s',s,a',a}N_{g',s',a'}. \tag{20}$$

Here $c_{g,s,a}\rho_{g,s,s',a,a'}$ is the rate of acquisition of new sexual partners by individuals who belong to g, s, a from g', s', a', and $c_{g,s,a}\rho_{g,s,s',a,a'}N_{g,s,a}$ is the total number of new partners acquired by g, s, a from g', s', a'. So the equation simply states that the total number of new partners acquired by g, s, a from g', s', a' must be equal to the total number of new partners acquired by g', s', a' from g, s, a.
Let

$$c_{g,s,a}\rho_{g,s,s',a,a'}N_{g,s,a} \neq c_{g',s',a'}\rho_{g',s',s,a',a}N_{g',s',a'}.$$

We want to find such a multiplier B that

$$B^{\theta_1}c_{g,s,a}\rho_{g,s,s',a,a'}N_{g,s,a} = B^{\theta_2}c_{g',s',a'}\rho_{g',s',s,a',a}N_{g',s',a'}.$$

Then

$$B^{\theta_1-\theta_2} = \frac{c_{g',s',a'}\rho_{g',s',s,a',a}N_{g',s',a'}}{c_{g,s,a}\rho_{g,s,s',a,a'}N_{g,s,a}}.$$

To keep things simple, let $\theta_1 - \theta_2 = 1$. Note that B serves as a degree of imbalance: the balance is established if $B = 1$. So, to ensure that (20) is true it is enough to introduce the group-specific rates $c^*_{g,s,s',a,a'}$ to be used instead of $c_{g,s,a}$:

$$c^*_{g,s,s',a,a'} = c_{g,s,a}B^{\theta_1}, \quad c^*_{g',s',s,a',a} = c_{g',s',a'}B^{\theta_1-1} \tag{21}$$

We limit the range of value of parameter θ_1 to $[0, 1]$. Suppose the supply and demand are not balanced and $\theta_1 = 0$. Then as follows from (21), $c_{g,s,s',a,a'} = c_{g,s,a}$ (the rates for gender g do not get adjusted), but $c_{g',s',s,a',a} = c_{g',s',a'}B^{-1}$ (the rates for gender g' are adjusted). If $\theta_1 = 1$ it is the other way around. Consequently, θ_1 indicates to what extent individuals of each gender adjusts their sexual partner acquisition rates (i.e. we may say, sexual behaviour) in case the available supply of sexual partners does not meet demand.
At this stage our sexual mixing matrix is fully formed.

Appendix 2 (sensitivity to the presence of sexual activity in individuals under 15 y.o.)

In this appendix we illustrate our findings regarding the sensitivity of our model calibration to possible sexual activity in individuals younger than 15 y.o. We introduced an additional age group into our model which includes 10-14 y.o. It is unreasonable to assume the population in this age group is completely sexually active, so we do not expect the relative new sexual partner change rate (PCR) in the group to be close to the yearly rate of 5.2755 assigned to the 15-19 y.o. age group.
Here we present the calibration plots corresponding to the PCR in group 1 equal to 1, 2 and 3 (Figures 3, 4, 5 correspondingly), where 1 is the base PCR (Table 6).

