

SCIENTIFIC REPORTS



OPEN

Geographic variation in pneumococcal vaccine efficacy estimated from dynamic modeling of epidemiological data post-PCV7

Erida Gjini 

Although mean efficacy of multivalent pneumococcus vaccines has been intensively studied, variance in vaccine efficacy (VE) has been overlooked. Different net individual protection across settings can be driven by environmental conditions, local serotype and clonal composition, as well as by socio-demographic and genetic host factors. Understanding efficacy variation has implications for population-level effectiveness and other eco-evolutionary feedbacks. Here I show that realized VE can vary across epidemiological settings, by applying a *multi-site-one-model* approach to data post-vaccination. I analyse serotype prevalence dynamics following PCV7, in asymptomatic carriage in children attending day care in Portugal, Norway, France, Greece, Hungary and Hong-Kong. Model fitting to each dataset provides site-specific estimates for vaccine efficacy against acquisition, and pneumococcal transmission parameters. According to this model, variable serotype replacement across sites can be explained through variable PCV7 efficacy, ranging from 40% in Norway to 10% in Hong-Kong. While the details of how this effect is achieved remain to be determined, here I report three factors negatively associated with the VE readout, including initial prevalence of serotype 19F, daily mean temperature, and the Gini index. The study warrants more attention on local modulators of vaccine performance and calls for predictive frameworks within and across populations.

Over the last 20 years, the global epidemiological dynamics of *Streptococcus pneumoniae* have been under intense investigation. Typically carried asymptotically in the human nasopharynx, pneumococcal bacteria also cause disease, such as otitis, pneumonia and meningitis, and are recognized as one of the most important causes of morbidity and mortality worldwide¹. Prevention of bacterial colonization critically reduces chances of spread and transmission, thus indirectly protecting from pneumococcal disease². Under this premise, since 2001, control of this pathogen has focused on host immunization with pneumococcal conjugate vaccines, via which protection applies not only to vaccinated individuals, but also to the rest of the population, a phenomenon known as herd immunity.

The first licensed multivalent pneumococcal conjugate vaccine (PCV7), confers protection against 7 of >90 pneumococcal capsular serotypes³, namely against 4, 6B, 9V, 14, 18C, 19F and 23F, by eliciting anti-capsular antibodies^{4,5}. This vaccine has now been followed by higher-valency vaccines, such as PCV10 and PCV13⁶. From early studies on PCV7^{7–10} the protective effect of pneumococcal vaccination became clear, in terms of reduction of colonization by vaccine serotypes (VT), as well as their associated disease. For example, results from an early vaccine trial¹¹ in Gambia, West Africa, showed that vaccine serotype carriage following vaccination decreased in children who had received two doses of the vaccine (odds ratio OR 0.22 [95% CI: 0.08–0.61]), but was accompanied by an increase in non-vaccine type (NVT) carriage (OR 0.65 [95% CI: 0.29–1.42]). A later community-randomized controlled trial¹² investigating the effect of PCV7 on nasopharyngeal colonization among American Indian infants showed that vaccinees were less likely to be colonized with VT (OR 0.40 [95% CI, 0.23–0.67]), but again were more likely to be colonized with NVT pneumococci (OR 1.67 [95% CI: 1.02–2.78]).

After mass childhood vaccination was introduced in routine immunization programs around the world, epidemiological surveys have documented the response in immunized host populations to this vaccine, e.g. in the USA^{13–15} in Switzerland¹⁶, England^{17,18}, the Netherlands¹⁹, and many other countries. These reports have

Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, Oeiras, Portugal. Correspondence and requests for materials should be addressed to E.G. (email: egjini@igc.gulbenkian.pt)

described the post-vaccination changes in prevalence of pneumococcal bacteria across settings, microbial population structure, serotype and clonal distribution and antibiotic resistance patterns, elucidating the great diversity and evolutionary potential of pneumococci. While vaccine serotypes declined, overall colonization rates were seen to remain generally stable, partly due to serotype replacement, the process by which non-vaccine serotypes rose in carriage as well as disease (reviewed in ref. 20). Although variation in vaccine effects and serotype replacement across communities has been recognized^{20,21}, the reasons for this variation are unclear. As the net benefit of vaccination critically depends on the balance between vaccine protection and serotype replacement, and the complex relationship between carriage and disease, an important challenge in pneumococcus epidemiology remains to understand and predict post-vaccination dynamics, going beyond descriptive approaches^{22,23}. This includes capturing variation between settings under the same mechanistic framework of pneumococcus transmission. Monitoring disease outcomes, which are often the endpoint in clinical trials²⁴, provides limited insight into the underlying mechanisms that determine herd immunity and serotype replacement^{25,26}. For this, carriage data are essential, needing new statistical approaches²⁷, and dynamic modeling²⁸. In this paper, I apply one such integrative approach to post-vaccination data from different settings, uncovering epidemiological parameters responsible for the differences in serotype replacement rates.

By fitting the same dynamic model to several studies reporting post-vaccine dynamics in day care centres, I address the question of what is the realized vaccine efficacy, consistent with the serotype replacement reports in different countries, and what may be the underlying local underlying factors. Although pre-licensure trials offer basic estimates of vaccine efficacy^{7,29}, they typically focus on disease outcomes in one particular population, which cannot be directly extrapolated to new untested contexts. After mass immunization takes place in another transmission setting, it is possible that a different net protection at the individual level may result. Such variation could be driven by environmental variability, local serotype and clonal composition, as well as by demographic, social or immuno/genetic factors in the host population. Accurate estimates of vaccine efficacy against serotype acquisition are crucial to subsequently predict or interpret impact on pneumococcal disease²².

Given a multitude of pneumococcus transmission models^{30–33}, the choice of the appropriate one for retrospective vaccine assessment is not straightforward. The type and detail of the data inevitably constrain the complexity of possible model structures. A solution is to start with simple models which afford questions at a general level, and then gradually build in complexity for finer questions when feasible. The advantage of coarser-grain descriptions is that parameter readouts are likely to be more robust and comparable across settings. In this spirit, the present study uses a simple epidemiological framework³⁴, based on a neutral model for pneumococcus dynamics, to integrate temporal observations pre- and post- PCV7. The model accounts for time of survey, transmission intensity and variation in coverage rates, recognized as strong sources of heterogeneity in meta-analyses³⁵. Moreover, this framework provides a tool to comparatively assess vaccine performance in field settings, incorporating the basic epidemiological feedbacks, including herd immunity and serotype replacement, typically neglected in vaccine trials and in purely statistical post-licensure approaches^{12,23,27,36}.

The data consist of cross-sectional prevalences of pneumococcal carriage in day care centres (DCCs) in Portugal, Norway, France, Greece, Hungary and Hong-Kong, from studies conducted before and after implementation of PCV7. DCCs are an important setting of transmission of pneumococcus in developed countries. Pre-school children display highest nasopharyngeal carriage rates, and act as the main reservoir for pneumococcal spread in the community^{37,38}, consequently becoming critical vectors of herd immunity effects. It has been shown that pneumococcal transmission may take place as micro-epidemics driven by the day care centres³⁹. Often the serotypes identified as dominant in some DCCs, i.e., those found to cause micro-epidemics, are not necessarily the most transmissible ones. It is thus probable that most serotypes transmit in a similar fashion in the child population, and micro-epidemic patterns and neutral micro-epidemic bacterial evolution⁴⁰ result primarily from heterogeneous transmission between interconnected host clusters.

The analysis of pneumococcus dynamics has persistently lingered between a serotype-specific approach and neutral models based on equivalent trait approximations. The empirical evidence appears mixed: some studies support slight variation in serotype acquisition, competition and clearance rates^{41–44}, others conclude that serotype differences are not significant^{45,46}, others find differences only in clearance rates and not in acquisition⁴⁷. Furthermore, when trying to match serotype trait hierarchies across epidemiological settings^{42,43}, inconsistent estimates emerge, questioning universally fixed serotype traits. Given the inconclusive picture, and the diversity of pneumococci beyond the serotype level, multiple models have adopted neutral formulations, assuming equal transmission and clearance rate across serotypes^{31,38,48–50}. Some of these describe pre-vaccination dynamics⁵⁰, others predict vaccination impact across host age classes⁴⁹, or interpret epidemiological dynamics after vaccination^{31,34}. Other models neglecting serotype differences focus on transmission fluxes between child care centers and the community⁵¹, pneumococcus interactions with influenza⁵², and seasonality⁵³.

In this paper, I also adopt the neutral perspective to model pneumococcus colonization dynamics in young children attending day care. Serotypes are partitioned in two functionally relevant groups: aggregated vaccine and non-vaccine serotypes, and equal traits across serotypes are assumed as a first order approximation, averaging out underlying clonal diversity. In this context, serotype replacement is exclusively driven by vaccination. With a *multi-site-one-model* approach, my aim here is threefold: (i) to use minimal assumptions to tie multiple strands of evidence together, making the most of sparse data, (ii) to gain a deeper retrospective insight into vaccine efficacy of PCV7, by estimating setting-specific parameters underlying dynamic serotype replacement, (iii) to assess variability in realized vaccine protection across settings, and generate hypotheses for this variation.

Results

Pneumococcus colonization data from different settings. Data were extracted from 6 different cross-sectional studies on pneumococcus prevalence before and after vaccination with PCV7: the studies are summarized in Tables 1 and 2. Because not all the data published in these papers are used in the model fitting,

Country	Survey period	Mean sample size	Mean age (range) in months	Time in DCC	Vaccine coverage % (ρ) (mean over all ages)	Co-col. (%)	Ref.
Portugal	2001–2007	359	48 (12–71)	24	17, 32, 56, 65, 63, 75	10%	54, 55
Norway	2006–2008	599	41 (12–60)	32	10, 40	14%	56, 57
France	1999–2008	321	23 (3–39)	17	0 (99–01), 37, 37, 68.4, 68.4, 90.1	—	58
Greece	2005–2009	662	47 (13–76)	27	12.9, 32.6, 70.1, 95.5	3.8%	59
Hungary	2001–2009	413	54 (36–72)	18	0, 10 (02–05), 21.5, 38.5, 38.5	—	60, 61
Hong Kong	2000–2010	2094	49 (24–72)	24	0 (00–05), 20.4, 26.7, 35.7, 43, 43	—	62

Table 1. Characteristics of the studies in daycare settings from different countries before and after PCV7 vaccination. Mean age of children in each study and expected duration of DCC attendance (see S2.2 for details) are given in months. Vaccination proportions per year were extracted from the children that that received at least one dose of PCV7 across the studies. When information on vaccination coverage for intermediate years was missing, the same value as in the preceding year was assumed.

