RESEARCH ARTICLE

Epidemiological consequences of enduring strain-specific immunity requiring repeated episodes of infection

Rebecca H. Chisholm[1,2], Nikki Sonenberg[2¤], Jake A. Lacey[3], Malcolm I. McDonald[4], Manisha Pandey[5], Mark R. Davies[6], Steven Y. C. Tong[3,7,8], Jodie McVernon[9], Nicholas Geard[9,10]*

1 Department of Mathematics and Statistics, La Trobe University, Melbourne, Victoria, Australia, **2** Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia, **3** Doherty Department University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Victoria, Australia, **4** Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Queensland, Australia, **5** Institute for Glycomics, Gold Coast Campus, Griffith University, Brisbane, Queensland, Australia, **6** Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia, **7** Victorian Infectious Diseases Service, The Royal Melbourne Hospital, at the Peter Doherty Institute for Infection and Immunity, Victoria, Australia, **8** Menzies School of Health Research, Charles Darwin University, Darwin, Australia, **9** Victorian Infectious Diseases Reference Laboratory Epidemiology Unit at the Peter Doherty Institute for Infection and Immunity, The University of Melbourne and Royal Melbourne Hospital, Victoria, Australia, **10** School of Computing and Information Systems, Melbourne School of Engineering, The University of Melbourne, Melbourne, Victoria, Australia

¤ Current address: The Alan Turing Institute, London, United Kingdom
* nicholas.geard@unimelb.edu.au

Abstract

Group A *Streptococcus* (GAS) skin infections are caused by a diverse array of strain types and are highly prevalent in disadvantaged populations. The role of strain-specific immunity in preventing GAS infections is poorly understood, representing a critical knowledge gap in vaccine development. A recent GAS murine challenge study showed evidence that sterilising strain-specific and enduring immunity required two skin infections by the same GAS strain within three weeks. This mechanism of developing enduring immunity may be a significant impediment to the accumulation of immunity in populations. We used an agent-based mathematical model of GAS transmission to investigate the epidemiological consequences of enduring strain-specific immunity developing only after two infections with the same strain within a specified interval. Accounting for uncertainty when correlating murine timeframes to humans, we varied this maximum inter-infection interval from 3 to 420 weeks to assess its impact on prevalence and strain diversity, and considered additional scenarios where no maximum inter-infection interval was specified. Model outputs were compared with longitudinal GAS surveillance observations from northern Australia, a region with endemic infection. We also assessed the likely impact of a targeted strain-specific multivalent vaccine in this context. Our model produced patterns of transmission consistent with observations when the maximum inter-infection interval for developing enduring immunity was 19 weeks. Our vaccine analysis suggests that the leading multivalent GAS vaccine may have limited impact on the prevalence of GAS in populations in northern Australia if strain-specific

immunity requires repeated episodes of infection. Our results suggest that observed GAS epidemiology from disease endemic settings is consistent with enduring strain-specific immunity being dependent on repeated infections with the same strain, and provide additional motivation for relevant human studies to confirm the human immune response to GAS skin infection.

Author summary

Group A *Streptococcus* (GAS) is a ubiquitous bacterial pathogen that exists in many distinct strains, and is a major cause of death and disability globally. Vaccines against GAS are under development, but their effective use will require better understanding of how immunity develops following infection. Evidence from an animal model of skin infection suggests that the generation of enduring strain-specific immunity requires two infections by the same strain within a short time frame. It is not clear if this mechanism of immune development operates in humans, nor how it would contribute to the persistence of GAS in populations and affect vaccine impact. We used a mathematical model of GAS transmission, calibrated to data collected in an Indigenous Australian community, to assess whether this mechanism of immune development is consistent with epidemiological observations, and to explore its implications for the impact of a vaccine. We found that it is plausible that repeat infections are required for the development of immunity in humans, and illustrate the difficulties associated with achieving sustained reductions in disease prevalence with a vaccine.

Introduction

The development of immunological memory following infection or vaccination against a particular pathogen enables a more rapid and enhanced immune response during subsequent infections. The characteristics of this immunological memory at an individual host level—such as the degree or duration of immune protection against subsequent pathogen encounters—impact epidemiological dynamics at the host population level [1, 2].

 Routine vaccination programs targeting pathogens comprised of a single serotype (*i.e.*, one immunologically-equivalent strain), such as the mumps and measles viruses, inhibit sustained transmission because they result in the accumulation of hosts with enduring immunological memory (herd immunity) effective against all pathogen genotypes [3, 4]. For pathogens with multiple serotypes (*i.e.*, multi-strain pathogens), such as *Neisseria meningitidis* [5], poliovirus [6], *Streptococcus pneumoniae* [7] and dengue virus [8], infection by one strain may lead to an immune response that is strain-specific, providing less, if any, protection against other strains (cross-strain immunity). As a result, the link between an individual's immune response and the accumulation of herd immunity at the host population-level can be more complex for multi-strain pathogens, posing challenges for understanding their transmission and for control [1, 2, 9–15].

 An important human pathogen with very high strain diversity is group A *Streptococcus* (GAS), which, globally, is comprised of over 230 molecular sequence types [16] and over 290 distinct genotypes [17]. GAS generally causes infections of the skin or throat that are mild and easily treated. However, mild GAS infection can also lead to more serious invasive and immune-mediated disease with high mortality rates [18]. Hence, populations with high rates

of mild GAS infections tend to also suffer from high rates of invasive disease and immune sequelae, such as acute rheumatic fever, rheumatic heart disease and acute post-streptococcal glomerulonephritis [18]. These GAS "hyper-endemic populations" also tend to have much higher strain diversity compared to populations with a low prevalence of GAS [19]. For example, dozens of strains of GAS are reported to co-circulate in the Indigenous communities of tropical northern Australia, where the median prevalence of GAS skin infections is 45% (IQR 34.0–49.2%) in children, and the incidence of acute rheumatic fever is among the highest reported in the world [18, 20–24]. There is a lack of population prevalence data of GAS skin infection in non hyper-endemic populations [20]. However, in the US, where rheumatic heart disease prevalence is estimated to be amongst the lowest levels in the world [25], just three strains accounted for over 50% of GAS throat isolates collected from children over a seven year period [26].

Despite the high global burden of GAS disease [18], currently there is no licensed GAS vaccine, although there are a number in the vaccine pipeline [27, 28]. A critical knowledge gap in GAS vaccine development is our limited understanding of how strain-specific immunity might prevent GAS infection (particularly skin infection) and, in turn, shape patterns of transmission across different populations. Epidemiological studies indicate that GAS skin infection is much less frequent in adults than in children [20, 24, 29, 30], suggesting that people may be able to acquire enduring immunity to particular GAS strains following skin infection. However, if enduring strain-specific immunity to GAS is possible, the high rates of repeat skin infections observed in children in hyper-endemic regions [30–32] suggest that it is slow to develop. Moreover, an association between the age-related immunity to GAS and the acquisition of GAS specific antibodies suggest the need for repeated GAS exposures for enduring immunity [33]. A recent study in mice showed evidence that sterilising strain-specific and enduring immunity required two skin infections by the same GAS strain within three weeks [34]. A single infection, or two infections by the same strain that occurred greater than three weeks apart did not result in the generation of memory B cells, but rather only short-lived strain-specific immunity. An analogous mechanism of acquiring enduring strain-specific immunity from GAS skin infection in humans may be a significant impediment to the accumulation of herd immunity, particularly in populations with high numbers of circulating strains.

In this work we develop an agent-based mathematical model that simulates the transmission of multiple strains of GAS in a population where hosts can only acquire enduring immunity protecting against reinfection by a particular strain if they experience two repeated episodes of infection by this strain within a specified time interval. To the best of our knowledge, this is the first time a transmission model of any pathogen has accounted for this type of strain-specific immunity. We simulate our model to (i) understand the population-level consequences of hosts requiring two episodes of infection within a given time frame to obtain enduring strain-specific immunity; (ii) determine whether epidemiological observations of GAS in an Australian Indigenous population are consistent with this type of immune response; and to (iii) investigate how one of the leading multivalent strain-specific GAS vaccines could potentially alter the prevalence of GAS in the Australian Indigenous context. Understanding generated may be crucial for predicting and understanding future population effects of GAS vaccines currently in development (see [27, 28] for reviews of the current state of GAS vaccine research).

Methods

In this section we describe our agent-based model of GAS transmission, the selection of model parameters based on available epidemiological studies, and our *in silico* experiments.

Model of GAS transmission

Our agent-based model simulates the transmission of $n(t)$ strains of GAS in a well-mixed host population (where agents correspond to hosts) of constant size N, in discrete time t. We assume the population is situated in a geographical region where n_{max} strains of GAS are in circulation so that $0 \leq n(t) \leq n_{max}$. Each strain is assumed to have on average identical transmissibility, cause infections with identical baseline average duration, and be equidistant to each other in 'antigenic strain space' (in which distance corresponds to antigenic dissimilarity, as was assumed in [35]) so that each strain prompts a distinct immune response in hosts.

The model tracks the age, infection and immunity status of each host through time. Changes in host infection and immunity status occur due to the clearance of infections, transmission events, and waning immunity (detailed below), and are updated synchronously at the end of each day. New susceptible individuals aged zero are introduced into the population at a per capita rate d to replace individuals that are lost due to natural death. We also model migration at a per capita rate of α (detailed below).

Infection. In high incidence settings, multiple strains of GAS have been concurrently detected in the same and different skin lesions of individuals [36]. Therefore, in our model, hosts can be co-infected by multiple strains. We assume that a host can have a maximum of κ infections at any one time (including multiple infections of the same strain), and that the susceptibility of hosts to infection decreases as the total number of infections in each host increases. These assumptions incorporate the effects of pathogen populations directly competing for space and resources within the host, or indirectly interacting via the host immune response. We calculate the relative susceptibility r of host i to an uninfected host as

$$r = \left(1 - \frac{g_i(t)}{\kappa}\right)^x, \tag{1}$$

where $g_i(t)$ is the total number of infections of host i at time t and $x > 0$ is a number scaling the level of resistance to acquisition of new infections due to the competitive advantage of already established infections. Clearly, if host i is uninfected then $r = 1$, and if the host is at infection carrying capacity κ then $r = 0$.

