Serotype cycles in cholera dynamics
Katia Koelle, Mercedes Pascual and Md Yunus

Proc. R. Soc. B 2006 273, 2879-2886
doi: 10.1098/rspb.2006.3668

References

This article cites 34 articles, 11 of which can be accessed free
http://rspb.royalsocietypublishing.org/content/273/1603/2879.full.html#ref-list-1

Article cited in:
http://rspb.royalsocietypublishing.org/content/273/1603/2879.full.html#related-urls

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

To subscribe to Proc. R. Soc. B go to: http://rspb.royalsocietypublishing.org/subscriptions

This journal is © 2006 The Royal Society
Serotype cycles in cholera dynamics

Katia Koelle¹*, Mercedes Pascual¹ and Md. Yunus²

¹The Department of Ecology and Evolutionary Biology, 2019 Kraus Natural Science Building, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048, USA
²International Centre for Diarrheal Disease Research, Dhaka 1000, Bangladesh

Interest in understanding strain diversity and its impact on disease dynamics has grown over the past decade. Theoretical disease models of several co-circulating strains indicate that incomplete cross-immunity generates conditions for strain-cycling behaviour at the population level. However, there have been no quantitative analyses of disease time-series that are clear examples of theoretically expected strain cycling. Here, we analyse a 40-year (1966–2005) cholera time-series from Bangladesh to determine whether patterns evident in these data are compatible with serotype-cycling behaviour. A mathematical two-serotype model is capable of explaining the oscillations in case patterns when cross-immunity between the two serotypes, Inaba and Ogawa, is high. Further support that cholera’s serotype-cycling arises from population-level immunity patterns is provided by calculations of time-varying effective reproductive rates. These results shed light on historically observed serotype dominance shifts and have important implications for cholera early warning systems.

Keywords: cholera serotypes; strain cycling; cross-immunity; cholera two-strain model; disease time series

1. INTRODUCTION

Patterns of strain diversity and evolution may result from intrinsic characteristics of diseases, such as the persistence of infection and the interaction of strain variants (Grenfell et al. 2004). Interactions between strain variants include priming each other through antibody-dependent enhancement (Ferguson et al. 1999), directly competing within a host in times of co-infection, interfering with each other through generating a convalescent period in the host (Earn et al. 2000) or inducing cross-immunity (Andreasen et al. 1997). When strain interaction is through cross-immunity, herd-immunity levels generate selection pressures on strain evolution and impact disease dynamics. In the case of pathogens such as influenza, continual antigenic drift results from partial cross-immunity between sufficiently similar strains and the large number of possible variants. For pathogens with only a few possible strain variants, theoretical studies predict strain cycling as the predominant pattern (Andreasen et al. 1997; Gupta et al. 1998; Dawes & Gog 2002). However, due to the multiplicity of strains that underlies many diseases and the lack of sufficiently long disease datasets, a clear empirical example of this theoretical expectation is still lacking.

Here, we analyse a cholera dataset from Matlab, Bangladesh to determine whether the serotype patterns observed in the time-series provide such an example. The disease is caused by the Gram-negative bacterium Vibrio cholerae, and consists of two pathogenic serogroups, 01 and 0139. These serogroups differ significantly in their antigenic lipopolysaccharide characteristics, with no apparent cross-immunity between each other (Mooi & Bik 1997). However, within the 01 serogroup classification, cholera cases are identified as being of a particular biotype (classical or El Tor) and of a particular serotype (Inaba or Ogawa, and very rarely Hikojima). Cholera biotypes are distinct phenotypes that differ with respect to the severity of their infections (and consequently their infection-to-case ratios), ability to survive outside the human host and seasonality patterns (Woodward & Mosley 1972; Glass et al. 1982; Reidl & Klose 2002; Koelle et al. 2005a). Serotypes differ from one another only with respect to antigenic determinants that are present on their O-antigen capsule. Inaba organisms have antigenic determinants identified as A and C, while Ogawa organisms have determinants A and B (Stroher et al. 1992). Differences in the serotypes’ antigenic determinants are not known to confer differences in any phenotypic characteristics, such as the duration of the infectious period, the strength of the hosts’ immune response or the recovery rate. However, these antigenic determinants play a critical role in inducing host immunity. Owing to the common presence of the ‘A’ antigenic determinant, cross-immunity between serotypes exists. Strong evidence for this cross-immunity comes from patients who, being infected with one serotype, undergo serological conversion against the heterologous serotype (Benenson et al. 1968). Even though present, cross-immunity is thought to be incomplete for several reasons: (i) rare within-host serotype conversions are occasionally associated with relapses of symptoms (Gangarosa et al. 1967); (ii) cholera patients infected with a recently appeared serotype are generally older than patients infected with the endemic serotype (ProMed Digest 2005); and (iii) founder serotypes are often replaced by the heterologous serotype in epidemic areas once considerable serotype-specific immunity has been generated in the population (Felsenfeld 1966).
2. DISEASE DATA