Country	Sampling months: (t)	Number of carriers: $n_V(t) + n_N(t)$	Carriers of PCV7 serotypes: $n_V(t)$	Sample sizes: $n(t)$
Portugal	0, 72	173, 319	102, 50	270, 449
Norway	0, 24	470, 474	212, 97	606, 592
France	0, 36, 60, 84, 108	161, 172, 182, 168, 155	123, 128, 89, 35, 6	298, 294, 334, 335, 343
Greece	0, 12, 24, 48	370, 206, 286, 334	132, 88, 33, 33	769, 494, 566, 820
Hungary	0, 78, 112	34, 38, 357	19, 17, 46	95, 121, 1022
Hong Kong	0, 120	383, 347	253, 190	1978, 2211

Table 2. Pneumococcus prevalence data from daycare settings in different countries. For Portugal and Norway more data were available and were used than shown here (see Methods 4).

in the interest of clarity and reproducibility of my analyses, here I summarize the key information that is useful in the context of this study. The criteria for selection of studies were: (i) focus on day care transmission setting; (ii) report on multiple time points; (iii) relatively large sample size; and (iv) no interference with more recent vaccines. Each study analyzed pneumococcus colonization in day care centres in a different country: in Oeiras, Portugal^{54,55}, in Oslo, Norway^{56,57}, in Alpes Maritimes, France⁵⁸, in Larissa, central Greece⁵⁹, in Szeged, Hungary^{60,61}, and in Hong Kong⁶². Out of these, only three studies reported co-colonization, respectively the Norwegian and Portuguese study, which have also been partly modelled previously^{34,50}, and the Greek study.

All studies used sensitive serotyping techniques on nasopharyngeal swabs collected from healthy children attending day-care. The settings varied in serotype and clonal composition, vaccination coverage and time of follow-up (Table 1). The study sample sizes varied between a minimum total of 240 children a maximum of 2094 children in Portugal and Hong-Kong respectively. Among other findings, all studies reported decreasing trends in the proportion of colonized hosts with serotypes included in PCV7 after implementation of vaccination. All studies reported no temporal trend with regards to overall colonization prevalence, except the Hong Kong and the French study, which showed slightly declining carriage over time.

The number of circulating serotypes across time periods was between 26 and 33 in all studies. The four most prevalent serotypes before vaccination in each country varied slightly in their ranking: 19F, 23F, 11A/D, 6B (Portugal), 6B, 19F, 15B/C, 23F, (Norway); 6B, 19F, 23F, 14 (France); 6A, 11A, 10A, 35F (Greece); 14, 19F, 23F, 6A/B (Hungary); 6B, 19F, 23F, 14 (Hong-Kong) (see Table S1 for details). The Portuguese and Norwegian study reported in addition specific colonization prevalences with PCV7 and non-PCV7 serotypes both in single and multiple carriage after vaccination. The number of time points per study is 2 (Portugal), 2 (Norway), 5 (France), 4 (Greece), 2 (Hong Kong) and 3 (Hungary). The exact number of day care centres (DCCs) varied slightly: from 11–12 DCCs in Portugal, to 27–29 in Norway, 25 in France and 21 in Greece. For model fitting, aggregated colonization data across DCCs are used, assuming homogeneous mixing between all the children. To facilitate comparison across settings, host age structure is neglected, approximating children attending day care, aged between 1–5 years as a compact age group with similar characteristics, spending on average about 2 years in the DCC environment (see S2.2 for details).

Fitted transmission model. The dynamics of pneumococcal colonization and co-colonization among young children attending day care are modelled with a previously published neutral model³⁴. The susceptible-infected-susceptible (SIS) formulation tracks six epidemiological compartments, including susceptible hosts, S , hosts colonized by any one vaccine serotype (here PCV7) I_V , hosts colonized by one non-vaccine serotype, I_N , and co-colonized hosts I_{VV} , I_{NN} , and I_{VN} with two vaccine serotypes, two non-vaccine serotypes or one of either group, respectively (Fig. 1). With vaccination, the model equations double, to account for vaccination

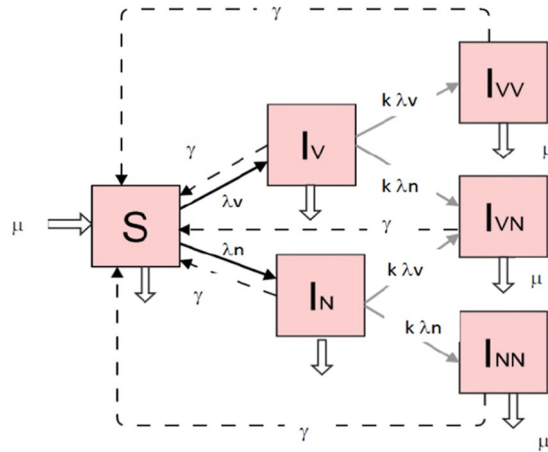


Figure 1. Model flow diagram in the absence of vaccination. S, susceptible, I_V , single carriers of any vaccine serotype, I_N , single carriers of any non-vaccine serotype, I_{VV} double carriers of any two vaccine serotypes, I_{NN} , double carriers of any two non-vaccine serotypes, I_{VN} , double carriers of a vaccine and non-vaccine serotype. With vaccination, the structure of the model is doubled to account for vaccination status of each epidemiological class. The parameters are defined in the text. See Methods 1–3 for details, and Supplementary Material S1 for full ODE equation formulation.

status of each epidemiological class. A fraction ρ of all new susceptible hosts, recruited at rate μ , is vaccinated, a quantity that can vary across geographical settings and also from year to year (Table 1). Vaccinated hosts display partially reduced susceptibility to VT, relative to non-vaccinated ones, given by the scaling parameter $1 - VE$, where vaccine efficacy against VT, denoted as VE ($0 \leq VE \leq 1$), describes the net per-capita protection aggregated over all serotypes included in the vaccine. Other studies in the literature^{31,33}, also represent aggregated vaccine efficacy with a single parameter, assuming similar and long-lasting protection on average across all PCV7 serotypes. Further details on model structure, features and assumptions are given in Methods 1–3 and Supplementary Material S1.

In line with previous neutrality assumptions^{31,38,48,50}, also here, phenotypic equivalence across serotypes, with respect to basic colonization capacity (transmission rate β), clearance rate (γ) and direct competition coefficient in co-colonization (k) is assumed. The pneumococcus force of infection for each serotype group (λ_v and λ_n) depends explicitly on their prevalence in the population, including contribution by single and double carriers. This feature is critically reflected in the model initial conditions, when calibrated to each dataset under the assumption of pre-vaccine endemic equilibrium (see Methods 2). In all settings, initial conditions are adjusted according to local serotype prevalences, linking the ensuing dynamics to reported VT/NVT distribution before vaccination. Once initialized, the vaccination model trajectories are fitted to the remaining cross-sectional observations, specific to each study (see Methods 4 and data in Table 2). Stochasticity due to sampling effects is accounted for using a multinomial distribution with a different sample size, for the likelihood of the data at each time point. Parameter estimation is performed in a Bayesian framework in MATLAB, using adaptive Metropolis-Hastings MCMC⁶³.

Estimated basic reproduction number R_0 and serotype competition parameter k . An endemic persistence equilibrium requires $\beta > \gamma + \mu$, corresponding to the classical threshold for the basic reproduction number⁶⁴: $R_0 = \frac{\beta}{\gamma + \mu} > 1$. Provided that $R_0 > 1$, the pneumococcus endemic colonization state is asymptotically stable. Because it is a measure of transmission intensity, the higher R_0 is, the higher the endemic prevalence of the pathogen, and the higher also the co-colonization. In particular, susceptibles, single colonization and co-colonization prevalences are given by:

$$S^* = 1/R_0, \quad I_1^* = I_V^* + I_N^* = \frac{1 - 1/R_0}{1 + k(R_0 - 1)}, \quad \text{and} \quad I_2^* = I_{VV}^* + I_{NN}^* + I_{VN}^* = k(R_0 - 1)I_1^*. \quad (1)$$

Serotype distribution at endemic equilibrium under neutrality is thus flexible regarding single and multiple carriage combinations, but globally satisfies $P_V^* + P_N^* = 1 - 1/R_0$, relative to VT and NVT prevalences, a conserved relationship even after vaccination. At a global level, in a symmetric multi-type system, if transmission rate β (or analogously R_0) stays constant over time, targeted vaccination is not expected to alter the overall magnitude of single and co-colonization; only shift the serotype composition towards non-vaccine types, a picture supported by serotype replacement patterns post-PCV7 across countries^{13,16,17}, and the majority of our data. The basic reproduction number in each setting was assumed constant over time, and estimated within the integrated fitting procedure for the entire time-series.

The values for R_0 estimated for these studies, range between 1.2 and 4.7 (Table 3), indicating variable transmission intensity within day-care attendees across geographical sites, lowest in Hong Kong and highest in Norway, respectively (see Fig. 2). Each child carrying pneumococcus within the DCC causes thus on average 2.5 new colonizations over his entire carriage period. The model estimates also a high competition between serotypes acting at

Country	Basic reproduction number R_0	Competition parameter k	Vaccine efficacy VE	Transmission rate β (month ⁻¹)
Portugal	3.18 (2.87, 3.53)	0.08 (0.06, 0.11)	0.11 (0.09, 0.14)	2.36 (2.13, 2.62)
Norway	4.77 (4.28, 5.30)	0.04 (0.03, 0.05)	0.43 (0.35, 0.50)	3.49 (3.13, 3.88)
France	2.22 (2.14, 2.31)	—	0.19 (0.16, 0.21)	1.68 (1.62, 1.74)
Greece	1.85 (1.79, 1.91)	0.045 (0.04, 0.05)	0.21 (0.17, 0.25)	1.36 (1.32, 1.41)
Hungary	1.55 (1.49, 1.62)	—	0.27 (0.20, 0.31)	1.17 (1.12, 1.22)
Hong Kong	1.23 (1.22, 1.24)	—	0.08 (0.05, 0.10)	0.91 (0.90, 0.92)

Table 3. Parameter estimates for pneumococcus transmission and PCV7 vaccine efficacy (VE) in day care settings. The means and the 95% credible intervals for each parameter are obtained from the posterior distributions. The competition coefficient k is estimated from co-colonization data in some cases, and set to a fixed value in others (—), equal to 0.05, a criterion used to calibrate model initial conditions (see Methods 2). The basic reproduction number is derived as: $R_0(i) = \frac{\beta(i)}{\gamma + \mu(i)}$. Setting-specific birth/death rate μ for the SIS model was assumed as the inverse of mean duration of DCC attendance for children in each study (see Table S2 for results based on 1/mean age instead).