Each day, each infecting strain will clear with probability $\Gamma = 1 - \exp(-\gamma/s)$, where $1/\gamma$ is the mean duration of infection of a host without prior immunity, and s is the expected relative duration of infection of a host compared to a host without prior immunity (detailed below). If a host has multiple infections of the same strain and this strain clears during a time step, then we assume that all infections of that strain in the host clear simultaneously.

Transmission. In the model, each host has on average c contacts with other hosts per day. The contacts of infected hosts are chosen uniformly at random from the population, and the outcomes of these contact events are then determined (*i.e.*, whether or not a transmission event occurs). We specify that transmission may only occur one-way from the infected host to their contacts. The probability of a contact resulting in transmission is $B = \beta r$, where β is the baseline probability of transmission, and r is the relative susceptibility of a host to an uninfected host (detailed above). If the infected host has more than one infection, only one of these co-infections can possibly transmit during a single contact event. For co-infected hosts, we choose one infection uniformly at random to attempt transmission. If this attempt fails, then the contact event does not result in transmission. These rules correspond to the assumption that co-infected hosts are not necessarily more infectious than hosts with a single infection. We also specify that a host may only contract a maximum of one infection per day.

With these assumptions, we can calculate the *basic reproduction number* $\mathcal{R}_0$, which is the expected number of secondary infections caused by a single infected host introduced into a

completely susceptible host population. A pathogen is expected to cause an outbreak or become endemic in a host population if $\mathcal{R}_0 > 1$. In our model, $\mathcal{R}_0$ is defined as

$$\mathcal{R}_0 = \frac{c\beta}{\gamma + d + \alpha}. \tag{2}$$

Immunity. Based on observations in the mouse model of GAS skin infection discussed above [34], we assume that the clearance of any host's first infection by a particular strain confers temporary immunity. This temporary immunity has a strain-specific effect of strength σ (where $0 \le \sigma \le 1$) and a cross-strain effect of strength ω_1 (where $0 \le \omega_1 \le \sigma$) that lasts for a duration w for all hosts and strains.

If a host has temporary strain-specific immunity to a particular strain and is reinfected by the same strain, clearance of this subsequent infection leads to enduring strain-specific immunity that prevents reinfection by this strain and confers enduring cross-strain immunity of strength ω_2 that is effective against strains that a host does not have temporary or enduring strain-specific immunity to. However, if this temporary immunity wanes, then a subsequent infection by this strain will only confer temporary immunity with the same characteristics as a first infection. This natural history of infection is summarised in Fig 1. Henceforth, we refer to the duration w, as the 'maximum inter-infection interval' that enables the development of enduring strain-specific immunity.

We note that in the model, it is possible for a host without any prior strain-specific immunity of a strain to experience multiple infections of a particular strain simultaneously. Due to our assumptions about strain clearance (detailed above), all infections by the same strain will clear simultaneously in the model when the host recovers from this strain, leading to a single immune response. We assume that such a clearance event only confers temporary immunity.
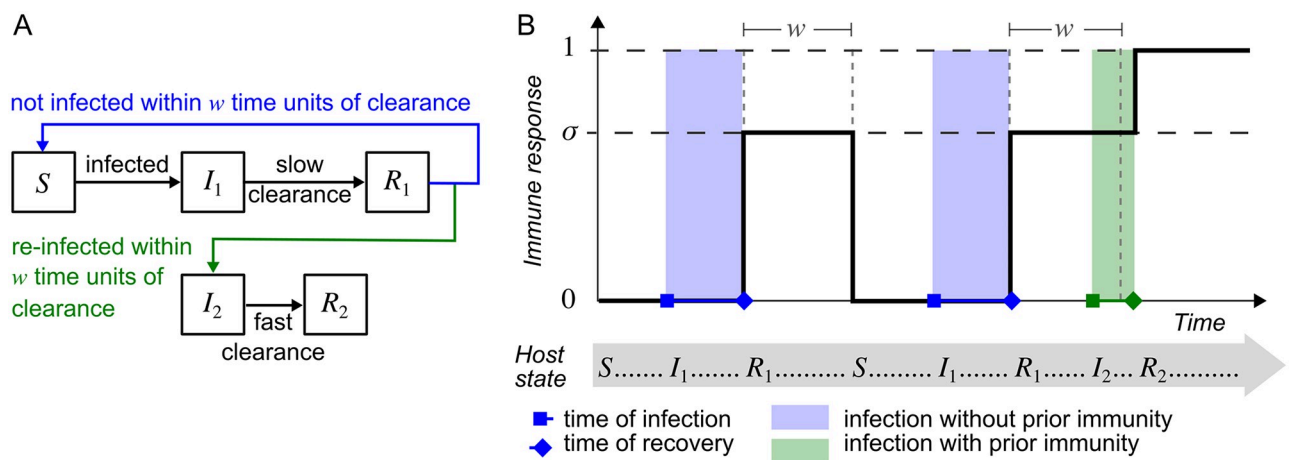

Fig 1. Model of the natural history of disease with respect to a single strain. A: Hosts without prior immunity to a particular strain (S) become infected by contacting infected hosts (I_1 or I_2). These infections (I_1) clear at an average rate γ which confers temporary immunity (R_1). This temporary immunity reduces the duration of a subsequent infection (I_2) by a factor dependent on the strength of temporary strain-specific immunity (σ) if the subsequent infection occurs within a short-enough time window (the maximum inter-infection interval, w) from the time of clearance (green line). If infection does not occur within this time frame (blue line), then temporary immunity wanes and a subsequent infection has the characteristics of a first infection. If temporary immunity does not wane before the next infection, then the clearance of this next infection occurs faster, and confers enduring immunity protecting against further infection (R_2). B: An example of a host's immune response (solid black line) following three episodes of infection by the same strain. Here, the temporary immunity acquired following the first infection wanes before the second infection. The clearance of the second infection leads again to temporary immunity. However this becomes enduring immunity following the clearance of the third, more timely, infection.

In the mouse model [34], the effect of the immune response was assessed by determining the number of colony forming units in skin and blood samples (bioburden) collected six days post inoculation. These showed a reduction in bioburden of approximately 90% for second infections of the same serotype compared to the first infection, provided that the second infection occurred within three weeks of the first. However, if the second infection was a different serotype, this reduction in bioburden ranged from approximately 0–30%. Our model does not explicitly represent bioburden within hosts. However, a reduction in bioburden during an infection could conceivably result in a reduced duration of infection and/or reduced infectiousness of a host during a contact event, or possibly prevent the host from ever being infectious (*i.e.*, the host is no longer susceptible to infection). In our model, we translate the reduction in bioburden due to host immunity into a reduction in the duration of infection. We note that a reduced duration of infection also corresponds to a reduction in the overall infectiousness of a host since a host will have less opportunities for transmission over the course of a shorter infection. Furthermore, with this assumed effect of immunity, further immune memory may be gained by a repeated exposure as if the host were totally naïve.

For each host i their expected relative duration of an infection by strain j compared to a host with no immunity is

$$s = \begin{cases} 1 - \sigma, & \text{if the host has temporary strain-specific immunity to strain } j, \\ 1 - \omega_1, & \text{if the host has only temporary strain-specific immunity to other strains,} \\ 0, & \text{if the host has enduring strain-specific immunity to strain } j, \\ 1 - \omega_2, & \text{if if the host has no strain-specific immunity to strain } j \text{ and} \\ & \text{enduring strain-specific immunity to other strains,} \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

Clearly, if a host has no immunity then the expected duration of an infection is not reduced from the baseline duration $1/\gamma$ (since $s = 1$ in this case). If a host has temporary strain-specific immunity of a strain at the time they are infected by this strain, then the expected duration of infection is reduced according to the strength of temporary strain-specific immunity σ (so that $s = 1 - \sigma$). A host with enduring strain-specific immunity to a strain is essentially completely protected against infection by this strain (since a subsequent infection by this strain will have zero duration). Without strain-specific immunity to a strain, a host may still have a shorter expected duration of infection by that strain if they have either temporary or enduring immunity of other strains at the time of infection (since either $s = 1 - \omega_1$ or $s = 1 - \omega_2$ in these cases).

Migration. In host settings where GAS disease is hyper-endemic and where high numbers of GAS strains typically co-circulate, different strains of GAS have been observed to move sequentially through communities rather than persist indefinitely [21–24]. The introduction of novel strains and previously circulating strains into these populations is thought to be enabled by host mobility [37, 38]. Therefore, in our model, each day A hosts (where A is a Poisson distributed random variable with mean αN, and α is the per capita migration rate) are chosen uniformly at random to be replaced by immigrants. Immigrants are assumed to have a similar immune profile to individuals in the population. This is implemented by specifying that an immigrant will have the same immune profile as an individual selected uniformly at random from the population. Immigrants may also be infected with up to one copy of infection of any strain (chosen uniformly at random from all n_{max} strains in the region). The prevalence of

infection in immigrants is set at 10% to be consistent with the asymptomatic carriage rate of GAS across all age groups and population settings [39].

Summary statistics

Two metrics are used to summarise transmission dynamics in our model at the population-level at time t: the diversity of strains $D(t)$, and the prevalence of infected hosts $P(t)$. We choose these summary statistics as they can be calculated from existing epidemiological data of GAS transmission [24]. Strain diversity is a measure of the total number of strains as well as how evenly strains are distributed across all infections in the host population. We calculate strain diversity using Simpson's reciprocal index, $D(t)$:

$$D(t) = \frac{M(t)(M(t) - 1)}{\sum_j m_j(t)(m_j(t) - 1)},\tag{4}$$

where $m_j(t)$ is the number of infections of strain j in the host population at time t, and $M(t)$ is the total number of infections in the host population at time t. The prevalence of infected hosts in a host population, $P(t)$, is calculated as

$$P(t) = \frac{1}{N}\sum_i \mathbb{1}_{\mathbb{Z}^+}(g_i(t)) \times 100\%,\tag{5}$$

where $\mathbb{1}_{\mathbb{Z}^+}(g_i(t))$ is the indicator function of the subset $\mathbb{Z}^+$ (the positive integers) of the set of all non-negative integers $\mathbb{Z}^0$ which takes the value of one when $g_i(t) \in \mathbb{Z}^+$ (*i.e.*, when host i has at least one infection) and zero otherwise. We define pathogen extinction to be the case where the prevalence $P(t) = 0$.