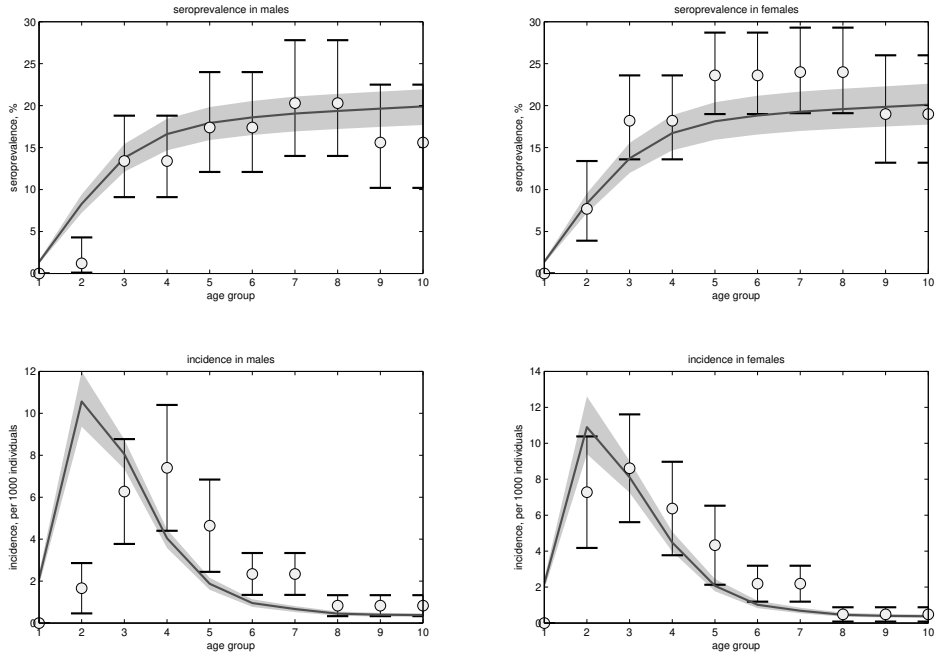

Figure 3. Calibration plot with an additional 10-14 y.o. age group (group 1); relative sexual activity in this group is 1.

We observe that the higher the PCR in group 1, the higher are both incidence and seroprevalence in this group. Also, the higher is the proportion of individuals from age group 2 (15-19 y.o.) who form partnerships with individuals from age group 1 instead of individuals from age group 3 (which is considerably more active than age group 1). Indeed, we should recall that the overall PRCs are pre-specified for each age group other than 1, so any partnerships with the new age group 1 are established only at the price of reduced number of relationships with age group 3. Therefore, we see a decline in incidence in age group 2. Similarly, there is a decline in incidence in age group 3, members of which now form more partnerships with age group 5, less active than age group 2. This effect spreads to the rest of the groups, though it becomes very weak in the groups with low PCRs. Overall, the corresponding changes in seroprevalence are hardly noticeable.
Therefore, we conclude that allowing for some nonzero level of sexual activity in under-15 population we would see the age specific genital warts indidence reduced, particularly in the most sexually active age groups.

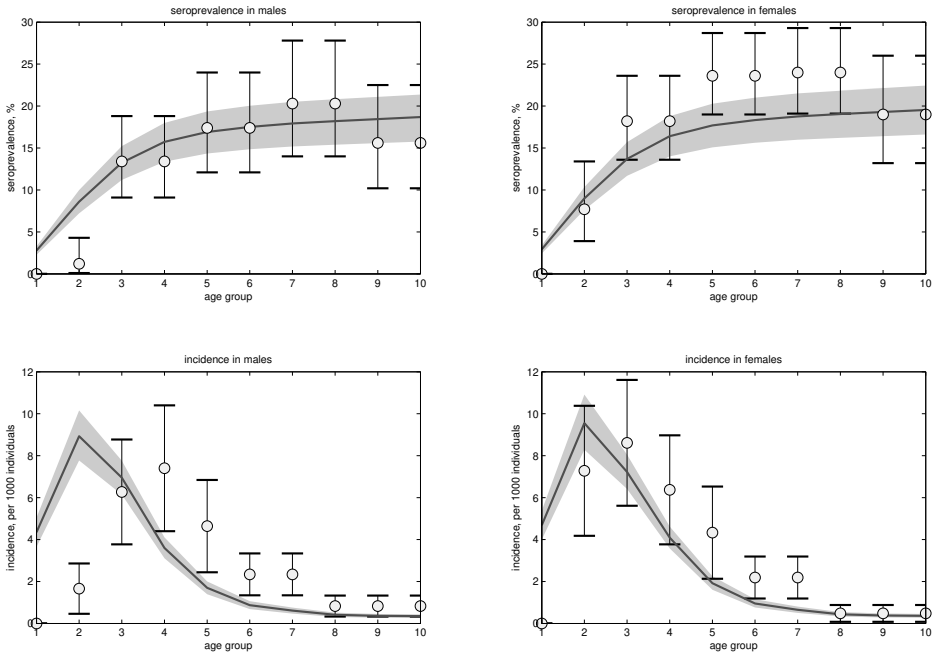

Figure 4. Calibration plot with an additional 10-14 y.o. age group (group 1); relative sexual activity in this group is 2.