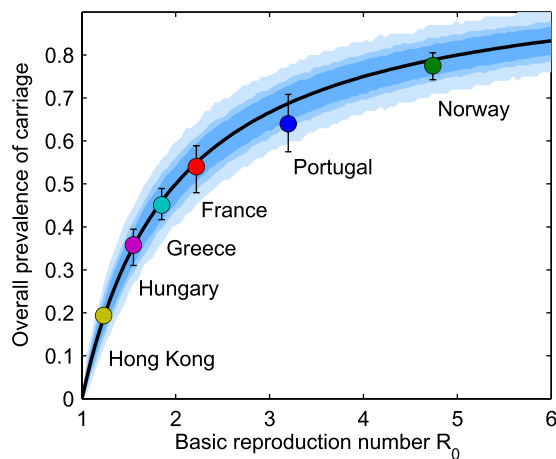


Figure 2. Transmission intensity vs. overall pathogen prevalence in a SIS framework. The mean R_0 's estimated from our datasets are matched with the expected overall prevalence of pneumococcal carriage in the pre-vaccine era. The bars reflect the 95% confidence interval for binomial proportions over 300 simulations with same R_0 , taking into account the mean sample size in each setting (Table 1). The line depicts the theoretical nonlinear relationship expected by the model (Eq. (1)). The shaded regions represent the 95% quantile of simulations under different sample sizes: $n = 100$ (light blue), $n = 250$ (darker blue) and $n = 500$ (darkest shade).

acquisition, in the range of 90–95% inhibition of co-colonization, independently from the datasets that reported co-colonization, namely the Norwegian, Portuguese and Greek settings.

Links with other studies. The R_0 values resemble R_0 estimates for other respiratory pathogens such as influenza⁶⁵. The highest R_0 was estimated for Norway, displaying highest carriage prevalence. This may be related to the low rate of antibiotic use and high rate of DCC attendance in Norway compared to other settings⁵⁶. Understanding R_0 variation is important when forecasting population-level impact of a particular vaccination programme⁶⁶. For example, when contemplating vaccines based on serotype-transcending antigens, depending on the value of R_0 in each setting, a different vaccination coverage with such universal vaccine would be needed to achieve the same impact. In addition, transmission intensity critically shapes the pathogen evolutionary potential, by affecting nonlinearly prevalence of carriage (Fig. 2) and multiple-carriage, the latter being key in bacterial recombination^{67, 68}.

Competition coefficient estimates are consistent with earlier studies from Denmark and England^{31, 69}. In fact, the pre-PCV7 Norwegian data from 2006 has also been analyzed previously⁵⁰, with a stochastic SIS Markov model for transmission within DCCs. Strikingly, with the simplified deterministic model in this paper, applied to both pre- and post-vaccination data, and neglecting infection hazard from the community, very similar estimates emerge for both transmission rate β (3.59, 95% CI: 2.8, 4.6) within DCC and competition between serotypes at acquisition k (0.1, 95% CI: 0.06, 0.14), suggesting a minor influence of stochastic effects on SIS model parameter estimation⁷⁰, at least at this particular sample size ($n \approx 600$). For the other studies not reporting co-colonization, k could not be estimated from the data. Instead, a fixed value for k (=0.05) was used to simulate their trajectories, in line also with earlier estimates for this parameter^{31, 50, 69}.

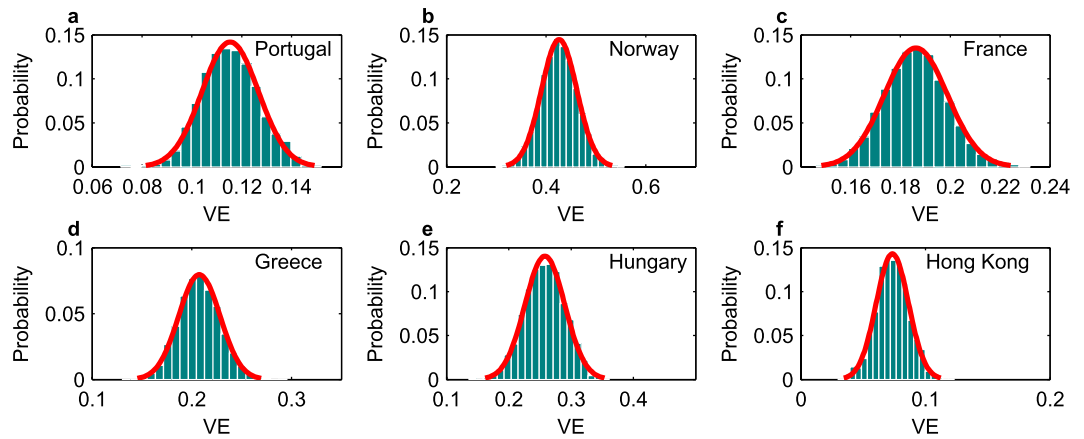


Figure 3. Posterior distributions for heptavalent vaccine efficacy against PCV7 serotypes, estimated in a setting-specific manner. Parameter inference is based on dynamic model fitting to cross-sectional data in day care settings from different countries, accounting for different transmission intensities, initial serotype distribution, and vaccine coverage. Red lines indicate normal distribution fits. Numerical results are detailed in Table 3.

The resulting monthly child-child transmission rate β for France in this study, around 1.68 (95% CI: 1.62, 1.74), was similar, although somewhat higher, to previous estimates obtained from a longitudinal French modeling study⁴⁵ with schoolchildren 3–6 years in 2000 (1.45, 95% CI: 1.30, 1.59). The difference could partly be explained by the fact that the children in the Alpes-Maritimes study⁵⁸, modelled here, represent a younger age group, more prone to colonization, and by the fact that the present epidemiological model neglects transmission from the non-DCC community, thus attributing all the force of infection to the DCC environment. Interestingly, the β estimated for Greece 1.36 (95% CI: 1.32, 1.41) compares better with the range of the French study⁴⁵, probably because the children age ranges match more closely. The lowest transmission rate estimated for the Hong-Kong day care setting, seems closer to the transmission rates within families in Bangladesh, between 0.64 and 0.84 per month estimated previously⁴⁶, while transmission rates for Hungary are consistent with another French study reporting within DCC transmission rates between 1.04 and 1.18⁴².

Vaccine efficacy against acquisition of PCV7 serotypes. The other parameter estimated with this model is vaccine efficacy against acquisition of PCV7 serotypes, aggregated over all VT (Fig. 3). This parameter reflects the reduced susceptibility in the acquisition of VT, either as primary or secondary colonizing serotypes, in vaccinated relative to non-vaccinated individuals. The 6 independent estimates for each dataset, obtained by fitting the dynamic model to post-vaccine observations, accounted for differences in vaccination coverage (Fig. S1), time of survey, and transmission intensity between settings. As summarized in Table 3, the reported values of VE suggest differential relative vaccine protection, experienced on average by each vaccinated child in the age-groups considered, across settings. According to this model, the four datasets from Portugal, France, Greece and Hong Kong are consistent with lower values for vaccine efficacy, around 10–20%, while the Norwegian and Hungarian data are consistent with a higher value of vaccine efficacy around 30–40% (see Fig. 3), with the 95% credible interval for Norway VE in the range 30–50%. Notice that the model can distinguish between differences in efficacies that are invariant to serotype distribution prior to vaccination, because the initial conditions for each study are calibrated to reported local serotype frequencies aggregated as VT and NVT prevalences, at the first time point (see Methods 2). The predicted trajectories for the epidemiological dynamics with the estimated parameters, drawn from the posteriors (Table 3), are shown in Fig. 4, matching well the pneumococcus prevalence observations in each country.

Sensitivity to model assumptions. The estimated VE values remain robust to eventual changes in clearance rate γ across settings (affecting only R_0), and to the value of serotype competition parameter k , affecting mainly the ratio between dual and single serotype carriage. However, they are slightly more sensitive to the assumed μ (susceptible host recruitment rate), which besides modulating R_0 , multiplies directly the vaccination coverage ρ . When μ increases, turnover rate in the modeled population increases, and vaccine efficacy estimates decrease slightly for the same coverage (Fig. S2). In the case of using $1/\text{mean age}$ for μ (Table 1), as an alternative to $1/\text{mean duration of DCC attendance}$, the estimated vaccine efficacies varied slightly, with Norway and Hungary still displaying highest values around 45%, and R_0 values kept unchanged (see Table S2 for results).

When using a continuous vaccination assumption instead of a step-wise increasing coverage, i.e. holding fixed the vaccination coverage, throughout the years of each survey, the sensitivity analysis for the estimated VE indices yielded somewhat lower values in absolute terms to the above ones reported here, yet preserving the rank-ordering between multiple sites, and showing little dependence on the assumed μ (Figs S3 and S4). When vaccination coverage is held constant, e.g. around the median of values in Table 1, the resulting VE estimates display an overall downward bias, because in this case, the model over-estimates the magnitude and sequential effects of the positive feedbacks that arise in the early stages of vaccination, due to a higher-than-realistic assumed coverage in year 1. As vaccine coverage ρ and vaccine efficacy VE can be traded-off against one another,