In areas where cholera is endemic, it has long been noticed that serotypes tend to fluctuate in dominance (Pollitzer 1959), with shifts occurring in the intervals between epidemics of the disease (Venkatraman 1947; Felsenfeld 1959), with shifts occurring in the intervals between epidemics of the disease (Venkatraman 1947; Felsenfeld 1959). However, the reason for these fluctuations has not been understood, and the identification of regularities of these patterns is absent (Bhaskaran & Gorrill 1957; ProMed Digest 2005). A long-term cholera dataset provided by the International Centre for Diarrheal Disease Research, Bangladesh allows for the quantitative study of the origin of these serotype patterns (figure 1a). The cholera case data indicate a temporal separation of serotype dominance: while the Inaba serotype was dominant between the years 1966 and 1983, Ogawa has been the predominant serotype since that time. The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The cholera case data indicate a temporal separation of serotype dominance: while the Inaba serotype was dominant between the years 1966 and 1983, Ogawa has been the predominant serotype since that time. The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.

The appearance of the heterologous serotype occurs through the very low frequency of mutation between the two serotypes (Manning et al. 1994). Owing to the simplicity of its serotype structure, cholera therefore provides an ideal empirical system for evaluating the existence of theoretically expected strain-cycling behaviour.
early 1990s and in 2002/2003 when Inaba cases, although few in absolute number, are highest in proportion. A less striking, but nevertheless important, pattern is the slow decrease in mean age of cases in the early 1980s, as the Ogawa serotype becomes established. The association between serotype shifts and age incidence patterns seen here is in agreement with the previous observations, for example, by medical personnel in nearby Calcutta, India (ProMed Digest 2005).

In addition to the observed cholera data shown in figure 1, we have detailed demographic data over this period, with monthly birth and death information and decadal census data (not shown). The size of the population in the surveillance area is growing at a rate of 0.77% per year, with a current population size of 225 000.

3. THE MATHEMATICAL MODEL

We turn to the hypothesis that the observed serotype dynamics arise from the immunological interaction between the two serotypes, and thereby constitute part of a strain cycle. To first determine whether strain cycling is theoretically expected for a two-serotype system such as cholera, and to be able to quantify characteristics of these dynamics, we revisit a published two-strain model by Kamo & Sasaki (2002). Within this mathematical framework, cross-immunity acts by reducing the ability of a host to become infected with the heterologous strain. An individual who has been previously infected can therefore still become infected with the heterologous serotype, but at a reduced rate. Alternative definitions and models for cross-immunity exist (e.g. Andreasen et al. 1997; Gupta et al. 1998; Gog & Swinton 2002); we use this representation because volunteer studies have shown that patients re-challenged with the heterologous serotype of cholera rarely become infected. When symptoms do occur, they are milder and V. cholerae excretion levels are reduced (Cash et al. 1974), arguing for a concomitant reduction in transmissibility, which is not incorporated in the model framework we use here. The empirical evidence for the type of cross-immunity between serotypes therefore does not unequivocally support a particular model, e.g. Kamo & Sasaki (2002) versus Gupta et al. (1998), but a combination of models. The particular choice of model, however, does not affect our general conclusion (results not shown).

Here, we use the model presented by Kamo & Sasaki (2002) owing to its analytical results. From Kamo & Sasaki (2002), we have

\[
\frac{dx}{dt} = \beta x (1 - \gamma) \gamma y - \sigma x y,  \quad (3.1a)
\]

\[
\frac{dy}{dt} = \beta x (1 - \gamma) \gamma y - \sigma x y,  \quad (3.1b)
\]

\[
\frac{dx}{dt} = \beta (x y + \gamma z) - \mu (1 - x n),  \quad (3.1c)
\]

\[
\frac{dy}{dt} = \beta (x y + \gamma z) - \mu (1 - x n),  \quad (3.1d)
\]

\[
\frac{dx}{dt} = \beta (x y + \gamma z) - \mu (1 - x n) - \mu x,  \quad (3.1e)
\]

\[
\frac{dy}{dt} = \beta (x y + \gamma z) - \mu (1 - x n) - \mu x,  \quad (3.1f)
\]
Here, $y_I$ is the proportion of individuals infected with the Inaba serotype, $y_O$, the proportion of individuals infected with the Ogawa serotype, $x_N$, the proportion of hosts naive to both the serotypes, $x_I$, the proportion of hosts who have or have had an Inaba infection, $x_O$, the proportion of hosts who have or have had an Ogawa infection and $x_M$, the proportion of hosts who have or have had both Inaba and Ogawa infections. The present or previous infection with a serotype provides permanent immunity against re-infection with that serotype. Since $x_N + x_I + x_O + x_M = 1$, we use only the first five equations (3.1a–e) to completely describe the system. The parameters in the model are duration of infectiousness, 1/$\beta$, population turnover rate, $\mu$, transmissibility, $\beta$, and degree of cross-immunity, $\gamma$.