In silico experimental approach

Since GAS is endemic in human populations, we only consider endemic transmission dynamics in our model. All simulations are run for at least 50 years to allow the epidemiological dynamics to reach a quasi-steady state where the level of immunity in the population reaches a stable level. The level of immunity in the population is determined by the distribution of the number of strains that hosts in the population currently have immunity to, $Y(t)$, the mean of which is given by $\hat{Y}(t)$. We define the quasi-steady state (where $\hat{Y}(t)$ is stable) as the endemic equilibrium. We also define P^*, D^* and $\hat{Y}^*$ to be, respectively, the endemic values of the summary statistics $P(t)$ and $D(t)$ and of the mean population immunity $\hat{Y}(t)$. These are calculated by taking the mean values of $P(t)$, $D(t)$ and $\hat{Y}(t)$ across the previous 5 years (that is, for $t \in [45, 50]$ years).

Selection of model parameters. Table 1 shows the parameters in our model and the values we considered in our simulations. Parameters were selected to reflect GAS transmission an Indigenous population of northern Australia, where GAS disease is hyper-endemic and the majority of GAS infections are skin infections [24].

The population size N and the number of strains circulating in the region $n_{\max}$ are set at 2500 and 40 respectively to be consistent with community sizes [24] and the number of strains circulating [19] among Indigenous populations of northern Australia. The mean duration of infection $1/\gamma$ is set at 14 days to be consistent with clinic data collected in this setting [22–24].

The number of daily contacts c is calculated using household contact data collected in remote Australian Indigenous communities [41]. In this setting, it is estimated that individuals make approximately 22 contacts per day on average in households. Due to a lack of data describing contact patterns outside of households in these populations, we make the

Table 1. Parameter values.

Symbol	Description	Values / Range	Ref.
N	Population size	2500	[24]
n_{max}	Number of strains circulating in region	40	[24]
d	Per capita death rate (years^{-1})	1/71	[40]
c	Mean number of daily contacts	33	[41]
α	Mean per capita rate of migrations (weeks^{-1})	0.002	[42]
σ	Temporary strain-specific immunity strength	0.9	[34]
ω_1	Temporary cross-strain immunity strength	0.1	[34]
ω_2	Enduring cross-strain immunity strength	0.1	[34]
$1/\gamma$	Baseline mean duration of infection (days)	14	[22–24]
w	Maximum inter-infection interval (weeks)	$\{3,19,104,420,\infty\}$	[43]
$\mathcal{R}_0$	Basic reproduction number	[1, 10]	-
κ	Maximum number of co-infections per host	$\{10, 20, 40\}$	-
x	Level of resistance to co-infection	$\{1,10,100\}$	-

assumption that an individual will have roughly half the number of contacts outside of households compared to within households (approximately 11 contacts per day), as has been assumed previously for a model of influenza transmission in this setting [41]. Therefore, we set the mean number of daily contacts c to be 33.

Migration patterns are not described in this settings. We set the per capita expected migration rate α to 0.002 per week which corresponds to an average of 5 migration events per week when the population size $N = 2500$. With the prevalence of infection in migrants set to 10%, infected migrants enter the population approximately once every two weeks, which is consistent with genomic analysis of GAS isolates collected across two Indigenous communities in Northern Australia [42].

Values for parameters relating to the effects of immunity are determined from the mouse model of GAS skin infection [34]. We set the strength of temporary strain-specific immunity σ to 0.9 and the strength of temporary and enduring cross-strain immunity to 0.1. These values are based on the respective observations of 90% and 0-30% reduction in bioburden in the mouse due to strain-specific and cross-strain immunity [34].

To date, $\mathcal{R}_0$ has not been calculated for GAS. We explore values of $\mathcal{R}_0$ ranging from 1–10 (detailed below). For each combination of the parameters $\{\mathcal{R}_0, 1/\gamma, c\}$ considered, the baseline transmission probability β is calculated using Eq (2).

There is also limited data on co-infection for GAS, which is not always accounted for during data collection or when typing GAS specimens. We consider nine different co-infection scenarios defined by different values of the co-infection carrying capacity κ and the level of resistance to co-infection x (detailed below).

What are the population-level consequences of enduring strain-specific immunity being contingent on repeat infections?. The maximum inter-infection interval w was estimated to be three weeks in the mouse model [34]. It is not clear how this timespan translates in humans. Based on comparisons in mice versus humans of lifespan, the time of weaning, and the age of adulthood onset, the equivalent 3-week timespan in humans could be estimated, respectively, as either 104 weeks, 19 weeks or 420 weeks, respectively [43]. Therefore, to understand the population-level consequences of enduring strain-specific immunity being contingent on repeated episodes of infection of the same strain, we consider all three of these estimates for the maximum inter-infection interval w in humans, as well the case where w

remains unchanged between the mouse and human, that is, when $w = 3$ weeks. We also consider the null case where there is no upper bound on the time allowed between the first and second infections for clearance of the second infection to confer enduring immunity, that is, when $w = \infty$.

We explore values of $\mathcal{R}_0$ in increments of 0.5 ranging from 1–10. This range includes values of $\mathcal{R}_0$ that are consistent with estimates for other pathogenic bacteria that occupy similar niches to GAS: *S. pneumoniae* [44, 45] and *Staphylococcus aureus* [46] ($\mathcal{R}_0 = 2$–3). It also allows for the possibility that GAS may have a higher than expected $\mathcal{R}_0$ in Indigenous populations of northern Australia, where factors such as household crowding [41] and poor access to clean water [47] may increase transmissibility.

Finally, we consider scenarios with $\kappa \in \{10, 20, 40\}$ and $x \in \{1, 10, 100\}$. These co-infection parameters affect the shape of the function r (Eq (1)) governing the relative probability of transmission to a host relative to an uninfected host during a contact event with an infected host, as shown in S1 Fig. With increasing x, the chances of a host acquiring additional infections decreases as their number of co-infections approaches κ.

For each value of w, $\mathcal{R}_0$, κ and x considered, we perform 80 simulations of our model. From each set of simulations, we obtain distributions for the values of the summary statistics at equilibrium (the endemic prevalence P^* and endemic strain diversity D^*) as well as the endemic mean population immunity $\hat{Y}^*$, from which we calculate their mean values, and 25%–75% quantiles.

Are our model outputs consistent with epidemiological data?.   Next, we determine whether data simulated from our model (with any of the estimates of w, $\mathcal{R}_0$, κ and x considered) is consistent with epidemiological data collected in a hyper-endemic population (a community indigenous to Northern Australia [24]). In this previous study, prospective surveillance of a population of approximately 2500 people was carried out monthly over a 23 month period. Swabs were taken from the throats of all participants and any skin sores of participants and GAS isolates underwent strain typing (according to *emm* sequence, which is the sequence at the 5′ end of a locus found in all GAS isolates that encodes the M-protein, a cell-surface protein). From this data we calculate the prevalence and strain diversity at each time point and use this to estimate the endemic prevalence and strain diversity in this setting. As this study did not collect serological data, we cannot estimate endemic population immunity.

For each parameter scenario, the comparison to real data is achieved by first simulating transmission until the dynamics reach equilibrium (for 50 years) before continuing the simulation for a further 22 months, sampling the data monthly in a manner reflective of the previous study's surveillance protocol [24]. Specifically, 548 people were enrolled in the study in this community and the number of consultations each month ranged from 21 to 211. For each model realisation we assign 548 hosts uniformly at random from the whole population into the study, and from this pool of hosts, each month we sample, uniformly at random, the same number of hosts that were seen in the corresponding month of the study. We then compare the distributions of sampled P^* and D^* to those calculated from the real data.

What is the potential impact of a multivalent vaccine?.   A number of GAS vaccines are in the vaccine pipeline, including multivalent vaccines targeted towards serotypes associated with pharyngitis and invasive disease in Northern America and Western Europe [48]. While these targeted multivalent vaccines are predicted to provide high strain coverage in their target populations, the coverage in other populations where disease burden is much greater is predicted to be much lower [17, 19]. For example, at the time of design, a leading multivalent GAS vaccine was estimated to target only 25% of the serotypes of GAS circulating in

Indigenous populations of Australia, and 85-90% of serotypes in Northern America (ignoring any potential cross reactivity between serotypes) [19].

To investigate how a targeted 30-valent vaccine could potentially alter the prevalence of GAS in the Indigenous Australian context, we simulate the effects of a vaccination program consisting of routine vaccination and a one-off catch up campaign. The routine vaccination program vaccinates children when they reach one-year of age. At the commencement of the intervention, a one-off catch-up campaign vaccinates primary school-aged children in the population (aged 5–11 years). In the absence of a currently licensed vaccine, or any real-world studies to determine optimal vaccination schedule or vaccine effectiveness, we explore the extreme assumption that vaccine immunity is life-long, protects against all strains in the vaccine, and has an effectiveness of 90% (which takes into account both imperfect vaccine protectiveness and imperfect program coverage). This allows us to assess the greatest possible impact of immunisation. Lesser impacts on population dynamics are anticipated for a vaccine with only temporary protection.