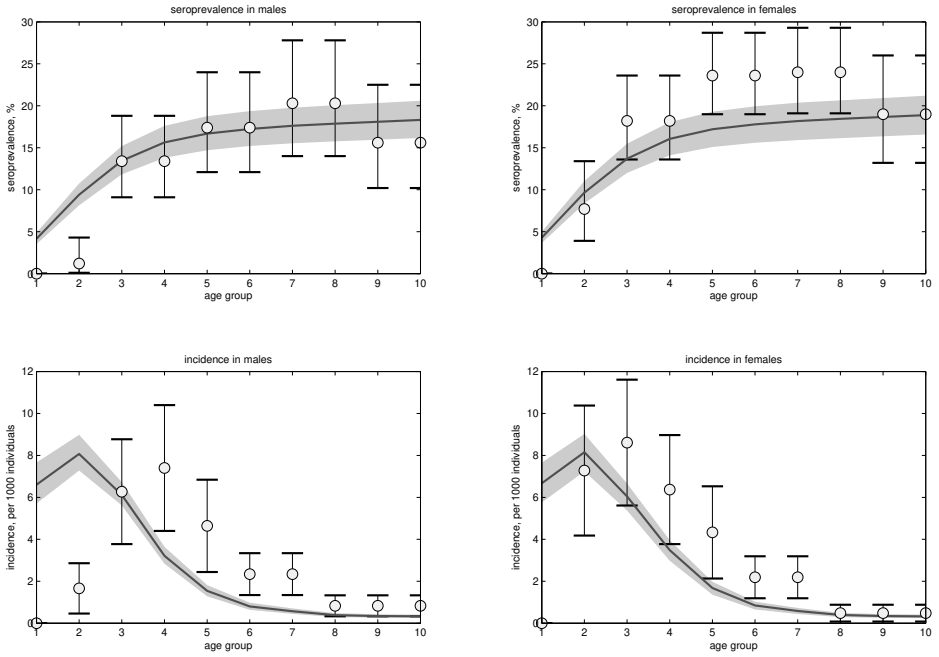

Figure 5. Calibration plot with an additional 10-14 y.o. age group (group 1); relative sexual activity in this group is 3.

Appendix 3 (sensitivity to the durations of natural immunity)

This Appendix studies the sensitivity to the assumption that the durations of natural immunity for males and females are life-long. We demonstrate that a reasonable reasonable choice was made for this

assumption with regard to behaviour of the model.

To assess the effect of varying durations of immunity on calibration we performed simulations for the durations of immunity ranging from 10 to 45 years. For the particular prior distributions on the durations, we first assumed that they are 10-15 years, then 15-20, years, and so on.

It is clear from Figures 6-8 that as immunity lasts longer, individuals remain in state RSP (the only state containing seropositive individuals) longer; hence, we observe higher seroprevalence. At the same time, these individuals contribute less to the infected state, so we see that genital warts incidence decreases.

Note that if the durations of immunity are very long, let's say 40-45 years (Figure 8), the level of seropositivity produced by the model is closer to the real-life reported level than under any other assumed durations of immunity (see Figures 6 and 7). Also, genital warts incidence is clearly closer to the reported values.

Therefore, our model favours long durations of immunity, and we simply set them to as long as possible to reduce the number of parameters in the model.