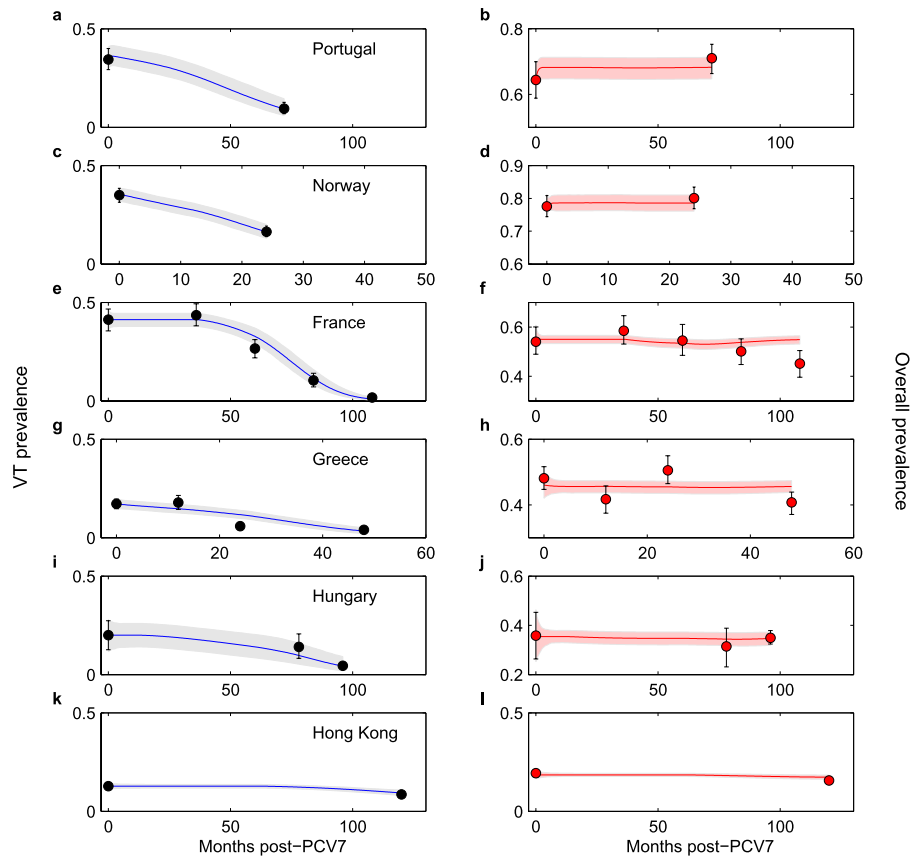


Figure 4. Observed and fitted prevalence of asymptomatic carriage of pneumococcus in day care centres by country. The model-predicted 95% credible envelopes for temporal trajectories of VT prevalence (left panels) and overall prevalence (right panels) after vaccination are superimposed on observed data in different day care settings at specific time-points. (a,b) Portugal. (c,d) Norway. (e,f) France. (g,h) Greece. (i,j) Hungary. (k,l) Hong Kong. The credible envelopes were computed using 500 simulations with different parameter combinations sampled from the joint posterior, and here account only for the uncertainty around parameter values and initial conditions. For the credible envelopes accounting for sampling uncertainty throughout time (with mean sample size as in Table 1) see Fig. S6.

potentially yielding indistinguishable model trajectories, it is important to realize that VE estimates given a certain dataset remain coupled to the originally assumed coverage rates.

While here the model-based vaccine efficacy estimates assume stable endemic transmission, in reality there may be cases where prevalence has declined. In particular the French dataset indicated a slight temporal trend for reduction in overall prevalence of pneumococcus, which suggests a time-varying R_0 for this country, probably driven by other interventions than vaccination (e.g. change in antibiotic use). Although, these changes were not addressed in the model-fitting, in Fig. S5, model-predicted serotype replacement dynamics for different values of vaccine efficacy and lower transmission are illustrated, and the difference appears minimal. Under a neutral model, such as the one applied here, changes in R_0 should not impact significantly on the relative rate of decline of targeted serotypes post-vaccination. By contrast, under a non-neutral model, for example where direct competition at co-colonization may be asymmetric, the magnitude of transmission intensity, R_0 , is expected to influence the hierarchical dynamics between types both in the absence⁷¹, and presence of vaccination⁷².

Possible factors of vaccine efficacy variability across settings. In spite of all simplifying assumptions, model estimates of vaccine efficacy for PCV7 emerging from this study are close to existing estimates in the literature^{31,73}. For example, Choi *et al.*³¹, using a similar framework, estimated vaccine efficacy in England $VE_c = 50\%$, close to the value for Norway estimated here. More recently, a meta-regression study⁷³, pooling data from different settings together, also reports that the aggregate VE_c for all PCV7 serotypes 6 months after completion of the vaccination schedule around 57% (95% CI: 50, 65%), tending to an aggregate VE_c of 42% (95% CI: 19, 54%) at 5 years. Compared to these earlier approaches, a novel effect emerging from the present study, calling attention, is the heterogeneity in vaccine efficacy against PCV7 serotypes across geographical settings, which requires a deeper look at the data and our model formulation. Among many possible factors, here I explore only three in a statistical sense, inviting further examination of the local modulators of vaccine performance in the field.

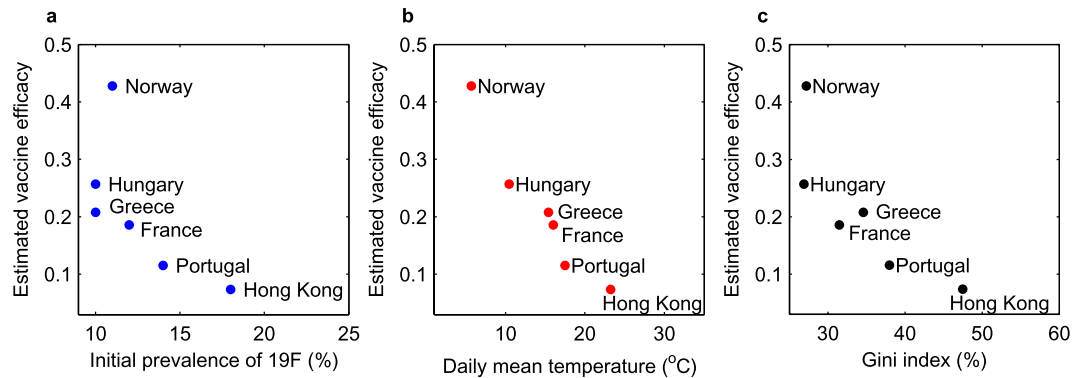


Figure 5. Possible factors for variation in vaccine efficacy against acquisition. **(a)** Prevalence of serotype 19F pre-PCV7 and model-based vaccine efficacy estimates. The decreasing trend suggests that high initial frequency of a serotype for which vaccine protection may be inferior (e.g. 19F) could influence model estimates of the net realized vaccine efficacy against PCV7 serotypes when aggregated together, ($R^2 = 47\%$, (R^2 -adj = 11%, GLM fit: $\rho = -0.0003$, $p = 0.10$ n.s.). **(b)** Temperature across countries correlates with vaccine efficacy values estimated by the model ($R^2 = 92\%$, R^2 -adj = 87%, GLM fit: $\rho = -0.019$, $p < 0.005$ significant), suggesting a possible environmental factor in immunomodulation and vaccination responses. **(c)** Income inequality, as a proxy for transmission heterogeneity, is associated negatively to realized vaccine efficacy extracted by the model ($R^2 = 68\%$, R^2 -adj = 46%, GLM fit: $\rho = -0.013$, $p < 0.05$ significant). More details are provided in the Supplementary Tables S4 and S5. The fitted regression lines are displayed in Fig. S7.

Serotype-specific efficacies and aggregation effects. The first is related to a main assumption of the aggregated (VT/NVT) model, namely that vaccine efficacy is equal against all 7 serotypes included in the vaccine, which although previously assumed, is clearly an approximation. Lower vaccine protection against acquisition of some PCV7 serotypes, for example against 19F, has been observed in some studies^{74,75}, and confirmed by more recent meta-analyses⁷³. Such observation implies that vaccine type composition prior to vaccination, in terms of individual frequencies in each country, could be important, besides their aggregated prevalence as a group. Since the dynamic model does not describe single serotypes and serotype-specific efficacies, aggregation effects may introduce bias in overall vaccine efficacy estimation. Indeed, when inspecting the VT composition in each setting and explicitly considering the frequency of serotype 19F prior to vaccination, it emerged that in those settings where 19F was initially more prevalent, model fitting led to a lower apparent vaccine efficacy against PCV7 serotypes (Fig. 5a), suggesting a role for the serotype aggregation effect. Although this trend did not reach statistical significance with these datasets, it is not surprising, given that serotype-specific vaccine protection and serotype-specific protection waning rates have been reported for PCV7⁷³.

Environmental gradients: the role of temperature. A second potential source of modulation of vaccine efficacy locally could be environmental factors, such as temperature, humidity, ultraviolet (UV) radiation, typically varying with geographical latitude. It has been suggested that exposure to UV radiation, as in sunlight, can modulate immune responses in animals and humans⁷⁶. This immunomodulation can sometimes be deleterious, as documented by experimental animal studies where UV exposure was shown to impair resistance to many infectious agents, including bacteria, viruses, and fungi⁷⁷⁻⁷⁹. When testing among our results for an association between setting-specific temperatures (daily mean, averaged over the year, (see Table S4)) and the model-estimated pneumococcus vaccine efficacies, a significant association emerged, whereby in cooler climates, PCV7 vaccine protection was realized at higher levels (see Fig. 5b). As temperature correlates with latitude and UV exposure, it is possible that environmentally-induced immunomodulation across geographical settings could interfere with the net individual protection against PCV7 serotypes conferred by the vaccine. Naturally, this hypothesis warrants further studies across more settings and experimental evidence for substantiation. Yet, geographical effects and gradients in auto-immunity have been established, indicating decreasing auto-immune disease along the north-south axis⁸⁰. Another strand of evidence has implicated temperature with many growth and virulence properties in bacteria⁸¹. In the particular case of pneumococci as respiratory tract colonizers, thermoregulation across mucosal surfaces might contribute to modulating their biological interaction with vaccine-induced anti-capsular antibodies.

Transmission heterogeneities in the host population. Third, the model employed here assumes a homogeneous underlying population with regards to pneumococcus transmission and focuses only on the day care setting. It is known that heterogeneity within host populations is very important for transmission dynamics of infectious diseases and for the design of control policies. A statistical pattern known as the 20/80 rule has been shown to apply to a variety of disease systems⁸², implying that control programs targeting the core 20% group of transmission are potentially highly effective in reducing levels of infection in the population as a whole, compared to programs that fail to reach all of this group. Mathematical models have shown that the impact of control measures such as vaccination depend crucially on the magnitude of heterogeneities in disease risk, and is lower when these heterogeneities are higher⁸³. It is not unreasonable to assume that such transmission heterogeneities may

be interwoven with economic disparities in a given population. Under this premise, using the Gini coefficient, a measure of income inequality⁸⁴, as a proxy for transmission heterogeneity, where a Gini coefficient of zero implies perfect equality in distribution of wealth, and a Gini coefficient of 1 (or 100%) expresses maximal inequality, I tested whether there was an association between the setting-specific Gini indices (Table S5) and the realized apparent PCV7 vaccine efficacy in each country, extracted by the model for the 6 pneumococcus DCC datasets. As shown in Fig. 5c, there was a significant negative association, where higher vaccine efficacy was found in more homogeneous countries like Norway and Hungary, and lower efficacy per individual was realized in more heterogeneous settings like Portugal and Hong Kong. This role of heterogeneity on vaccine efficacy estimation has also been predicted by earlier theoretical arguments in the context of malaria and vaccine efficacy trials⁸⁵. Such a relationship hints at a new possible predictor of PCV impact on pneumococcal carriage and disease, besides the ones already known²³, namely, heterogeneity in the underlying host population, which may turn out extremely relevant when interpreting reports of vaccination effects in low- and middle-income countries, starting only recently to become available.