A region-wide vaccination program will likely alter the overall prevalence of vaccine versus non-vaccine strains in the region. Therefore, strains infecting immigrants are no longer chosen uniformly at random from all n_{max} strains in the region. Instead, we define the probability $p_v(t)$ to be the probability that a strain infecting an immigrant will be a vaccine strain at time t. This is calculated as

$$p_v(t) = (n_v(t) + 1)/(n_v(t) + n_{nv}(t) + 2),\qquad(6)$$

where $n_v(t)$ and $n_{nv}(t)$ are, respectively, the number of vaccine strains and non-vaccine strains present in the population at time t. This expression for $p_v(t)$ is chosen so that (1) there is a small chance that an infected migrant will be carrying a vaccine strain when there are no vaccine strains currently present in the population (since $p_v(t) > 0$); and (2) there is a small chance that an infected migrant will be carrying a non-vaccine strain when there are no non-vaccine strains currently present in the population (since $p_v(t) < 1$). Since we are unsure how the vaccine will affect the overall prevalence of infection, we make the conservative assumption that the prevalence of infection in immigrants remains unchanged at 10%. For every infected immigrant, if it is determined (via the probability p_v) that their infecting strain is a vaccine strain, then this strain is chosen uniformly at random from the set of all vaccine strains. Conversely, if it is determined that their infecting strain is a non-vaccine strain, then this strain is chosen uniformly at random from the set of all non-vaccine strains. The vaccination status of any immigrants coming into the population are determined in the same way as their immune profiles—by specifying that the immune profile and vaccination status be the same as that of individuals in the population sampled uniformly at random.

We assess a range of vaccine scenarios that vary by the extent to which the 30-valent vaccine is tailored to the Australian Indigenous population context. We consider scenarios where the vaccine protects against infection by 25% of GAS strains circulating in the region (10 strains), an intermediate case where there is 50% strain coverage (20 strains), and a best-case scenario where all 30 strains targeted by the vaccine are strains that are currently circulating in the region (corresponding to 75% strain coverage). For each of these scenarios, we also explore the effect of further tailoring the vaccine to the population by choosing the vaccine strains to be the most-prevalent strains at the commencement of the intervention, as opposed to a random selection of strains (which might arise if the vaccine were tailored to another population setting).

We compare the base-line (pre-vaccine) endemic epidemiological dynamics with those calculated post-vaccine (after a further 100 years to allow the epidemiological dynamics to re-

equilibrate). We also consider the short-term impact of the vaccine during the first two years of implementation. The intervention scenarios considered are further broken down into those where routine vaccination is, or is not, supplemented by the one-off catch-up campaign targeting primary school aged children.

Results

The total prevalence of infection and strain diversity are maintained by the successive reintroduction of strains

In our model, endemic transmission is characterised by continuous strain turnover rather than the persistence of individual strains over long periods of time, which is consistent with GAS epidemiological observations within endemic settings [21–24]. When individual strains appear in the population, they either fade out quickly or cause an outbreak that can last for a period of months before going locally extinct and then reappearing some time later due to a re-importation. Outbreaks of individual strains can also partially overlap, but this overlap is reduced for larger outbreaks (Fig 2A).

Despite the unstable nature of individual strains, a positive overall prevalence of infection $P(t)$ and diversity of strains $D(t)$ can be maintained in the population over long periods of time (Fig 2B and 2C) if the maximum inter-infection interval w and the basic reproduction number $\mathcal{R}_0$ are appropriately specified (this is expanded upon below). In such cases, $P(t)$ and $D(t)$ oscillate around stable positive values at endemic equilibrium as individual strains sporadically
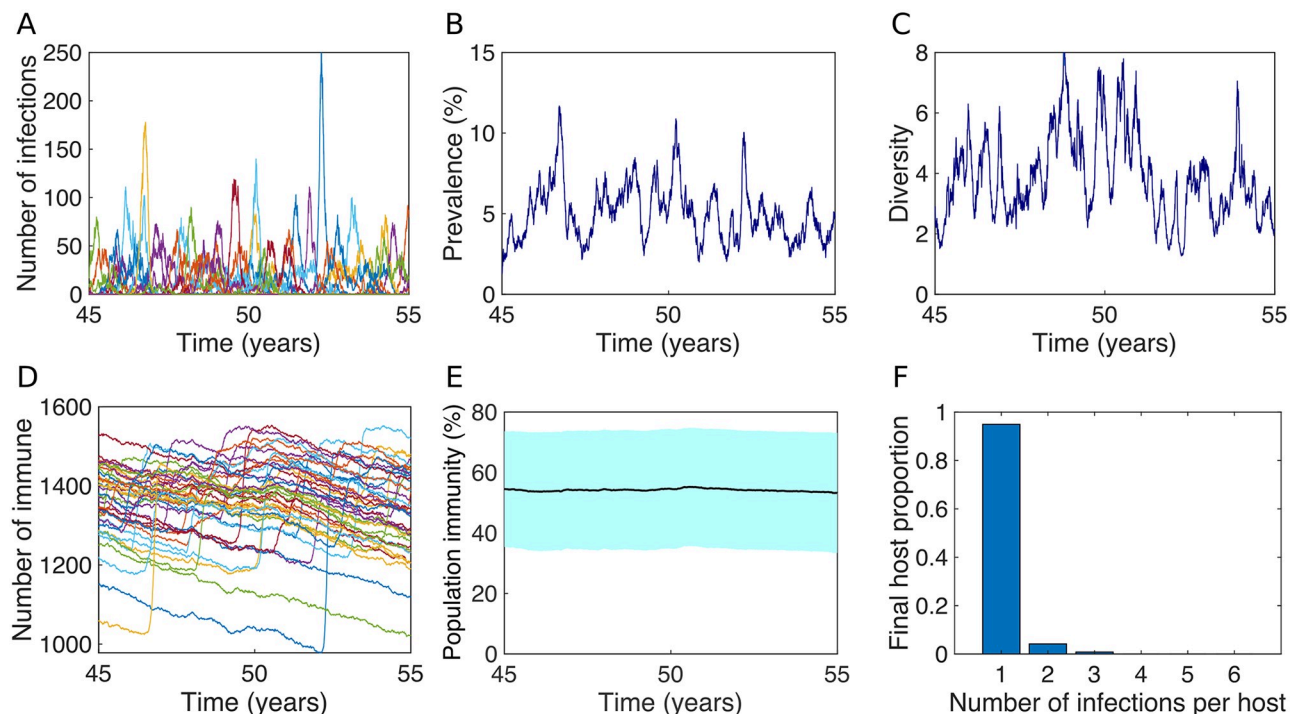

Fig 2. The prevalence of infection and strain diversity are maintained by the successive reintroduction of strains. Output from one realisation of the GAS transmission model over a ten-year period that follows the population reaching a quasi-steady, *i.e.*, an endemic equilibrium (after 45 years). (A) The number of infections of each strain, (B) the total prevalence of infected hosts $P(t)$, (C) strain diversity $D(t)$, (D) the number of hosts immune to each strain, (E) the mean (line) and inter-quartile range (shading) of the number of strains that hosts have immunity to (shown here as a percentage of the total number of strains in circulation in the region, $Y(t)/n_{max} * 100\%$), and (F) the final distribution of the number of infections per infected hosts (at $t = 55$ years). Here, $\mathcal{R}_0 = 3$, $1/\gamma = 2$ weeks, $c = 33$ per day, $\alpha = 0.002$ per capita per week, $n_{max} = 40$, $n(0) = 35$, $N = 2500$, $w = 19$ weeks, $x = 10$, $\sigma = 0.9$, and $\omega_1 = \omega_2 = 0.1$.

appear, cause an outbreak, and then fade out. This outbreak-type behaviour of individual strains is due to successive rapid accumulations and slow decays of the number of hosts immune to each strain following strain re-importations (Fig 2D). Despite the unstable nature of population immunity with respect to individual strains, the mean number of strains that hosts are immune to, $\hat{Y}(t)$, does not undergo oscillations at endemic equilibrium (Fig 2E). Instead, it is maintained at close to a constant level $\hat{Y}^*$. With $\mathcal{R}_0 = 3$, hosts have enduring immunity to approximately 50% of the 40 strains in circulation at endemic equilibrium (mean 21.4, IQR 13.6–29.2).

For a fixed value of the inter-infection interval w, increasing $\mathcal{R}_0$ above unity initially causes both an increase in the endemic prevalence P^* and strain diversity D^* until their maxima are achieved somewhere between $2 < \mathcal{R}_0 < 6$ for all values of w considered (Fig 3). Further increases to $\mathcal{R}_0$ result in a slow decrease for both of these quantities. Therefore, a non-monotonic relationship exists between the basic reproduction number $\mathcal{R}_0$ and both the endemic prevalence P^* and strain diversity D^*. For a fixed value of $\mathcal{R}_0$, increasing w from the value estimated in the mouse model of infection (3 weeks), to the smallest estimate of the equivalent timespan in humans (19 weeks) has a substantial effect on reducing both the endemic prevalence P^* and strain diversity D^* for all values of $\mathcal{R}_0 > 1$ considered (Fig 3). Further increases to w (beyond 19 weeks) correspond to increasingly smaller reductions in the endemic prevalence P^* and strain diversity D^* for all values of $\mathcal{R}_0 > 1$ considered.

Variation to either of the co-infection parameters κ and x over the ranges considered here has a limited effect on the described relationships between $\mathcal{R}_0$, w and both the endemic prevalence P^* and strain diversity D^* (S2 and S3 Figs). Even if there is low resistance to co-infection ($x = 1$ or $x = 10$), the low endemic prevalence of infected hosts (generally $P^* < 15\%$), means there is limited opportunity for infected hosts to contact each other and acquire additional infections (Fig 2F).