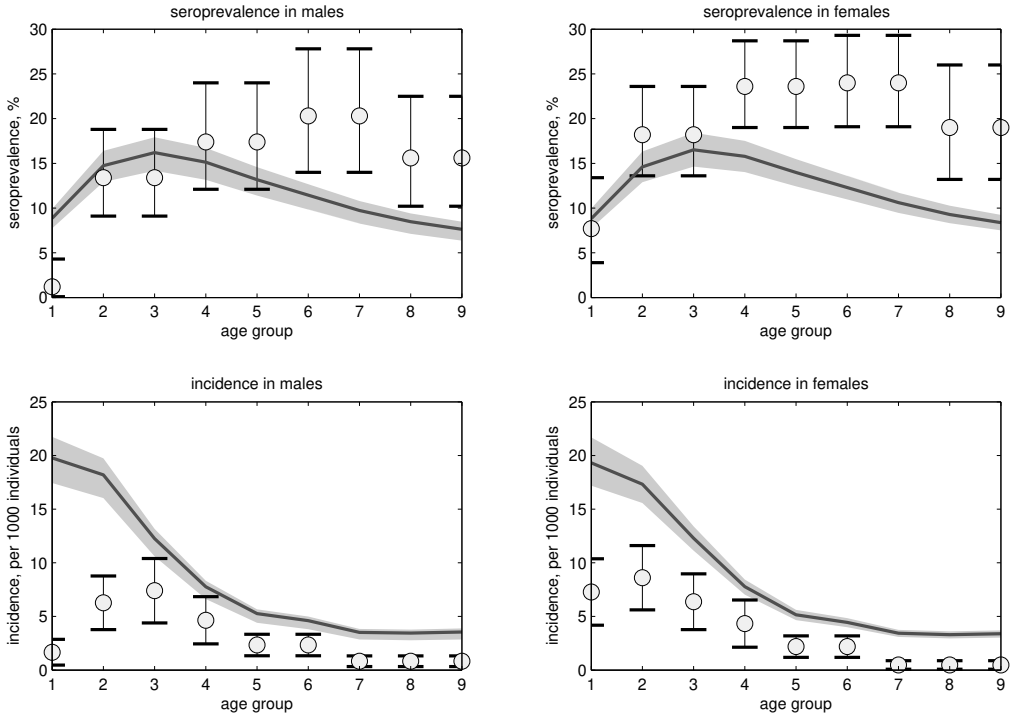

Figure 6. Calibration plot if the durations of natural immunity for both males and females are 10-15 years.

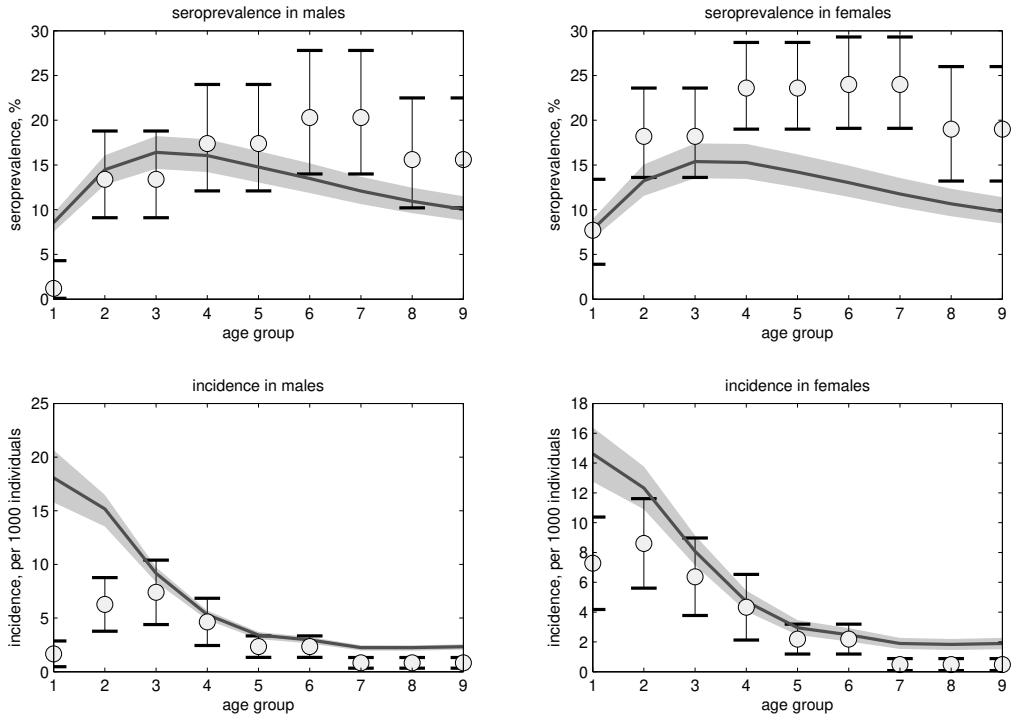

Figure 7. Calibration plot if the durations of natural immunity for both males and females are 20-25 years.