These findings point to a clear need for a precise mathematical understanding of post-vaccine dynamics variation across settings. Even though with the current model we cannot fully resolve the underlying causes of the variation in these 6 datasets, the study calls attention on an important aspect of pneumococcus vaccination dynamics, previously overlooked: namely, the host-pathogen-vaccine interaction across a gradient of host populations. While the details of how this variable vaccine effect is achieved remain to be addressed in follow-up studies of a larger scope, variation in itself may have important consequences for site-specific responses to current and future pneumococcus vaccines, and for other interventions.

Discussion

The effect of vaccines in immunization programmes is often described by pooling data from different populations together^{21,73}. In this study, I have taken an alternative route, by applying a unified modelling framework in parallel to pneumococcus datasets from different countries. Interpolation of colonization data before and after vaccination via dynamic model fitting provides quantitative insight on the mechanistic coupling between serotype competition for colonization, transmission intensity, and vaccine efficacy in a setting-specific manner. This allows a deeper interpretation of serotype replacement dynamics, and serves as a first step for comparative analyses. Similar *multi-country-one-model* approaches have been insightful in other epidemiological systems, for example when investigating tuberculosis transmission⁸⁶ and malaria dynamics⁸⁷. In our case, emergent heterogeneities in individual direct protection, after factoring out differences in vaccination coverage and initial serotype distribution across sites, raise interesting questions.

New insights from multi-site-one-model approaches. A novelty of the present study is the challenge to the notion that vaccine protection can be described by a single number or point estimate, equal across different populations and geographic settings. The difficulties of conducting pre-licensure vaccine trials in each setting should not preclude efforts to interpret retrospectively post-vaccine dynamics in a context-specific manner. There may be multiple factors that can affect the local performance of a vaccine, and mechanistic approaches, applied retrospectively to well-sampled, detailed, cross-sectional data, can help to disentangle these, addressing not only pathogen-intrinsic variability, as typically done, but also variability of host populations. In this study, focusing on the first pneumococcus conjugate vaccine, PCV7, and fitting the same transmission model individually to six observational datasets, describing daycare colonization patterns in different countries, I uncover variable estimates for vaccine efficacy across settings. Some of these, e.g. VE in the range 30–40% (Hungary and Norway), are close to existing estimates from PCV7 vaccine trial analysis²⁷ reporting VE against acquisition around 42%, and models of cross-sectional data from England³¹, reporting a vaccine efficacy against acquisition of 50%. However, the other datasets analyzed here suggest also that a lower realized vaccine efficacy, around 12–20%, may apply in countries like Portugal, France and Greece, closer to the lower end of the range (95% CI: 0.24, 0.56), estimated with previous statistical approaches using carriage as an endpoint²⁷.

Recognizing its multi-factorial nature as a parameter, realized vaccine efficacy in these studies could be statistically related to three plausible factors, given the resolution afforded by the present datasets and model formulation. The first regards vaccine protection assumed uniform for all serotypes included in the vaccine, and described by a single parameter. When accounting for the fact that vaccine-induced immunity against serotype 19F is weaker than that against other PCV7 serotypes^{74,75}, a negative trend became apparent between variable initial prevalence of 19F (the difficult serotype) at the beginning of the study period, and the resulting variable apparent protection against PCV7 serotypes when pooled together. Secondly, a stronger and significant negative relationship emerged when considering temperature and variability in PCV7 efficacy across countries, warranting further studies about environmental factors in immune responses to pneumococcal conjugate vaccines. The third association related negatively a demographic factor across populations, namely the Gini indices, a measure of income inequality and plausible proxy for transmission heterogeneity, with the realized vaccine efficacy. These results have to be taken with caution as the power to detect significant correlations is realistically limited with only 6 datasets considered here. While correlation does not imply causation, many avenues for future investigation of mechanisms remain.

The pneumococcus diversity challenge. Although pneumococcal serotypes are identified by the same polysaccharide capsule code across geographical settings, the clones corresponding to those serotypes can be vastly different, displaying potentially different growth, competition and antibiotic resistance phenotypes. For example, the proportion of isolates resistant to erythromycin was 77% in the pre-vaccination period in Hong Kong⁶², 70% in France⁵⁸, but only 7.6% in Greece⁵⁹, and even lower, 5.9%, in Norway⁵⁷. While not all aspects of

clonal diversity can be depicted in a practical mathematical model, it is possible that different circulating genotypes across geographical settings, by creating a unique ecological niche, can lead to a different net vaccine protection at the individual level, when serotype or serotype group is chosen as the common denominator. Crucially, antibiotic resistance phenotypes might need to be modelled in combination with serotype diversity to fully capture vaccination effects, as theoretically suggested⁸⁸. In fact, the local basic reproduction number R_0 is likely to be also a function of antibiotic use, antibiotic resistance patterns, and the local pool of circulating pneumococci in each setting.

Back to a serotype-specific view, the fact that the dominant non-vaccine serotypes vary across sites, and likely display different interaction with vaccine serotypes, could also affect the model readout of vaccine efficacy⁷². The present formulation treats NVT as the same entity in all studies, while in reality the majority target and non-target serotypes might compete differently in each case. One could therefore expect that when the NVT serotype pool is more similar across sites, then the vaccine efficacy estimates should be more similar. Without enough unequivocal evidence to inform serotype-specific parameters *a-priori*, one can try to evaluate this grouping effect *a-posteriori*, inspecting the aggregated model results, and a simple comparative test afforded by the data, supported this expectation (see Fig. S8). Estimating asymmetric VT-NVT competition across settings requires 3 additional parameters in the co-colonization model, as exemplified previously for the Portuguese data³⁴. Unfortunately, such model structures, applicable to all settings in a standard manner, need better temporal and especially co-colonization resolution, unfeasible with the current data. Pneumococcus ecological niche similarity and vaccine efficacy similarity across sites could be studied in various ways beyond simply serotype comparisons.

Inevitably all models, whether mathematical or experimental, are a simplification of reality, and serotype symmetry is clearly unrealistic for polymorphic pneumococci. Yet, here we see that the neutral model can reproduce efficiently the post-vaccine cross-sectional observations, despite omitting some biological realism; an indication that slight serotype differences in life-history traits do not radically alter global dynamics³⁴. Reassuringly, this model follows theoretical arguments^{89,90}, in favour of an unbiased formulation, incorporating same strain co-infection, here at the group level VT-VT and NVT-NVT, leading to a pyramid structure, in particular also advocated for pneumococcal dynamics³¹. The importance of neutral null models, which avoid mechanisms of stable co-existence for indistinguishable strains, has been previously highlighted⁴⁸. Thus, although seemingly a strong assumption, symmetry across VT and NVT often features in earlier models^{31,38,48,50}, and is in principle not limited to 2 serotype groupings. Besides its analytical advantages, the symmetry approximation allows a standardized application to temporal observations in different host populations; one step closer to their integration under a unified mechanistic framework. The definition of vaccine efficacy is also more directly interpretable in such a transmission model, reflecting the relative measure of reduced susceptibility to VT acquisition per vaccinated individual, when compared to a non-vaccinated one (in this case, in a day care setting). Despite it perhaps being a slightly biased metric, due to the unrealistic symmetry assumption, it will most likely be precise and informative, thanks to the simple model structure, enabling reproducible comparison across settings and a useful start for further exploration.

The host diversity challenge. Differences in realized net protection per vaccinated individual across settings are not unlikely. They may result from variable vaccine-induced immunity in different host backgrounds, where genetic factors of immunosuppression⁹¹, or local environmental factors such as nasopharyngeal microbiota (for respiratory pathogens)^{92,93}, might play a role. Different implementations of the immunisation programme can also be a source of heterogeneity, e.g. the effects of vaccination schedule, age of vaccination⁹⁴, or prior pathogen exposure⁹⁵. In addition, populations realistically vary along global transmission-modulating factors such as temperature, crowding, DCC attendance, age distribution, antibiotic use, and climatic conditions. Child nutritional status and interference between concurrent vaccines might also be important. All of these population-level processes, by way of interacting and often hidden nonlinear feedbacks, may exert top-down influences on the individual-level parameter readout (per-capita protection per vaccinated individual: VE), obtained when relatively simpler models are applied to data.

Revisiting the heterogeneity argument, it is important to note that a critical aspect of a dynamic model for vaccine assessment is the natural incorporation of herd immunity effects, whereby unvaccinated individuals receive indirect protection, through the immediate lower prevalence of target serotypes in vaccinated individuals. Given the data, the present model accounts for herd immunity only within the DCC environment, neglecting transmission with the outside community. Although its contribution to epidemiological DCC dynamics is not expected to be very big^{50,69,96}, with a rate of 0.5–0.6 per month, this might vary across settings⁵¹. In addition, herd immunity effects outside the DCC reservoir could affect the pace of serotype replacement observed within the DCC post-vaccination. For example, higher levels of DCC attendance in a population, coupled with high vaccination coverage in this reservoir, and a certain mixing between DCC attendees and the rest of the population could eventually speed up serotype replacement, by creating more positive feedbacks in host clusters⁹⁷.

Thus, variable DCC attendance among children across different geographical areas, might scale differently the importance of this epidemiological setting in the spread of pneumococci, and modulate differently the net herd immunity effects that feed back into the DCC environment. Such effects would necessarily shape the serotype dynamics ‘seen’ by our model. Hence, what emerges as variability in vaccine efficacy estimates, from modeling exclusively the DCC compartment, could in reality hint at different coupling, between one part and the rest of a structured population⁶⁶. Complemented datasets including DCCs, and samples from the children’s families or rest of the community, as well as from other relevant reservoirs of transmission of pneumococci, would be necessary to properly account for these heterogeneities in a dynamic model for vaccine assessment. For wider age ranges and age-stratified colonization data, matched across studies, model extensions could also incorporate temporal heterogeneities, e.g. waning of vaccine protection, by adding explicit host age structure to the epidemiological dynamics.

Next-generation analyses for next-generation vaccines. Considering all the epidemiological, demographic, and environmental factors underlying pneumococcus vaccine efficacy variation falls beyond the scope of this paper, and we cannot rule out any explanation with the present datasets. While the details of how this effect is achieved remain to be determined in the future, variable efficacies have been observed and studied in other pathogen systems with vaccination, e.g. in *Mycobacterium tuberculosis* and the BCG vaccine⁹⁸. The aim of the present study is to illustrate a new way of looking at pneumococcus vaccine protection realized in the field, by adopting a setting-specific lens of analysis under a unified dynamic model. Such retrospective frameworks post-licensure could offer an additional tool to identify and quantify first-order differences in vaccine efficacy across host populations, and then lead to detailed and systematic exploration of second-order causal links and data-driven model extensions. Once the dominant sources of variability across communities are uncovered, it would be interesting to test whether they apply also to other, more recent vaccines.