The co-infection parameters κ and x also have a limited effect on population immunity at endemic equilibrium (S4 Fig). In contrast to P^* and D^*, only $\mathcal{R}_0$ has a substantial affect on the
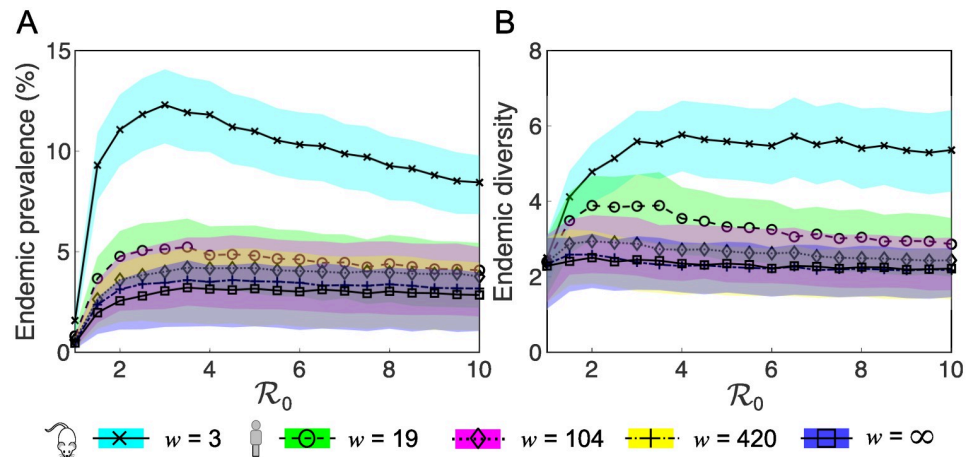

Fig 3. The relationship between $\mathcal{R}_0$ and w, and the endemic prevalence P^* and strain diversity D^*. The mean (lines) and the interquartile ranges (shaded regions) of (A) the total endemic prevalence of infected hosts, and (B) endemic strain diversity D^*, from 80 simulations of the model, when the maximum inter-infection interval w is the value estimated in the mouse model of GAS skin infection (3 weeks), when it is equal to three estimates of the equivalent timespan in humans (19, 104 and 420 weeks), and when there is no maximum inter-infection interval specified ($w = \infty$ weeks), as a function of the basic reproduction number $\mathcal{R}_0$ (horizontal axis). Here, $\mathcal{R}_0 \in [1, 10]$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{\max} = 40$, $n(0) = 30$, $N = 2500$, $x = 10$, $\sigma = 0.9$, and $\omega_1 = \omega_2 = 0.1$.

mean endemic level of population immunity $\hat{Y}^*$. In all model scenarios considered, $\hat{Y}^*$ increases monotonically with $\mathcal{R}_0$ and is relatively insensitive to w (S4 Fig). This is likely due to the rapid outbreak-type behaviour of individual strains. While the dynamics of a strain outbreak are controlled by all model parameters, the long-term enduring immunity dynamics are largely driven by the effective reproduction number, which increases slowly as enduring immunity is lost in the population due to migration and the birth of new susceptible hosts.

Model outputs are consistent with epidemiological data collected in a hyper-endemic population

The distributions of the endemic prevalence P^* and strain diversity D^* obtained from the sampled simulated data as well as from the real data collected in a hyper-endemic setting all show large variation (Fig 4, S4 and S5 Figs) which likely reflects the sparse sampling of these quantities as well as their oscillating nature at endemic equilibrium (as illustrated in Fig 2B and 2C). For both P^* and D^*, there is little difference in the distributions simulated with $w > 19$ weeks, suggesting that the maximum inter-infection w is only identifiable if it is sufficiently small when using these statistics to summarise transmission dynamics.

When we compare the real and simulated distributions of the endemic prevalence P^*, we find that there is the greatest overlap of the interquartile ranges when the values of the inter-infection interval w are set to either $w = 3$ weeks or $w = 19$ weeks, and when the basic reproduction number $\mathcal{R}_0$ is set between $2 \leq \mathcal{R}_0 \leq 5$ (Fig 4A). This is true for all values of the co-infection parameters κ and x considered (S4 Fig). The corresponding simulated distributions of endemic strain diversity D^* show substantial overlap with that of the real data for all values of w considered (Fig 4B). Again, this pattern does not change if we alter the co-infection parameters κ and x (S5 Fig). Therefore, we conclude that epidemiological data collected in a hyper-endemic population is most consistent with our simulated data generated when the inter-infection interval is between three and nineteen weeks and $2 \leq \mathcal{R}_0 \leq 5$.
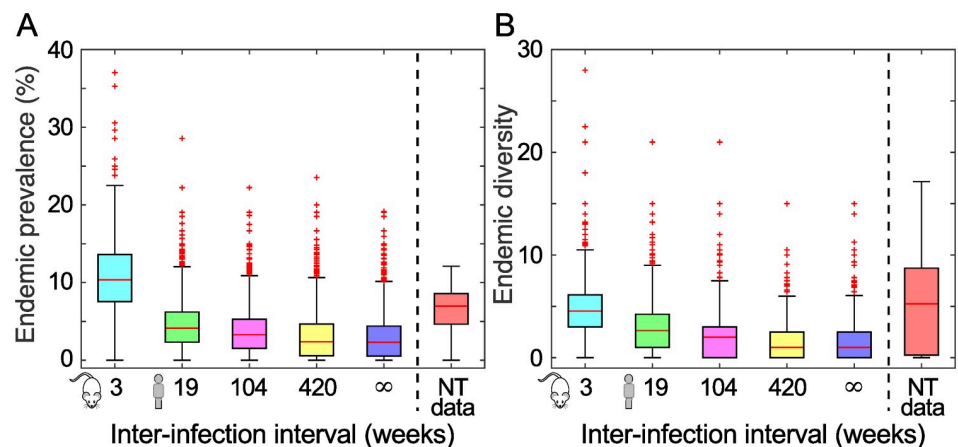

Fig 4. Model outputs are consistent with epidemiological data collected in a hyper-endemic population. The distribution of (A) the total endemic prevalence P^* of infected hosts and (B) endemic strain diversity D^*, from 80 simulations of the model, when the maximum inter-infection interval w is the value estimated in the mouse model of GAS skin infection (3 weeks), when it is equal to three estimates of the equivalent timespan in humans (19, 104 and 420 weeks), and when there is no maximum inter-infection interval specified (∞ weeks). Results are compared to population data (red) collected in one Indigenous community in the Northern Territory (NT) of Australia [24]. Here, $\mathcal{R}_0 = 2.5$, $1/\gamma = 2$ weeks, $c = 33$ per day, $\alpha = 0.002$ per capita per week, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $x = 10$, $\sigma = 0.9$, and $\omega_1 = \omega_2 = 0.1$. Similar results are obtained for $2 \leq \mathcal{R}_0 \leq 5$ for all values of w considered, as evidenced by the results shown in Fig 3, and so these results are not shown.

https://doi.org/10.1371/journal.pcbi.1007182.g004

Impact of a targeted multivalent vaccine is dampened by strain replacement

When we consider the impact of a targeted multivalent (serotype-specific) vaccine on transmission, we find that the effects of the vaccine program in the short term (over the first 2 years post introduction) and in the long term (once the system reaches a new endemic equilibrium) depend on the number of distinct strains in circulation that the vaccine protects against (Fig 5). Only short-term vaccine impact is dependent on the prevalence of each vaccine strain at the commencement of the intervention, and the choice of whether or not to implement the one-off catch-up campaign.

Specifically, in vaccine scenarios with the one-off catch-up campaign, prevalence $P(t)$ is quickly reduced following commencement of the vaccine program compared to equivalent scenarios without the catch-up campaign (Fig 5A–5C). This reduction in prevalence is greater when the vaccine is targeted towards the most-prevalent strains in the population at the time of the intervention, particularly for strain coverage less than 50% (Fig 5D–5F). However, prevalence rebounds in the months following the catch-up campaign to levels that are seen in equivalent scenarios without the catch-up campaign, particularly for low vaccine coverage. On average, this occurs within a year when the strain coverage in the vaccine is 25%. When the coverage is 50% or 75%, on average, this process takes greater than two years.

In the long term, we find that the vaccine reduces the endemic prevalence P^* by an amount that is less than the percentage of circulating strains targeted by the vaccine. Specifically, with 25%, 50% and 75% strain coverage in the vaccine, the median endemic prevalence P^* is reduced by 20%, 39% and 66% respectively. The failure to fully sustain initial reductions in prevalence following vaccine introduction is due to the partial replacement of vaccine strains with non-vaccine strains, as evidenced by the corresponding small reductions in median endemic strain diversity D^* of 14%, 20% and 38%, respectively.

Discussion

Incomplete understanding of the the immune response to GAS infection in individuals and the development of herd immunity in host populations represents a key barrier to the development of a globally effective GAS vaccine. Current consensus is that the immune response to GAS infection is largely strain (serotype)-specific [27]. Recent evidence in a murine model of GAS skin infection raises the possibility that the longevity of this immune response may be contingent on individuals experiencing a repeat episode of infection by the same strain within a narrow time window [34]. As yet, there is no direct evidence for an analogous immune response to GAS infection in humans.