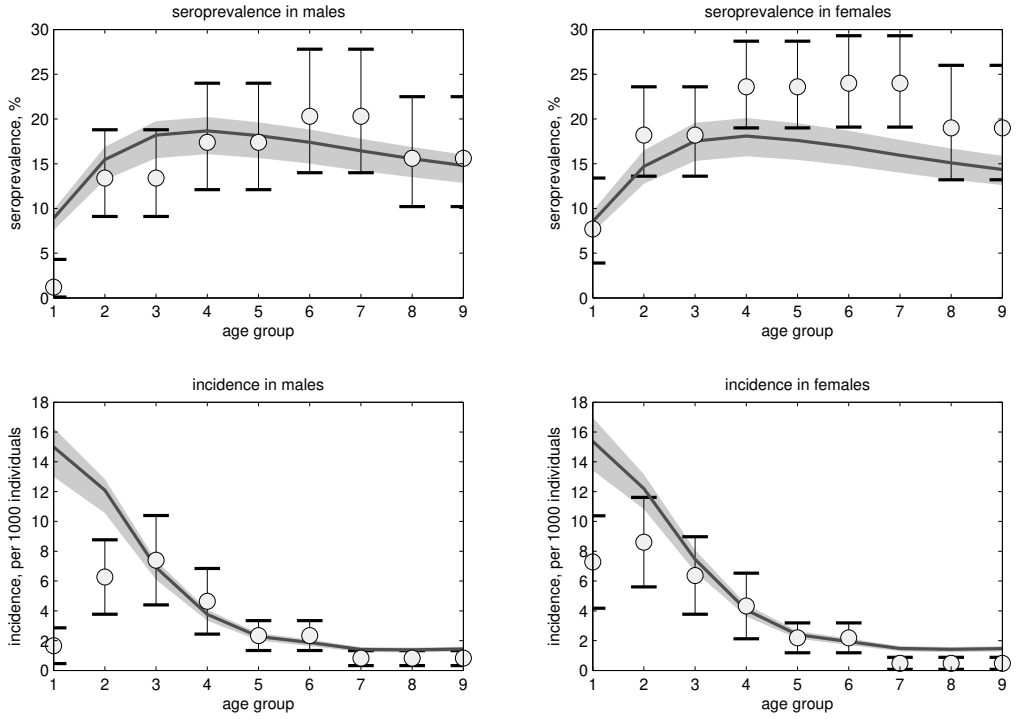

Figure 8. Calibration plot if the durations of natural immunity for both males and females are 40-45 years.

Appendix 4 (sensitivity to demographic changes)

This Appendix studies the sensitivity to the simplified population structure utilised in our model. In particular we made it more realistic by incorporating the birth and death data for Australia published by Australian Bureau of Statistics (http://www.abs.gov.au). We assumed that age-specific death rates for males and females are constant and equal to the reported rates for 2010. This is acceptable since they did not change significantly between 2005 and 2010 (see Table 8). We also assumed a constant birth rate at 12 births per population of 1000. There was no apparent need to analyse the case when the number of males in the population is significantly different from the number of females because the official data says that the difference is not notable.