Next-generation vaccines for protection against bacterial pathogens remain an area of active research⁶, for which an accurate interpretation and forecast of intervention outcomes are needed. Whenever possible, their efficacy should be evaluated in a context-specific manner, and re-assessed via multiple approaches and available data. Linking pre-licensure vaccine studies with follow-up observational surveys, and standardized dynamic models, will yield a better understanding of the underlying mechanisms of vaccine protection, their sources of variability, and downstream effects on population-level dynamics. Site-specific and comparative analyses should not be seen as questioning the importance or public health benefit of vaccination programs, but rather as a way to learn from the differences between settings⁹⁹, and to use this understanding for better disease control and empowered vaccines globally.

Methods

Pneumococcus transmission model pre-vaccine. The dynamics of pneumococcus colonization and co-colonization in young children is described through a susceptible-infected-susceptible (SIS) epidemiological model, tracking the proportion of: susceptible hosts, S , hosts colonized by vaccine serotypes I_V , hosts colonized by non-vaccine serotypes, I_N , and co-colonized hosts I_{VV} , I_{NN} , and I_{VN} with vaccine serotypes, non-vaccine serotypes or a combination of the two, respectively. Notice that $S + \sum I = 1$. Upon exposure, a susceptible host can become a single carrier of a VT or NVT pneumococcal serotype. The forces of infection (FOI) are: $\lambda_V(t) = \beta(I_V + I_{VV} + I_{VN}/2)$ for VT, and $\lambda_N(t) = \beta(I_N + I_{NN} + I_{VN}/2)$ for NVT. Single and dual carriers contribute equally to the FOI, and hosts carrying two serotypes transmit either with equal probability. β is the per-capita transmission rate. Single carriers can acquire an additional serotype at a reduced rate, modified by a relative competition coefficient $k \in (0, 1)$ between the resident and the challenge serotype. Pneumococcal carriage is cleared at rate γ , regardless of whether hosts are single or dual carriers. Susceptibles are recruited at constant rate μ , equal to the per-capita departure rate from the given age class. Immune memory to either set of serotypes is assumed negligible in the time-frame of day care attendance, while serotype-specific immunity is relatively minor compared to the pool of circulating serotypes. Homogeneous mixing is assumed among day-care attendees, and transmission dynamics with the outside community are neglected.

Calibrating the model to serotype prevalences before vaccination. Under this neutral model, prior to any interventions, vaccine and non-vaccine serotypes coexist at neutrally stable endemic equilibria³⁴, which satisfy:

$$\begin{aligned} S^* &= \frac{\mu + \gamma}{\beta}, & I_1^* &= I_V^* + I_N^* = \frac{(\beta - \gamma - \mu)(\gamma + \mu)}{\beta[\gamma + \mu + k(\beta - \gamma - \mu)]}, \\ I_2^* &= I_{VV}^* + I_{NN}^* + I_{VN}^* = \frac{(\beta - \gamma - \mu)^2 k}{\beta[\gamma + \mu + k(\beta - \gamma - \mu)]}, \end{aligned} \quad (2)$$

for the proportion of hosts susceptible at equilibrium, singly-colonized with one serotype, and co-colonized with two serotypes, respectively, where $\beta > \gamma + \mu$, equivalently $R_0 > 1$ ⁶⁴. When considering serotype distribution, eq. (2) show that the overall pathogen prevalence can be flexibly divided between serotypes in many ways. In a multi-type permutable system, there is however one constraint at a global endemic equilibrium: as soon as the prevalence of one serotype group in single carriage is fixed, (e.g. VT) all other equilibrium prevalences follow as direct functions of it, enabling correlated hierarchies. Thus, although serotypes behave broadly independently, they are coupled through competition for available hosts. Under this model, if the pre-vaccination data are assumed to reflect stationarity, one can infer competition and transmission parameters by fitting endemic prevalence and relative prevalences of single and double carriage. Using $R_0 = \beta/(\gamma + \mu)$, and $S^* = 1/R_0$, $I_2^* = k(R_0 - 1)I_1^*$ and $I_1^* + I_2^* = 1 - 1/R_0$, at the theoretical endemic equilibrium, the most likely estimates for basic reproduction number and serotype interaction parameter k can be obtained from data as:

$$\hat{R}_0 = \frac{1}{S^*}, \quad \hat{k} = \frac{I_2^*}{I_1^*} \frac{1}{\frac{1}{S^*} - 1}. \quad (3)$$

If, besides overall carriage, the relative prevalence of one serotype group is reported at stationarity, e.g. for VT, $P_V^* = I_V^* + I_{VV}^* + I_{VN}^*/2$, one can use the above equations in another way, namely to infer the expected proportions of hosts in all other epidemiological classes. In particular, plugging-in the 'observed' P_V , and a known competition parameter k one gets:

$$I_V^* = \frac{P_V^*}{1 + k(R_0 - 1)} \tag{4}$$

The other prevalences I_{VV}^* , I_N^* , I_{NN}^* and I_{VN}^* follow from I_V^* iteratively:

$$I_N^* = \frac{R_0[1 - I_V^*(1 + k(R_0 - 1))]}{R_0[1 + k(R_0 - 1)]} \tag{5}$$

$$I_{VV}^* = (I_V^*)^2 k R_0 [1 + k(R_0 - 1)] \tag{6}$$

$$I_{NN}^* = (I_N^*)^2 k R_0 [1 + k(R_0 - 1)] \tag{7}$$

$$I_{VN}^* = 1 - \frac{1}{R_0} - (I_V^* + I_N^* + I_{VV}^* + I_{NN}^*) \tag{8}$$

From a practical point of view, eqs (4–8) represent an analytical trick that can be used to inform *full* initial conditions of a transmission model, in those cases where explicit co-colonization data may be unobserved. In fact, in most datasets considered in this study, except for Norway and Portugal, co-colonization data were not quantified in detail, or not at all. To account for the possibility of underlying co-colonization, and to appropriately initialize model variables in these cases, I treated the pre-vaccination prevalences in France, Hungary, Hong-Kong and Greece, as if they were the endemic prevalences under the above constraints, starting with P_V and fixing k , and subsequently unfolding all epidemiological variables from P_V as delineated above.

Dynamic model with vaccination. In the presence of a vaccination programme, the number of model compartments doubles to 12, but the dynamical equations preserve their basic structure. A fraction ρ of all susceptibles is vaccinated at birth. The relative reduction in susceptibility to VT of vaccinated hosts (superscript 1) is given by $w = 1 - VE \in (0, 1)$, compared to those non-vaccinated. It is assumed that vaccine protection acts both at primary and secondary acquisition of a vaccine serotype. The equations are similar for the non-vaccinated hosts (superscript 0), who replenish the susceptible pool, S^0 , at rate $\mu(1 - \rho)$ and experience no protection against pneumococcus. The FOIs are now given by: $\lambda_V(t) = \beta(I_V^1 + I_{VV}^1 + I_{VN}^1/2 + I_V^0 + I_{VV}^0 + I_{VN}^0/2)$ for VT, and similarly $\lambda_N(t) = \beta(I_N^1 + I_{NN}^1 + I_{VN}^1/2 + I_N^0 + I_{NN}^0 + I_{VN}^0/2)$, for NVT. We still have $S^0 + \sum I^0 + S^1 + \sum I^1 = 1$. Full model equations are given in Supplementary material S1.

Linking the model to data: statistical inference. This model is fitted to the cross-sectional prevalence data for children attending day care reported in 6 different countries (Table 2). In cases where full resolution of single and double carriage is available (Portugal and Norway), the explicit pre-vaccine data are directly input as initial conditions of the model, which is subsequently run for each parameter combination. In particular, for Portugal we have for 2001 the following epidemiological observations: [97 77 66 7 5 18] for $[n_S, x_V, x_N, x_{VV}, x_{VN}, x_{VN}]$ (numbers of children in each class), and for 2007: [70 13 125 0 10 2] among vaccinated children, and [60 21 117 1 17 13] among non-vaccinated ones. For Norway, we have the following epidemiological resolution for 2006: [136, 188, 226, 9, 17, 30] for $[n_S, x_V, x_N, x_{VV}, x_{NN}, x_{VN}]$, and for 2008: [118, 82, 328, 2, 36, 26] at the aggregated level over all children in the study of each year. In these two cases, $\theta = (R_0, k, w)$ can be estimated accounting for information across all years of the survey, without needing the stationarity assumption for pre-vaccine observations.

In epidemiological studies where only overall prevalence and VT prevalence are reported, the pre-vaccine data are interpreted as a global equilibrium in the absence of intervention, and then, using Eqs (4–8), I derive explicitly initial conditions relative to all classes required by the model, for each value of R_0 . In these cases, joint estimation of $\theta = (R_0, w)$ is performed, but R_0 is more tightly coupled to the initial observations pre-vaccination. The equations are solved numerically for different times of follow-up in each study. The dynamic variables $S^1(t), I_V^1(t), \dots, S^0(t), I_V^0(t)$ are then compared to the epidemiological observations post-PCV7. When only data on prevalence of vaccine serotypes are available: $n_V = x_V + x_{VV} + x_{VN}/2$ and $n_N = x_N + x_{NN} + x_{VN}/2$, model predictions for $P_V(t) = I_V^1 + I_{VV}^1 + I_{VN}^1/2 + I_V^0 + I_{VV}^0 + I_{VN}^0/2$ are used for the fitting, instead of all 6 epidemiological variables. Accounting for vaccination coverage and sample sizes across studies, parameter inference is performed using a multinomial likelihood under a Bayesian framework³⁴. The likelihood of the data (France, Greece, Hungary, Hong-Kong) is given by:

$$L(n_S, n_V, n_N | \theta) = \prod_i \left(\frac{n!}{n_S! n_V! n_N!} S^{n_S} P_V^{n_V} P_N^{n_N} \right) (t_i) \tag{9}$$

where the model variables satisfy: $S(t) + P_V(t) + P_N(t) = 1$ and the observed data satisfy: $n_S(t) + n_V(t) + n_N(t) = n(t)$. The expression is similar for Portugal and Norway, but we account for the finer epidemiological resolution of the host population afforded by the data. We used uniform priors in [1, 6] and [0, 1] for R_0 and w/k respectively. As all children in the studies were day care attendees, we assumed a monthly clearance rate γ equal to 0.7. This corresponds to a mean duration of a carriage episode of about 5.7 weeks, as previously documented for this age group³⁸. The deterministic model skeleton is used to obtain temporal trajectories for expected prevalence of each carriage type among day-care attendees, which are then integrated within the sampling process, as done in an

earlier study³⁴. Concerning the entry rate μ , I considered the mean duration of DCC attendance per child per setting (Table 1), accounting for the proportions in different age groups and the maximal age of daycare attendees. This resulted in these values of $\mu = 1/\text{durationDCC}$ across settings: $\mu = 0.042$ (Portugal), 0.031 (Norway), 0.058 (France), 0.037 (Greece), 0.055 (Hungary) and 0.042 (Hong Kong). The results for model-fitting under a different assumption for μ (1/mean age), confirmed similar values for vaccine efficacy, preserving in general the ranking between settings and are detailed in Table S2 and Fig. S4.