Indirect evidence for immunity being contingent on repeat infections

The results of our mathematical modelling study indicate that epidemiological observations of GAS infections in a population with high rates of GAS disease are consistent with enduring strain-specific immunity being contingent on repeated infection with the same strain. Both epidemiological observations [21–24] and the data simulated from our model with a sufficiently long maximum inter-infection interval w are reflective of there being a continuous turnover of GAS strains in the population rather than individual strains persisting over long periods of time (see Fig 2). In our model, this strain cycling is enabled by (1) infected hosts migrating into the population and triggering outbreaks of new or previously-circulating strains; (2) the accumulation of hosts with enduring immunity which causes these strains to go locally extinct; and (3) the loss of sufficient herd immunity due to the continual influx of
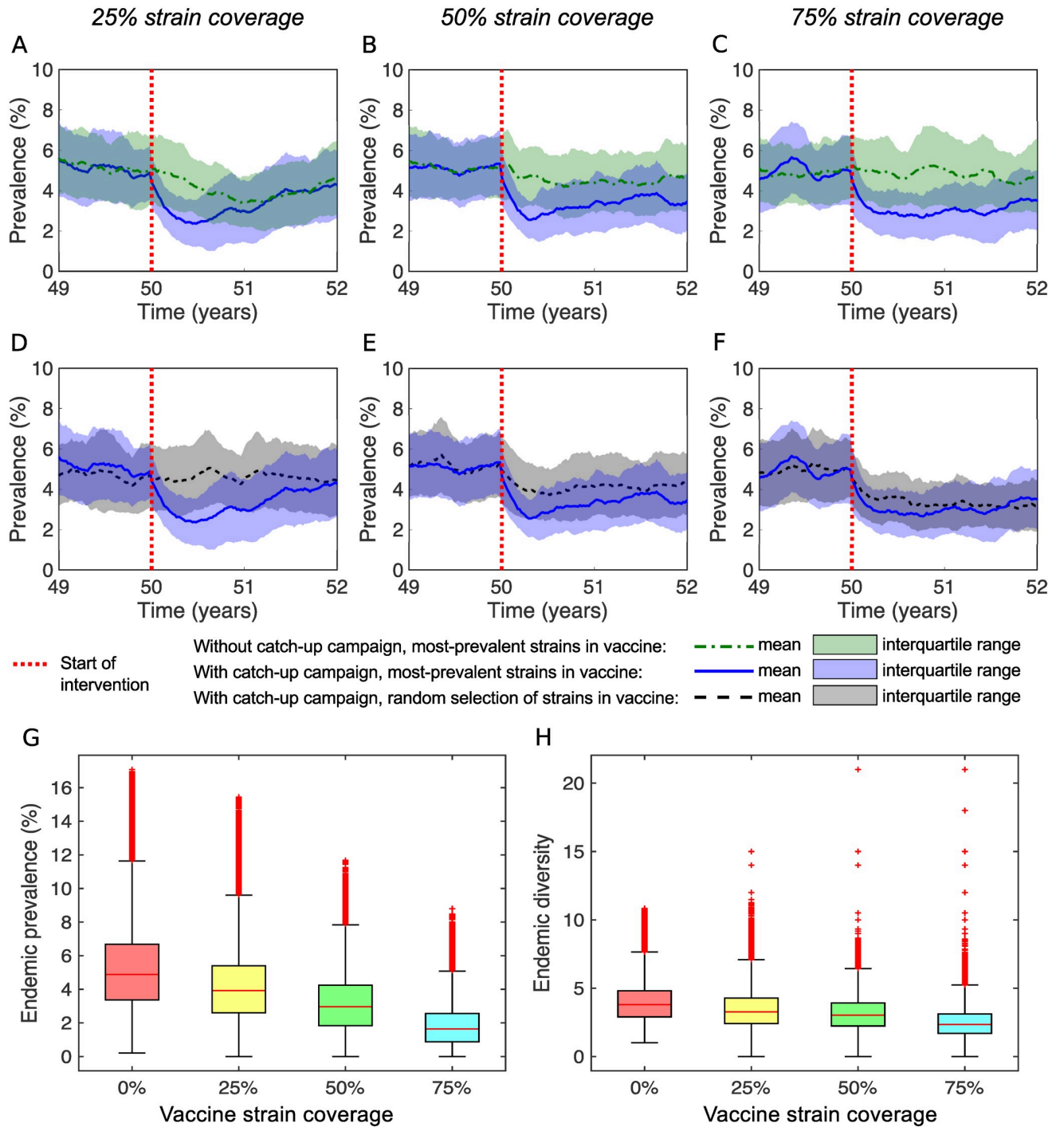
Fig 5. Impact of a targeted multivalent vaccine is dampened by strain replacement. (A–F) The mean (lines) and interquartile range (shaded regions) from 80 simulations of the model showing the prevalence over time, before and after the initiation of a vaccine intervention with (A,D) 25% strain coverage; (B,E) 50% strain coverage; and (C,F) 75% strain coverage. (A–C) Strains targeted by the vaccine are the most-prevalent strains at the initiation of the intervention. Scenarios with routine vaccination only (green), are compared against those where routine vaccination is supplemented with a one-off catch-up campaign (blue). (D–F) Routine vaccination is supplemented with a one-off catch-up campaign. Scenarios where strains targeted by the vaccine are the most-prevalent strains at the initiation of the intervention (blue), are compared against those where vaccine strains are randomly selected (black/grey). (G–H) The distributions of the (G) prevalence; and (H) strain diversity, calculated at endemic equilibrium pre vaccination (red boxplot) compared against those calculated post vaccination when there is (yellow) 25%, (green) 50% and (light blue) 75% strain coverage in the vaccine, when those strains targeted by the vaccine are the most-prevalent strains at the initiation of the intervention, and when there is a one-off catch up campaign as well as routine vaccination. Here, $\mathcal{R}_0 = 3$, $w = 19$ weeks, $1/\gamma = 2$ weeks, $c = 33$, $\alpha = 0.002$ per capita per week, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $x = 10$, $\sigma = 0.9$, and $\omega_1 = \omega_2 = 0.1$.

https://doi.org/10.1371/journal.pcbi.1007182.g005

susceptible hosts into the population (through birth and migration) which eventually allows a future reimportation to trigger another outbreak. We found that our model best matches real epidemiological data when the maximum inter-infection interval w is between 3 and 19 weeks, and if the basic reproduction number $\mathcal{R}_0$ is between 2 and 5.

An alternative hypothesis of the immune response to GAS skin infection is that enduring strain-specific immunity can be acquired through the clearance of a single infection. Mathematical models of other multi-strain pathogens that incorporate this type of immune response can also exhibit high strain turnover in host populations and result in a non-monotonic relationship between $\mathcal{R}_0$ and the endemic prevalence [1, 49], similar to what is observed in our model. However, this hypothesis precludes individuals experiencing repeated infections by the same GAS strain, which has been observed in children in high-incidence settings [22]. Another alternative hypothesis is that skin infection can never lead to enduring strain-specific immunity, but only temporary strain-specific immunity, thus allowing repeat infection by the same strain once immunity has waned. Future modelling work could consider whether there are conditions under which such a model is also consistent with GAS epidemiological data collected in high-incidence settings.

Epidemiological consequences of immunity being contingent on repeat infections

Our study demonstrates the broader epidemiological consequences of enduring strain-specific immunity being contingent on repeated episodes of infection. Pathogen transmissibility has competing effects on the likelihood of hosts acquiring enduring immunity in our model, which leads to a complex relationship between transmissibility and prevalence.

Increasing the basic reproduction number $\mathcal{R}_0$ from small values initially corresponds to a rise in the endemic prevalence of infection P^* due to increased transmission. This increase in P^* continues until transmission reaches a critical level whereupon it becomes more feasible for hosts to encounter the same strain twice within the required time window w and acquire enduring immunity. In this regime, further increases to $\mathcal{R}_0$ correspond to increased levels of herd immunity that eventually lead to reductions in the endemic prevalence P^* for further increases to $\mathcal{R}_0$. However, these further increases to $\mathcal{R}_0$ correspond to increasingly smaller reductions in the endemic prevalence P^*, possibly because the reduction in duration of outbreaks of individual strains (which coincide with increases to transmissibility) impacts the extent to which hosts can experience multiple episodes of infection of the same strain during a single outbreak. This is supported by the corresponding convergence of population immunity $\hat{Y}^*$ towards a maximum value for high values of $\mathcal{R}_0$.

A possible consequence of a non-monotonic relationship existing between $\mathcal{R}_0$ and the endemic prevalence is that interventions designed to reduce $\mathcal{R}_0$ (*e.g.*, via social interventions to improve household crowding or access to healthcare or running water) may lead to different outcomes in populations characterised with different baseline $\mathcal{R}_0$. For example, an intervention that leads to a substantial reduction in prevalence in one population may lead to very little change or even an increase in prevalence in a different population that has a higher baseline $\mathcal{R}_0$.

Short and long-term benefits of tailoring a multivalent vaccines to target populations

Our study also demonstrated how our model can be used to interpret and predict the effects of a targeted multivalent-vaccine intervention in a high-incidence setting. A key determinant of long-term vaccine impact is the number of strains that the vaccine protects against that are

circulating in the greater geographic region of the population. The greatest long-term reductions in prevalence occur when all strains in the vaccine are those in circulation, indicating the importance of customising a multivalent vaccine to particular host settings, or incorporating more conserved antigens with multivalent formulations [17].

Nevertheless, the high strain turnover that characterises transmission is likely to limit the long-term effectiveness of a targeted multivalent vaccine that does not protect against *every* strain in circulation. In our model, the replacement of vaccine strains with non-vaccine strains occurred within a few years of the implementation of the vaccine intervention. This occurred even when there was 75% vaccine strain coverage, and following significant short-term reductions in prevalence. While it may be the case that initial reductions in prevalence following the introduction of the vaccine cannot be sustained, it may be possible for long-term benefits to arise if the vaccine is rolled out in combination with other interventions designed to reduce transmission. It will be crucial to conduct surveillance for a number of years following vaccine introduction to evaluate short- and long-term vaccine impact.

Limitations and future work

In our model of GAS transmission in an Indigenous population of northern Australia, we assume that all GAS infections lead to the same type of immune response—that which was was observed in the mouse model of GAS skin infection [34]—since the majority of mild GAS infections in this setting are skin infections [23, 24]. However, in lower incidence settings, current consensus is that GAS causes throat infections more frequently than skin infections [18]. Furthermore, GAS can also be carried in the nose and throat of hosts without symptoms, and, less frequently, cause invasive disease [18]. It is not clear whether these other types of GAS infections cause an analogous immune response. If so, future modelling work could consider transmission and the effect of interventions in populations where other or multiple types of immune responses to GAS infection occur.

We have assumed that all GAS strains in the model have identical epidemiological characteristics. Further empirical work is needed to determine the validity of this assumption. Given that all strains share the same ecological niche, any differences in the competitive ability of strains will likely alter the level of strain diversity that can be sustained over short and long timescales in populations [50]. Furthermore, perturbations to pathogen population structure through the implementation of a vaccine targeting a subset of strains is likely to also depend on the epidemiological characteristics of targeted strains relative to non-targeted strains [50].

There are parallels between our simulation results and observed responses to the multivalent vaccines targeting another highly diverse human pathogen, *S. pneumoniae*. *S. pneumoniae* has over 90 different serotypes, and the multivalent pneumococcal conjugate vaccines (PCVs) targeted the most prevalent *S. pneumoniae* serotypes responsible for severe disease in different populations. The response to the PCVs varied across subgroups within these populations [51, 52]. However, generally there was a decrease in detection of vaccine strains and an increase in detection of non-vaccine strains following the implementation of PCV programs [53]. This is speculated to be due, in part, to strain replacement [53], similar to what occurred in our simulations. However, evolutionary factors such as serotype switching [54] and selection dynamics associated with the accessory genome, which remained relatively unchanged pre and post the implementation of the PCVs [55], may also have played a role in the observed vaccine response. Future work could consider exploring similar factors in the context of a GAS vaccine by incorporating evolutionary dynamics, such as mutation and recombination, into our model.