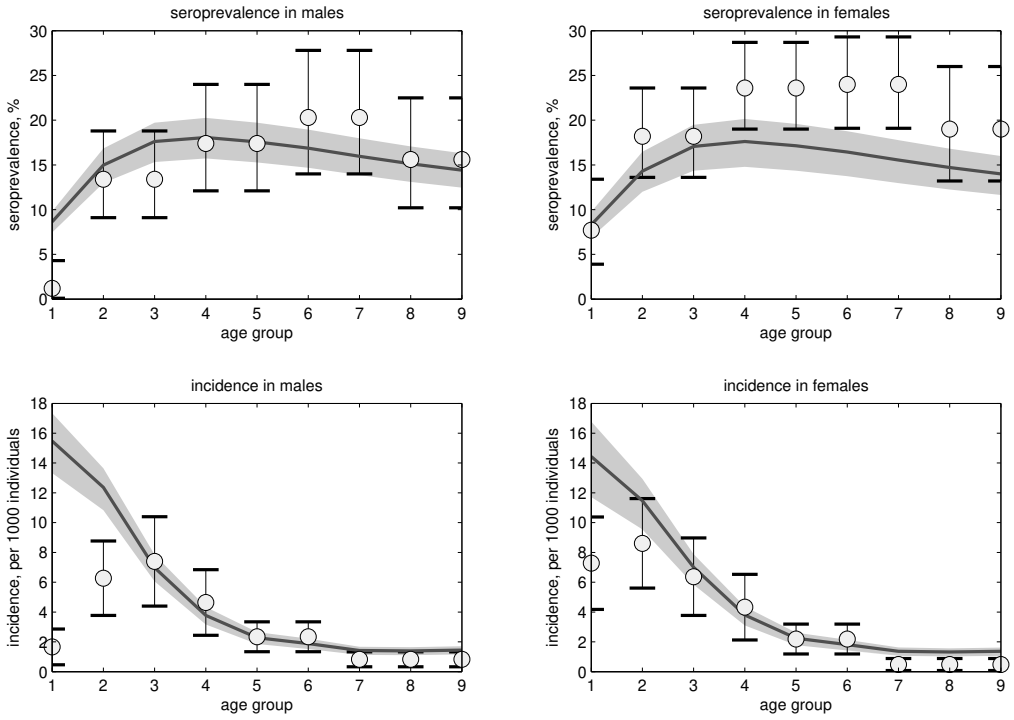

Figure 9. Calibration plot corresponding to the simplified population structure implemented in our model.

As it can be seen from Figures 9 and 10, the effect of the added complexity of population structure is clearly minor and mainly visible in the youngest age group. Considering that the observational data we deal with are particularly ambiguous for age group 1, we felt that the simplified population structure would be appropriate.

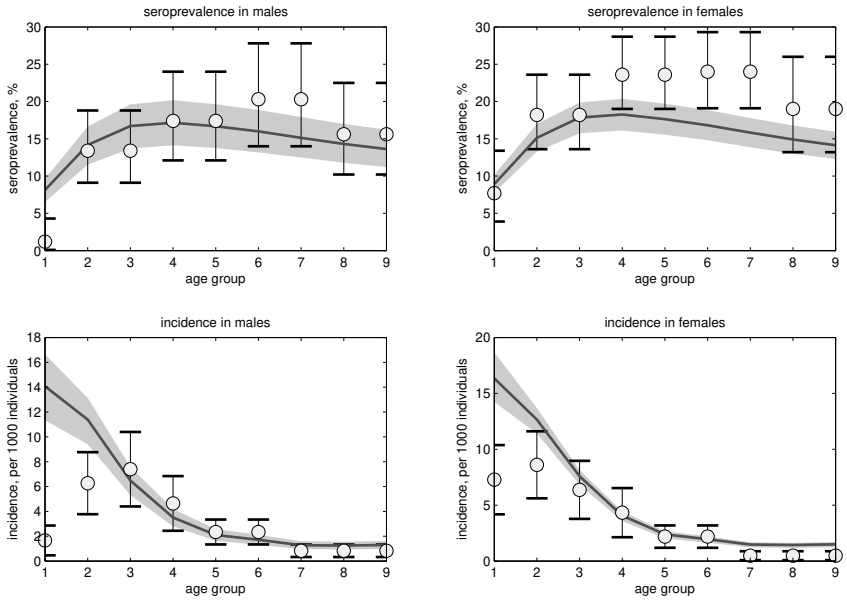

Figure 10. Calibration plot corresponding to the case when the simplified population structure implemented in our model is extended via incorporation of age-specific Australian death rates based on the figures reported by Australian Bureau of Statistics for 2005-2010 and national birth rate estimated for the same period. Data detailed in Table 8.

Appendix 5 (prior distributions)

In this Appendix the prior specifications are developed for Case 1 (that is, based on the data available in the literature, see Table 2).

TR (transmission term)

This parameter is often understood as probability of transmission of HPV infection per sexual partnership of a susceptible male with an infected female or vice versa. There is currently no conclusive evidence to narrow down the range of possible values of this parameter to anything noticeably different from $[0, 1]$. Therefore, we chose uniform prior distributions for both TRm and TRf: $TRm \sim U(0, 1),\ TRf \sim U(0, 1)$.

WIP (genital warts incubation period)

We chose a gamma distribution to be used as prior on this parameter, i.e. for males we have $WIPm \sim \Gamma(\alpha_m, \beta_m)$ and for females $WIPf \sim \Gamma(\alpha_f, \beta_f)$. Since the most frequent incubation period observed during the studies we use the data from was not within the first month, we set the shape parameters α_m and α_f in such a fashion that the mode of the prior is not the origin, simply by $\alpha_m = \alpha_f = 2$. The remaining scale parameter is then utilised to adjust the prior to the study results reporting median durations of incubation period as detailed next.