The initial conditions for the post-vaccination dynamics are split between vaccinated and non-vaccinated children in the proportions ρ and $1 - \rho$, for initial coverage. When a different coverage is known for each year of follow-up (Table 1), the system is numerically integrated stepwise from year to year, using a different ρ , and initializing the system for the next round of integration at the solution obtained up to the last time period. Parameter estimation was performed using the mcmcstat package⁶³ in Matlab. Convergence of two independent MCMC chains to the stationary posterior distributions was assessed using the Gelman-Rubin statistic¹⁰⁰.

References

- O'Brien, K. L. *et al.* Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *The Lancet* **374**, 893–902 (2009).
- Bogaert, D., de Groot, R. & Hermans, P. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *The Lancet infectious diseases* **4**, 144–154 (2004).
- Park, I. H. *et al.* Discovery of a new capsular serotype (6c) within serogroup 6 of *Streptococcus pneumoniae*. *Journal of Clinical Microbiology* **45**, 1225–1233 (2007).
- Dagan, R. *et al.* Serum serotype-specific pneumococcal anticapsular immunoglobulin g concentrations after immunization with a 9-valent conjugate pneumococcal vaccine correlate with nasopharyngeal acquisition of pneumococcus. *Journal of Infectious Diseases* **192**, 367–376 (2005).
- Millar, E. V. *et al.* Anticapsular serum antibody concentration and protection against pneumococcal colonization among children vaccinated with 7-valent pneumococcal conjugate vaccine. *Clinical infectious diseases* **44**, 1173–1179 (2007).
- Feldman, C. & Anderson, R. Review: current and new generation pneumococcal vaccines. *Journal of Infection* **69**, 309–325 (2014).
- Black, S. *et al.* Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. *The Pediatric infectious disease journal* **19**, 187–195 (2000).
- Klugman, K. P. *et al.* A trial of a 9-valent pneumococcal conjugate vaccine in children with and those without HIV infection. *New England Journal of Medicine* **349**, 1341–1348 (2003).
- O'Brien, K. L. *et al.* Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial. *Lancet* **362**, 355 (2003).
- Cutts, F. *et al.* Efficacy of nine-valent pneumococcal conjugate vaccine against pneumonia and invasive pneumococcal disease in The Gambia: randomised, double-blind, placebo-controlled trial. *The Lancet* **365**, 1139–1146 (2005).
- Obaro, S. K., Adegbola, R., Banya, W. & Greenwood, B. Carriage of pneumococci after pneumococcal vaccination. *The Lancet* **348**, 271–272 (1996).
- O'Brien, K. L. *et al.* Effect of pneumococcal conjugate vaccine on nasopharyngeal colonization among immunized and unimmunized children in a community-randomized trial. *Journal of Infectious Diseases* **196**, 1211–1220 (2007).
- Huang, S. S. *et al.* Post-pcv7 changes in colonizing pneumococcal serotypes in 16 massachusetts communities, 2001 and 2004. *Pediatrics* **116**, e408–e413 (2005).
- Huang, S. S. *et al.* Continued impact of pneumococcal conjugate vaccine on carriage in young children. *Pediatrics* **124**, e1–e11 (2009).
- Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nature genetics* **45**, 656–663 (2013).
- Brugger, S., Frey, P., Aebi, S., Hinds, J. & Muhlemann, K. Multiple colonization with *S. pneumoniae* before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. *PLoS One* **5**, e11638 (2010).
- Flasche, S. *et al.* Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in england: A cross-sectional study. *PLoS Med* **8**, e1001017 (2011).
- Miller, E., Andrews, N. J., Waight, P. A., Slack, M. P. & George, R. C. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in england and wales: an observational cohort study. *The Lancet infectious diseases* **11**, 760–768 (2011).
- Spijkerman, J. *et al.* Carriage of streptococcus pneumoniae 3 years after start of vaccination program, the netherlands. *Emerg Infect Dis* **17**, 584–591 (2011).
- Weinberger, D. M., Malley, R. & Lipsitch, M. Serotype replacement in disease after pneumococcal vaccination. *The Lancet* **378**, 1962–1973 (2011).
- Feikin, D. R. *et al.* Serotype-specific changes in invasive pneumococcal disease after pneumococcal conjugate vaccine introduction: a pooled analysis of multiple surveillance sites. *PLoS Med* **10**, e1001517 (2013).
- Weinberger, D. M. *et al.* Using pneumococcal carriage data to monitor postvaccination changes in invasive disease. *American journal of epidemiology* **178**, 1488–1495 (2013).
- Flasche, S., de Waroux, O. L. P., O'Brien, K. L. & Edmunds, W. J. The serotype distribution among healthy carriers before vaccination is essential for predicting the impact of pneumococcal conjugate vaccine on invasive disease. *PLoS Comput Biol* **11**, e1004173 (2015).
- Fletcher, M. A. & Fritzell, B. Pneumococcal conjugate vaccines and otitis media: an appraisal of the clinical trials. *International journal of otolaryngology* **2012** (2012).
- Halloran, M. E., Struchiner, C. J. & Longini, I. M. Study designs for evaluating different efficacy and effectiveness aspects of vaccines. *American journal of epidemiology* **146**, 789–803 (1997).
- Halloran, M. E., Haber, M., Longini, I. M. & Struchiner, C. J. Direct and indirect effects in vaccine efficacy and effectiveness. *American Journal of Epidemiology* **133**, 323–331 (1991).
- Rinta-Kokko, H., Dagan, R., Givon-Lavi, N. & Auranen, A. Estimation of vaccine efficacy against acquisition of pneumococcal carriage. *Vaccine* **27**, 3831–3837 (2009).
- Gjini, E. & Gomes, M. G. M. Expanding vaccine efficacy estimation with dynamic models fitted to cross-sectional prevalence data post-licensure. *Epidemics* **14**, 71–82 (2016).
- Eskola, J. *et al.* Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *New England Journal of Medicine* **344**, 403–409 (2001).
- Melegaro, A. *et al.* Dynamic models of pneumococcal carriage and the impact of the heptavalent pneumococcal conjugate vaccine on invasive pneumococcal disease. *BMC Infectious Diseases* **10**, 90 (2010).
- Choi, Y. H. *et al.* 7-valent pneumococcal conjugate vaccination in england and wales: is it still beneficial despite high levels of serotype replacement. *PLoS One* **6**, e26190–e26190 (2011).