Supporting information

S1 Fig. The effect of co-infection parameters on the shape of the function governing the relative probability of transmission to an infected host compared to an uninfected host. The relative probability $r(g)$ of transmission to an infected host compared to an uninfected host is a function of the number of infections in the host g (horizontal axis). The co-infection carrying capacity κ and the level of resistance to co-infection x determine the shape of $r(g)$. Here, $r(g)$ is shown for $0 \leq g \leq 10$, $\kappa \in \{10, 20, 40\}$ and $x \in \{1, 10, 100\}$.
(TIF)

S2 Fig. The relationship between $\mathcal{R}_0$, w, κ and x and the endemic prevalence P^*. The mean (lines) and the interquartile ranges (shaded regions) of the total endemic prevalence of infected hosts P^* from 80 simulations of the model, as a function of the basic reproduction number $\mathcal{R}_0$ (horizontal axis), for different values of the maximum inter-infection interval w (varied within each figure panel), the co-infection carrying capacity κ (varied across columns) and the level of resistance to co-infection x (varied across rows). Here, $\mathcal{R}_0 \in [1, 10]$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $\sigma = 0.9$, $\omega_1 = \omega_2 = 0.1$, $w \in \{3, 19, 104, 420, \infty\}$ weeks, $\kappa \in \{10, 20, 40\}$ and $x \in \{1, 10, 100\}$. Note that the interquartile ranges overlap for $w > 3$ weeks.
(TIF)

S3 Fig. The relationship between $\mathcal{R}_0$, w, κ and x and the endemic diversity D^*. The mean (lines) and the interquartile ranges (shaded regions) of the endemic strain diversity D^* from 80 simulations of the model, as a function of the basic reproduction number $\mathcal{R}_0$ (horizontal axis), for different values of the maximum inter-infection interval w (varied within each figure panel), the co-infection carrying capacity κ (varied across columns) and the level of resistance to co-infection x (varied across rows). Here, $\mathcal{R}_0 \in [1, 10]$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $\sigma = 0.9$, $\omega_1 = \omega_2 = 0.1$, $w \in \{3, 19, 104, 420, \infty\}$ weeks, $\kappa \in \{10, 20, 40\}$ and $x \in \{1, 10, 100\}$. Note that the interquartile ranges overlap for $w > 3$ weeks.
(TIF)

S4 Fig. The relationship between $\mathcal{R}_0$, w, κ and x and the mean endemic level of population immunity $\hat{Y}^*$. The mean (lines) and the interquartile ranges (shaded regions) of the mean endemic level of population immunity $\hat{Y}^*$ from 80 simulations of the model, as a function of the basic reproduction number $\mathcal{R}_0$ (horizontal axis), for different values of the maximum inter-infection interval w (varied within each figure panel), the co-infection carrying capacity κ (varied across columns) and the level of resistance to co-infection x (varied across rows). Here, $\mathcal{R}_0 \in [1, 10]$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $\sigma = 0.9$, $\omega_1 = \omega_2 = 0.1$, $w \in \{3, 19, 104, 420, \infty\}$ weeks, $\kappa \in \{10, 20, 40\}$ and $x \in \{1, 10, 100\}$. Note that all interquartile ranges overlap.
(TIF)

S5 Fig. Comparison of the endemic prevalence P^* estimated from population data collected in one Australian Indigenous community, to P^* estimated by sampling model outputs. Model outputs are generated for a range of values of co-infection parameters κ (varied across columns) and x (varied across rows), and the inter-infection infection interval w (varied within each figure panel). Distributions of P^* for each parameter combination were obtained from 80 simulations of the model. Here, $\mathcal{R}_0 = 2.5$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{max} = 40$, $n(0) = 30$, $N = 2500$, $\sigma = 0.9$, $\omega_1 = \omega_2 = 0.1$, $\kappa \in \{10, 20, 40\}$, $x \in \{1, 10,$

100}, and $w \in \{3, 19, 104, 420, \infty\}$ weeks.
(TIF)

S6 Fig. Comparison of the endemic diversity D^* estimated from population data collected in one Australian Indigenous community, to D^* estimated by sampling model outputs. Model outputs are generated for a range of values of co-infection parameters κ (varied across columns) and x (varied across rows), and the inter-infection infection interval w (varied within each figure panel). Distributions of D^* for each parameter combination were obtained from 80 simulations of the model. Here, $\mathcal{R}_0 = 2.5$, $1/\gamma = 2$ weeks, $\alpha = 0.002$ per capita per week, $c = 33$, $n_{\max} = 40$, $n(0) = 30$, $N = 2500$, $\sigma = 0.9$, $\omega_1 = \omega_2 = 0.1$, $\kappa \in \{10, 20, 40\}$, $x \in \{1, 10, 100\}$, and $w \in \{3, 19, 104, 420, \infty\}$ weeks.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Rebecca H. Chisholm, Manisha Pandey, Mark R. Davies, Steven Y. C. Tong, Jodie McVernon, Nicholas Geard.

Data curation: Malcolm I. McDonald.

Formal analysis: Rebecca H. Chisholm, Nikki Sonenberg.

Writing – original draft: Rebecca H. Chisholm.

Writing – review & editing: Rebecca H. Chisholm, Nikki Sonenberg, Jake A. Lacey, Malcolm I. McDonald, Manisha Pandey, Mark R. Davies, Steven Y. C. Tong, Jodie McVernon, Nicholas Geard.

References

1. Abu-Raddad LJ, Ferguson NM. The impact of cross-immunity, mutation and stochastic extinction on pathogen diversity. Proc R Soc B. 2004; 271(1556):2431–2438. https://doi.org/10.1098/rspb.2004.2877 PMID: 15590592

2. Buckee CO, Koelle K, Mustard MJ, Gupta S. The effects of host network structure on pathogen diversity and strain structure. Proc Nat Acad Sci USA. 2004; 101:10839–10844. https://doi.org/10.1073/pnas.0402000101 PMID: 15247422

3. Krugman S, Giles JP, Friedman H, Stone S. Studies on immunity to measles. J Pediatr. 1965; 66 (3):471–488. https://doi.org/10.1016/S0022-3476(65)80112-3 PMID: 14264306

4. Rubin S, Mauldin J, Chumakov K, Vanderzanden J, Iskow R, Carbone K. Serological and phylogenetic evidence of monotypic immune responses to different mumps virus strains. Vaccine. 2006; 24 (14):2662–2668. https://doi.org/10.1016/j.vaccine.2005.10.050 PMID: 16309801

5. Johswich KO, McCaw SE, Strobel L, Frosch M, Gray-Owen SD. Sterilizing immunity elicited by Neisseria meningitidis carriage shows broader protection than predicted by serum antibody cross-reactivity in CEACAM1-humanized mice. Infection and immunity. 2015; 83(1):354–363. https://doi.org/10.1128/IAI.02495-14 PMID: 25368118

6. Katrak K, Mahon BP, Minor P, Mills K. Cellular and humoral immune responses to poliovirus in mice: a role for helper T cells in heterotypic immunity to poliovirus. Journal of general virology. 1991; 72 (5):1093–1098. https://doi.org/10.1099/0022-1317-72-5-1093 PMID: 1851808

7. Weinberger DM, Dagan R, Givon-Lavi N, Regev-Yochay G, Malley R, Lipsitch M. Epidemiologic evidence for serotype-specific acquired immunity to pneumococcal carriage. The Journal of infectious diseases. 2008; 197(11):1511–1518. https://doi.org/10.1086/587941 PMID: 18471062

8. Murphy BR, Whitehead SS. Immune response to dengue virus and prospects for a vaccine. Annual review of immunology. 2011; 29:587–619. https://doi.org/10.1146/annurev-immunol-031210-101315 PMID: 21219187

9. Gupta S, Maiden MC, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. Nature medicine. 1996; 2(4):437–442. https://doi.org/10.1038/nm0496-437 PMID: 8597954

10. Gupta S, Ferguson N, Anderson R. Chaos, Persistence, and Evolution of Strain Structure in Antigenically Diverse Infectious Agents. Science. 1998; 280:912–915. https://doi.org/10.1126/science.280.5365.912 PMID: 9572737

11. Dawes JHP, Gog JR. The onset of oscillatory dynamics in models of multiple disease strains. J Math Biol. 2002; 45:471–510. https://doi.org/10.1007/s00285-002-0163-9 PMID: 12439588

12. Koelle K, Pascual M, Yunus M. Serotype cycles in cholera dynamics. Proc R Soc B. 2006; 273:2879–2886. https://doi.org/10.1098/rspb.2006.3668 PMID: 17015366

13. Buckee C, Danon L, Gupta S. Host community structure and the maintenance of pathogen diversity. Proc R Soc B. 2007; 274:1715–1721. https://doi.org/10.1098/rspb.2007.0415 PMID: 17504739

14. Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, et al. Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. Proc Nat Acad Sci USA. 2008; 105:15082–15087. https://doi.org/10.1073/pnas.0712019105 PMID: 18815379

15. Cobey S, Lipsitch M. Pathogen Diversity and Hidden Regimes of Apparent Competition. Am Nat. 2013; 181:12–24. https://doi.org/10.1086/668598 PMID: 23234842

16. Bessen DE, McShan WM, Nguyen SV, Shetty A, Agrawal S, Tettelin H. Molecular epidemiology and genomics of group A Streptococcus. Infection, Genetics and Evolution. 2015; 33:393–418. https://doi.org/10.1016/j.meegid.2014.10.011 PMID: 25460818

17. Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, Towers RJ, et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. Nature genetics. 2019; p. 1.

18. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis. 2005; 5(11):685–694. https://doi.org/10.1016/S1473-3099(05)70267-X PMID: 16253886

19. Smeesters PR, McMillan DJ, Sriprakash KS, Georgousakis MM. Differences among group A *Streptococcus* epidemiological landscapes: consequences for M protein-based vaccines? Expert Review of Vaccines. 2009; 8(12):1705–1720. https://doi.org/10.1586/erv.09.133 PMID: 19905872

20. Bowen AC, Mahé A, Hay RJ, Andrews RM, Steer AC, Tong SY, et al. The global epidemiology of impetigo: a systematic review of the population prevalence of impetigo and pyoderma. PLoS One. 2015; 10 (8):e0136789. https://doi.org/10.1371/journal.pone.0136789.

21. Gardiner DL, Sriprakash KS. Molecular epidemiology of impetiginous group A streptococcal infections in aboriginal communities of northern Australia. Journal of Clinical Microbiology. 1996; 34(6):1448–1452. https://doi.org/10.1128/JCM.34.6.1448-1452.1996 PMID: 8735096

22. Bessen DE, Carapetis JR, Beall B, Katz R, Hibble M, Currie BJ, et al. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. The Journal of infectious diseases. 2000; 182(4):1109–1116. https://doi.org/10.1086/315842

23. McDonald MI, Towers RJ, Fagan P, Carapetis JR, Currie BJ. Molecular typing of Streptococcus pyogenes from remote Aboriginal communities where rheumatic fever is common and pyoderma is the predominant streptococcal infection. Epidemiology & Infection. 2007; 135(8):1398–1405. https://doi.org/10.1017/S0950268807008023

24. McDonald M, Towers R, Andrews R, Benger N, Fagan P, Currie B, et al. The dynamic nature of group A streptococcal epidemiology in tropical communities with high rates of rheumatic heart disease. Epidemiology & Infection. 2008; 136(4):529–539. https://doi.org/10.1017/S0950268807008655

25. Watkins DA, Johnson CO, Colquhoun SM, Karthikeyan G, Beaton A, Bukhman G, et al. Global, regional, and national burden of rheumatic heart disease, 1990–2015. New England Journal of Medicine. 2017; 377(8):713–722. https://doi.org/10.1056/NEJMoa1603693 PMID: 28834488

26. Shulman ST, Tanz RR, Dale JB, Beall B, Kabat W, Kabat K, et al. Seven-year surveillance of North American pediatric group A streptococcal pharyngitis isolates. Clinical Infectious Diseases. 2009; 49 (1):78–84. https://doi.org/10.1086/599344 PMID: 19480575

27. Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, et al. Status of research and development of vaccines for *Streptococcus pyogenes*. Vaccine. 2016; 34:2953–2958. https://doi.org/10.1016/j.vaccine.2016.03.073 PMID: 27032515

28. Fischetti VA. Vaccine approaches to protect against Group A *Streptococcal* pharyngitis. Microbiology Spectrum. 2019; 7(3):GPP3–0010–2018. https://doi.org/10.1128/microbiolspec.GPP3-0010-2018

29. Danchin MH, Rogers S, Kelpie L, Selvaraj G, Curtis N, Carlin JB, et al. Burden of acute sore throat and group A streptococcal pharyngitis in school-aged children and their families in Australia. Pediatrics. 2007; 120(5):950–957. https://doi.org/10.1542/peds.2006-3368 PMID: 17974731

30. McDonald MI, Towers RJ, Andrews RM, Benger N, Currie BJ, Carapetis JR. Low rates of streptococcal pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clinical infectious diseases. 2006; 43(6):683–689. https://doi.org/10.1086/506938 PMID: 16912939

31. Rantz LA, Maroney M, Di Caprio JM. Infection and reinfection by hemolytic streptococci in early childhood. The American Journal of Medicine. 1952; 13(1):98–99. https://doi.org/10.1016/0002-9343(52)90108-3

32. Raynes JM, Frost HR, Williamson DA, Young PG, Baker EN, Steemson JD, et al. Serological evidence of immune priming by Group A streptococci in patients with acute rheumatic fever. Frontiers in microbiology. 2016; 7:1119. https://doi.org/10.3389/fmicb.2016.01119 PMID: 27499748

33. Brandt E, Hayman W, Currie B, Carapetis J, Wood Y, Jackson D, et al. Opsonic human antibodies from an endemic population specific for a conserved epitope on the M protein of group A streptococci. Immunology. 1996; 89(3):331–337. https://doi.org/10.1046/j.1365-2567.1996.d01-754.x PMID: 8958044

34. Pandey M, Ozberk V, Calcutt A, Langshaw E, Powell J, Rivera-Hernandez T, et al. Streptococcal Immunity Is Constrained by Lack of Immunological Memory following a Single Episode of Pyoderma. PLoS Pathog. 2016; 12(12):e1006122. https://doi.org/10.1371/journal.ppat.1006122 PMID: 28027314

35. Gog JR, Grenfell BT. Dynamics and selection of many-strain pathogens. Proceedings of the National Academy of Sciences. 2002; 99(26):17209–17214. https://doi.org/10.1073/pnas.252512799

36. Carapetis J, Gardiner D, Currie B, Mathews JD. Multiple strains of Streptococcus pyogenes in skin sores of aboriginal Australians. Journal of clinical microbiology. 1995; 33(6):1471–1472. https://doi.org/10.1128/JCM.33.6.1471-1472.1995 PMID: 7650169

37. Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. Clinical infectious diseases. 2006; 42(4):441–447. https://doi.org/10.1086/499812 PMID: 16421785

38. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. The Lancet infectious diseases. 2009; 9(10):611–616. https://doi.org/10.1016/S1473-3099(09)70178-1 PMID: 19778763

39. Oliver J, Wadu EM, Pierse N, Moreland NJ, Williamson DA, Baker MG. Group A Streptococcus pharyngitis and pharyngeal carriage: A meta-analysis. PLoS neglected tropical diseases. 2018; 12(3): e0006335. https://doi.org/10.1371/journal.pntd.0006335 PMID: 29554121

40. Australian Bureau of Statistics. Life Tables for Aboriginal and Torres Strait Islander Australians, 2015-2017, 'Table 1.6: Life tables for Aboriginal and Torres Strait Islander Australians, Northern Territory–2015-2017', data cube: Excel spreadsheet, cat. no. 3302.0.55.003; 2018. Available from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/3302.0.55.003Main+Features12015-2017?OpenDocument [cited 8 January 2019].

41. Vino T, Singh GR, Davison B, Campbell PT, Lydeamore MJ, Robinson A, et al. Indigenous Australian household structure: a simple data collection tool and implications for close contact transmission of communicable diseases. PeerJ. 2017; 5:e3958. https://doi.org/10.7717/peerj.3958 PMID: 29085755

42. Marcato A, Lacey JA, Davies MR, Campbell PT, McDonald MI, Price DJ, et al.. Combining whole genome sequencing and epidemiology to investigate group A *Streptococcus* transmission in Indigenous communities; 2019, May. ASID Annual Scientific Meeting, Darwin, Australia.

43. Dutta S, Sengupta P. Men and mice: relating their ages. Life sciences. 2016; 152:244–248. https://doi.org/10.1016/j.lfs.2015.10.025 PMID: 26596563

44. Hoti F, Erästö P, Leino T, Auranen K. Outbreaks of Streptococcus pneumoniae carriage in day care cohorts in Finland–implications for elimination of transmission. BMC Infectious Diseases. 2009; 9 (1):102. https://doi.org/10.1186/1471-2334-9-102 PMID: 19558701

45. Gjini E. Geographic variation in pneumococcal vaccine efficacy estimated from dynamic modeling of epidemiological data post-PCV7. Scientific Reports. 2017; 7(1):3049. https://doi.org/10.1038/s41598-017-02955-y PMID: 28607461

46. Hogea C, Van Effelterre T, Acosta C. A basic dynamic transmission model of *Staphylococcus aureus* in the US population. Epidemiology & Infection. 2014; 142(3):468–478. https://doi.org/10.1017/S0950268813001106

47. Clifford H, Pearson G, Franklin P, Walker R, Zosky G. Environmental health challenges in remote Aboriginal Australian communities: clean air, clean water and safe housing. Australian Indigenous Health Bulletin. 2015; 15(2):1–13.

48. Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B, Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. Infection and immunity. 2002; 70(4):2171–2177. https://doi.org/10.1128/IAI.70.4.2171-2177.2002 PMID: 11895984

49. Abu-Raddad L, Van der Ventel B, Ferguson N. Interactions of multiple strain pathogen diseases in the presence of coinfection, cross immunity, and arbitrary strain diversity. Physical review letters. 2008; 100 (16):168102. https://doi.org/10.1103/PhysRevLett.100.168102 PMID: 18518250

50. Cobey S, Lipsitch M. Niche and Neutral Effects of Acquired Immunity Permit Coexistence of Pneumococcal Serotypes. Science. 2012; 6074:1376–1380. https://doi.org/10.1126/science.1215947

51. Collins DA, Hoskins A, Snelling T, Senasinghe K, Bowman J, Stemberger NA, et al. Predictors of pneumococcal carriage and the effect of the 13-valent pneumococcal conjugate vaccination in the Western Australian Aboriginal population. Pneumonia. 2017; 9(1):14. https://doi.org/10.1186/s41479-017-0038-x PMID: 29021946

52. Hammitt LL, Bruden DL, Butler JC, Baggett HC, Hurlburt DA, Reasonover A, et al. Indirect effect of conjugate vaccine on adult carriage of Streptococcus pneumoniae: an explanation of trends in invasive pneumococcal disease. The Journal of infectious diseases. 2006; 193(11):1487–1494. https://doi.org/10.1086/503805 PMID: 16652275

53. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. The Lancet. 2011; 378(9807):1962–1973. https://doi.org/10.1016/S0140-6736(10)62225-8

54. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, et al. Selective and genetic constraints on pneumococcal serotype switching. PLoS genetics. 2015; 11(3):e1005095. https://doi.org/10.1371/journal.pgen.1005095 PMID: 25826208

55. Corander J, Fraser C, Gutmann MU, Arnold B, Hanage WP, Bentley SD, et al. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. Nature ecology & evolution. 2017; 1(12):1950. https://doi.org/10.1038/s41559-017-0337-x