To specify the prior distribution for the genital warts incubation period for males ($WIPm$) we used the results of a cohort study of 18-21 y.o. male students attending the University of Washington [19]. 18 out of 418 men had incident HPV-6 infection and developed warts during the duration of study (the mean follow-up time was 24.6 months). Among these men the median time between incident infection and detection of genital warts was reported to be 11 months (IQR: 0-16.1 months). We re-calculate these

durations in years by dividing them by 12 months which produced a median of 0.916 (IQR: 0-1.341) years.

The quantile function CDF^{-1} (also known as inverse cumulative distribution function (CDF)) of the gamma distribution has no closed form. However, according to the method of medians we can find β_m solving numerically the following equation: $0.916 = CDF^{-1}(0.5; \alpha_m = 2, \beta_m)$. This gives us $\beta_m = 0.546$.

The duration of genital warts incubation period in females ($WIPf$) was provided by a study of 18-20 y.o. female students of University of Washington [20]. 28 out of 41 women with incident HPV-6 infection (no HPV-11 infection) developed warts. The median incubation period was 2.9 months (IQR: 0-5.7 months). In years this is 0.241 (IQR: 0-0.475) years. As above, according to the method of medians this would involve solving numerically the equation $0.241 = CDF^{-1}(0.5; \alpha_f = 2, \beta_f)$ for β_m . Hence we obtain $\beta_f = 0.144$.

DWT (duration of treatment for genital warts)

To specify a prior distribution for this parameter, we used the data from a study [21] based on analysis of private insurance claims associated with genital warts in the United States (total 3,664,686 claims sampled, 5095 cases of genital warts identified). The mean "duration of episode" for men was reported to be 102.6 days (95% CI: 77.8-127.4), with 237 men considered, or in years (if we divide these numbers by 365 days) this is 0.281 (95% CI: 0.213-0.349).

We assume that DWT for men, $DWTm \sim \Gamma(k_m, \theta_m)$. To specify the shape parameter k_m and scale parameter θ_m we again note that the mode of the distribution would not be located at the origin and so we set the shape parameter $k_m = 2$ and solve for the scale parameters by matching $\theta_m = 0.14$.

The same study [21] reports the mean duration of treatment for women 84.8 days (95% CI: 67.5-102.1), with 299 women considered. In years, the mean is 0.232 (95% CI: 0.185-0.28). We again assume that $DWTf \sim \Gamma(k_f, \theta_f)$ and perform the same procedure to get $k_f = 2$ and $\theta_f = 0.12$.

DAI (duration of asymptomatic HPV infection)

We assign a Gamma distribution to this parameter: $DAIm \sim \Gamma(a_m, b_m)$ and $DAIf \sim \Gamma(a_f, b_f)$. As for the previous parameter, we set the shape parameters a_m and a_f in such a fashion that the mode of the prior is not the origin, simply by $a_m = a_f = 2$. The remaining scale parameter is then utilised to adjust the prior to previous study results reporting the median as detailed next where we apply the method of medians.

A study of the natural history of HPV infections among men aged 18 - 44 years was conducted in Tucson, Arizona [22]. In total 290 men were under observation. The findings showed that the median value of DAI for males ($DAIm$) for HPV-6/-11 was equal to 5.4 months (95% CI: 5.1-5.7 months). In years this is 0.45 (95% CI: 0.425-0.475).

In the same manner as described for WIP, we specify $b_m = 0.268$ as the solution to $0.45 = CDF^{-1}(0.5; a_m = 2, b_m)$.

To specify DAI for females ($DAIf$) we employed the results of a study conducted in São Paulo, Brazil [23]. The recruited 2462 females, 18-60 y.o., who attended a maternal and child health program in a low-income neighbourhood between 1993 and 1997 were followed for up to 10 years. It was found that the median duration of asymptomatic HPV-6/-11 infection was 6.0 months (95% CI: 5.7-6.9), which in years is 0.5 (95% CI: 0.425-0.575). The mean duration was 9.5 months (95% CI: 6.9-12.1) or 0.7916 (95% CI: 0.575-1.0) years. Similarly to the way we calculated b_m, $b_f = 0.472$ is obtained as the solution to $0.7916 = CDF^{-1}(0.5; a_f = 2, b_f)$.

DI (duration of natural immunity)

Almost nothing is known about duration of natural immunity. We investigated a wide range of possible values for this parameter assuming it is distributed uniformly (see the sensitivity analysis in Appendix 3).

PSC (probability of becoming seropositive)

We decided to assign a prior distribution $Beta(2, 2)$ to this parameter since it denotes a probability.