32. Cobey, S. & Lipsitch, M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science (New York, NY)* **335**, 1376–1380 (2012).
33. Flasche, S. *et al.* The impact of specific and non-specific immunity on the ecology of streptococcus pneumoniae and the implications for vaccination. *Proceedings of the Royal Society B: Biological Sciences* **280** (2013).
34. Gjini, E., Valente, C., Sa-Leao, R. & Gomes, M. How direct competition shapes coexistence and vaccine effects in multi-strain pathogen systems. *Journal of Theoretical Biology* (2015).
35. Schönberger, K., Kirchgässner, K., Riedel, C. & von Kries, R. Effectiveness of 2 + 1 pcv7 vaccination schedules in children under 2 years: A meta-analysis of impact studies. *Vaccine* **31**, 5948–5952 (2013).
36. Andrews, N. J. *et al.* Serotype-specific effectiveness and correlates of protection for the 13-valent pneumococcal conjugate vaccine: a postlicensure indirect cohort study. *The Lancet Infectious Diseases* **14**, 839–846 (2014).
37. Givon-Lavi, N., Dagan, R., Fraser, D., Yagupsky, P. & Porat, N. Marked differences in pneumococcal carriage and resistance patterns between day care centers located within a small area. *Clinical infectious diseases* **29**, 1274–1280 (1999).
38. Pessoa, D. *et al.* Comparative analysis of Streptococcus pneumoniae transmission in Portuguese and Finnish day-care centres. *BMC Infectious Diseases* **13**, 180 (2013).
39. Leino, T., Hoti, F., Syrjänen, R., Tanskanen, A. & Auranen, K. Clustering of serotypes in a longitudinal study of streptococcus pneumoniae carriage in three day care centres. *BMC infectious diseases* **8**, 1 (2008).
40. Fraser, C., Hanage, W. P. & Spratt, B. G. Neutral microepidemic evolution of bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1968–1973 (2005).
41. Abdullahi, O. *et al.* Rates of acquisition and clearance of pneumococcal serotypes in the nasopharynxes of children in kilifi district, kenya. *Journal of Infectious Diseases* **206**, 1020–1029 (2012).
42. Domenech De Celles, M. *et al.* Intrinsic epidemicity of Streptococcus pneumoniae depends on strain serotype and antibiotic susceptibility pattern. *Antimicrob Agents Chemother* **55**, 5255–5261 (2011).
43. Tigoï, C. *et al.* Rates of acquisition of pneumococcal colonization and transmission probabilities, by serotype, among newborn infants in kilifi district, kenya. *Clin Infect Dis.* **55**, 180–188 (2012).
44. Lipsitch, M. *et al.* Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. *Epidemiology* **23**, 510–519 (2012).
45. Cauchemez, S. *et al.* S. pneumoniae transmission according to inclusion in conjugate vaccines: Bayesian analysis of a longitudinal follow-up in schools. *BMC Infect Dis* **6**, 14 (2006).
46. Erasto, P. *et al.* Modelling multi-type transmission of pneumococcal carriage in bangladeshi families. *Epidemiol Infect* **138**, 861–872 (2010).
47. Hill, P. C. *et al.* Nasopharyngeal carriage of Streptococcus pneumoniae in Gambian infants: a longitudinal study. *Clinical infectious diseases* **46**, 807–814 (2008).
48. Lipsitch, M., Colijn, C., Cohen, T., Hanage, W. P. & Fraser, C. No coexistence for free: neutral null models for multistrain pathogens. *Epidemics* **1**, 2–13 (2009).
49. Nurhonen, M., Cheng, A. C. & Auranen, K. Pneumococcal transmission and disease in silico: a microsimulation model of the indirect effects of vaccination. *PloS one* **8**, e56079 (2013).
50. Numminen, E., Cheng, L., Gyllenberg, M. & Corander, J. Estimating the transmission dynamics of Streptococcus pneumoniae from strain prevalence data. *Biometrics* **69**, 748–757 (2013).
51. Huang, S. S., Finkelstein, J. A. & Lipsitch, M. Modeling community- and individual-level effects of child-care center attendance on pneumococcal carriage. *Clinical Infectious Diseases* **40**, 1215–1222 (2005).
52. Shrestha, S. *et al.* Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. *Science translational medicine* **5**, 191ra84–191ra84 (2013).
53. Numminen, E. *et al.* Climate induces seasonality in pneumococcal transmission. *Scientific reports* **5** (2015).
54. Sá-Leão, R. *et al.* Changes in pneumococcal serotypes and antibiotypes carried by vaccinated and unvaccinated day-care centre attendees in portugal, a country with widespread use of the seven-valent pneumococcal conjugate vaccine. *Clinical Microbiology and Infection* **15**, 1002–1007 (2009).
55. Valente, C. *et al.* Decrease in pneumococcal co-colonization following vaccination with the seven-valent pneumococcal conjugate vaccine. *PLoS ONE* **7**, e30235 (2012).
56. Vestrheim, D. F., Høiby, E. A., Aaberge, I. S. & Caugant, D. A. Phenotypic and genotypic characterization of streptococcus pneumoniae strains colonizing children attending day-care centers in norway. *Journal of clinical microbiology* **46**, 2508–2518 (2008).
57. Vestrheim, D. F., Høiby, E. A., Aaberge, I. S. & Caugant, D. A. Impact of a pneumococcal conjugate vaccination program on carriage among children in norway. *Clinical and Vaccine Immunology* **17**, 325–334 (2010).
58. Dunais, B., Bruno-Bazureault, P., Carsenti-Dellamonica, H., Touboul, P. & Pradier, C. A decade-long surveillance of nasopharyngeal colonisation with streptococcus pneumoniae among children attending day-care centres in south-eastern france: 1999–2008. *European journal of clinical microbiology & infectious diseases* **30**, 837–843 (2011).
59. Grivea, I. N., Tsantouli, A. G., Michoula, A. N. & Syrogiannopoulos, G. A. Dynamics of streptococcus pneumoniae nasopharyngeal carriage with high heptavalent pneumococcal conjugate vaccine coverage in central greece. *Vaccine* **29**, 8882–8887 (2011).
60. Tóthpál, A. *et al.* A marked shift in the serotypes of pneumococci isolated from healthy children in szeged, hungary, over a 6-year period. *Acta microbiologica et immunologica Hungarica* **58**, 239–246 (2011).
61. Tóthpál, A. *et al.* Radical serotype rearrangement of carried pneumococci in the first 3 years after intensive vaccination started in hungary. *European journal of pediatrics* **1–9** (2014).
62. Ho, P.-L. *et al.* Changes in nasopharyngeal carriage and serotype distribution of antibiotic-resistant streptococcus pneumoniae before and after the introduction of 7-valent pneumococcal conjugate vaccine in hong kong. *Diagnostic microbiology and infectious disease* **71**, 327–334 (2011).
63. Haario, H., Laine, M., Mira, A. & Saksman, E. DRAM: Efficient adaptive MCMC. *Statistics and Computing* **16**, 339–354 (2006).
64. Heesterbeek, J. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation, vol. 5 (Wiley, 2000).
65. Vynnycky, E. & White, R. *An introduction to infectious disease modelling* (Oxford University Press, 2011).
66. Scarbrough Lefebvre, C. D., Terlinden, A. & Standaert, B. Dissecting the indirect effects caused by vaccines into the basic elements. *Human vaccines & immunotherapeutics* **11**, 2142–2157 (2015).
67. Fraser, C., Hanage, W. P. & Spratt, B. G. Recombination and the nature of bacterial speciation. *Science* **315**, 476–480 (2007).
68. Feil, E. J., Smith, J. M., Enright, M. C. & Spratt, B. G. Estimating recombinational parameters in streptococcus pneumoniae from multilocus sequence typing data. *Genetics* **154**, 1439–1450 (2000).
69. Auranen, K., Mehtala, J., Tanskanen, A. & Kalltoft, S. M. Between-strain competition in acquisition and clearance of pneumococcal carriage: Epidemiologic evidence from a longitudinal study of day-care children. *American Journal of Epidemiology* **171**, 169–176 (2010).
70. Chalub, F. A. & Souza, M. O. Discrete and continuous sis epidemic models: a unifying approach. *Ecological Complexity* **18**, 83–95 (2014).
71. Gjini, E. & Madec, S. A slow-fast dynamic decomposition links neutral and non-neutral coexistence in interacting multi-strain pathogens. *Theoretical Ecology* **1–13** (2016).

72. Gjini, E. & Gomes, M. G. M. Expanding vaccine efficacy estimation with dynamic models fitted to cross-sectional prevalence data post-licensure. *Epidemics* (2015).
73. De Waroux, O. L. P., Flasche, S., Prieto-Merino, D., Goldblatt, D. & Edmunds, W. J. The efficacy and duration of protection of pneumococcal conjugate vaccines against nasopharyngeal carriage: a meta-regression model. *The Pediatric infectious disease journal* **34**, 858–864 (2015).
74. Lipsitch, M. *et al.* Strain characteristics of streptococcus pneumoniae carriage and invasive disease isolates during a cluster-randomized clinical trial of the 7-valent pneumococcal conjugate vaccine. *Journal of Infectious Diseases* **196**, 1221–1227 (2007).
75. Moore, M. R. *et al.* Impact of a conjugate vaccine on community-wide carriage of nonsusceptible streptococcus pneumoniae in alaska. *Journal of Infectious Diseases* **190**, 2031–2038 (2004).
76. Sleijffers, A., Garssen, J. & Van Loveren, H. Ultraviolet radiation, resistance to infectious diseases, and vaccination responses. *Methods* **28**, 111–121 (2002).
77. Goettsch, W. *et al.* Effects of ultraviolet-b exposure on the resistance to listeria monocytogenes in the rat. *Photochemistry and photobiology* **63**, 672–679 (1996).
78. Ryan, L. K. *et al.* Exposure to ultraviolet radiation enhances mortality and pathology associated with influenza virus infection in mice. *Photochemistry and photobiology* **72**, 497–507 (2000).
79. Ward, M., Sailstad, D., Andrews, D., Boykin, E. & Selgrade, M. Effects of ultraviolet radiation (uvr) on the allergic respiratory responses of balb/c mice to a fungal allergen. *Toxicol. Sci. S* **60**, 207 (2001).
80. Tobón, G. J., Youinou, P. & Saraux, A. The environment, geo-epidemiology, and autoimmune disease: rheumatoid arthritis. *Journal of autoimmunity* **35**, 10–14 (2010).
81. Lam, O., Wheeler, J. & Tang, C. M. Thermal control of virulence factors in bacteria: A hot topic. *Virulence* **5**, 852–862 (2014).
82. Woolhouse, M. E. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences* **94**, 338–342 (1997).
83. Gomes, M. G. M. *et al.* End tb strategy: the need to reduce risk inequalities. *BMC infectious diseases* **16**, 1 (2016).
84. Gini, C. On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series* **208**, 73–79 (1936).
85. White, M. T., Griffin, J. T., Drakeley, C. J. & Ghani, A. C. Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. *Malaria journal* **9**, 82 (2010).
86. Gomes, M. G. M. *et al.* How host heterogeneity governs tuberculosis reinfection? *Proceedings of the Royal Society of London B: Biological Sciences* **2473–2478** (2012).
87. Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimates of the changing age-burden of plasmodium falciparum malaria disease in sub-saharan africa. *Nature communications* **5** (2014).
88. Mitchell, P. K., Lipsitch, M. & Hanage, W. P. Carriage burden, multiple colonization and antibiotic pressure promote emergence of resistant vaccine escape pneumococci. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **370**, 20140342 (2015).
89. van Baalen, M. & Sabelis, M. W. The dynamics of multiple infection and the evolution of virulence. *American Naturalist* **881–910** (1995).
90. Alizon, S. Co-infection and super-infection models in evolutionary epidemiology. *Interface focus* **3**, 20130031 (2013).
91. Yoshikawa, T. *et al.* Susceptibility to effects of uvb radiation on induction of contact hypersensitivity as a risk factor for skin cancer in humans. *Journal of Investigative Dermatology* **95**, 530–536 (1990).
92. Valdez, Y., Brown, E. M. & Finlay, B. B. Influence of the microbiota on vaccine effectiveness. *Trends in immunology* **35**, 526–537 (2014).
93. Biesbroek, G. *et al.* Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. *Emerg Infect Dis* **20**, 201–10 (2014).
94. Park, D. E. *et al.* The differential impact of coadministered vaccines, geographic region, vaccine product and other covariates on pneumococcal conjugate vaccine immunogenicity. *The Pediatric infectious disease journal* **33**, S130–S139 (2014).
95. Dagan, R., Givon-Lavi, N., Greenberg, D., Fritzell, B. & Siegrist, C.-A. Nasopharyngeal carriage of streptococcus pneumoniae shortly before vaccination with a pneumococcal conjugate vaccine causes serotype-specific hypo-responsiveness in early infancy. *Journal of Infectious Diseases* **201**, 1570–1579 (2010).
96. Hoti, F., Erasto, P., Leino, T. & Auranen, K. Outbreaks of Streptococcus pneumoniae carriage in day care cohorts in Finland—implications for elimination of transmission. *BMC Infectious Diseases* **9**, 102 (2009).
97. Reluga, T. C. & Shim, E. Population viscosity suppresses disease emergence by preserving local herd immunity. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20141901 (2014).
98. Fine, P. E. Variation in protection by bcg: implications of and for heterologous immunity. *The Lancet* **346**, 1339–1345 (1995).
99. Yusuf, S. & Wittes, J. Interpreting geographic variations in results of randomized, controlled trials. *New England Journal of Medicine* **375**, 2263–2271 (2016).
100. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639 (2002).

Acknowledgements

The author would like to thank Brigitte Dunais for her kind sharing of the French data and Didrik F. Vestheim and Anneke Steens for their helpful assistance on the Norwegian data in September 2014. The author also thanks Louise Matthews for providing useful feedback on previous versions of the manuscript.

Author Contributions

E.G. conceived and conducted the study, analysed the results and wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-02955-y

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017