To see whether this model can generate the cholera patterns of figure 1, we first estimated parameter values for $\sigma$ and $\mu$ from Matlab’s demographic and cholera literature (figure 2). Two unknown parameters remained: the transmission rate, $\beta$, and the degree of cross-immunity, $\gamma$. Estimates of $\beta$ can be computed from the basic reproductive rate, $R_0$, by $\beta = R_0(\sigma + \mu)$. ($R_0$ for a two-strain system is defined as the number of secondary cases one infected individual begets in a population that is completely susceptible to both strains.) $R_0$ can be estimated by the equation $G/A$ for a single endemic strain in a growing population (May & Anderson 1985), with $G$ being the reciprocal of the birth rate, and $A$, the average age of first infection. This equation naturally also holds for a two-strain system with complete cross-immunity. For a two-strain–system in which there is no cross-immunity, $R_0$ is estimated by $G/2A$ (Gupta et al. 1994). For cholera, because some cross-immunity between the serotypes exists, we (for now) simply bound $R_0$ by these upper and lower limits (figure 2).

The degree of cross-immunity is a less easily identifiable parameter, because it is not consistently defined in the literature and owing to the limited number of volunteer re-challenge experiments and field studies. However, the studies that have taken place indicate that serotype cross-immunity is almost complete (Levine 1978; Levine et al. 1979). Additional patient studies show that a significant rise in vibriocidal antibodies to the heterologous serotype occurs in approximately 90% of infections (Benenson et al. 1968), further supporting a high level of cross-immunity, since vibriocidal levels are indicative of protection against cholera infection (Mosley 1969). Although considered high, cross-immunity is thought to be incomplete for the reasons previously mentioned (within-host serotype conversions, age-incidence patterns and serotype replacement patterns in epidemic areas).

Given our estimates of $\sigma$ and $\mu$, we first look at the sensitivity of the model’s behaviour as a function of the lesser known parameters, $R_0$ (proportional to $\beta$) and cross-immunity, $\gamma$. A mathematical analysis of this two-strain model (Kamo & Sasaki 2002) showed that this system has three fixed points. Two are the simple cases of one serotype being present and the other serotype being absent. The endemic fixed point is the only stable fixed point when $R_0 > 1$, with solutions given by Kamo & Sasaki (2002). At this stable fixed point, we can define $s_I$ as the equilibrium fraction of hosts susceptible to becoming infected with the Inaba serotype; mathematically, $s_I = x_N + (1 - \gamma)x_M$, as is evident from equation (3.1a).

Similarly, the proportion susceptible to becoming infected with Ogawa is $s_O = x_N^O + (1 - \gamma)x_M$. Since the serotypes Inaba and Ogawa are symmetrical in their parameter values, $s_I = s_O$, let $s$ equal this value. From equations (3.1a,b), we find that $s^* = 1/R_0$ and thereby independent of the degree of cross-immunity, $\gamma$. In figure 2a, we plot these analytically computed values for $s^*$ as a function of $R_0$. We can now use the serological data presented in figure 1c along with age-structured demographic data (Mosley et al. 1968) to compute cholera’s $s_1^*$ and $s_2^*$ in 1965: $s_1^* = 0.36$ and $s_2^* = 0.32$. Taking $s^*$ to be an average of these values, our estimate of $R_0$ can be approximated as $R_0 = 2.9$. This value is on the high end of our $R_0$ range and is reasonable in light of the expected high degree of cross-immunity seen from volunteer re-challenge studies.

Through a linearization of the system and an analysis of the resulting eigenvalues, Kamo and Sasaki also identified two natural frequencies present at this fixed point, $\omega_1$ and $\omega_2$; $\omega_1$ corresponded to the natural frequency of the synchronized dynamics of the two strains, while $\omega_2$ corresponded to the natural frequency of the anti-phase dynamics of the two strains. The anti-phase natural frequency, $\omega_2$, and its associated period, $2\pi/\omega_2$, can be computed directly from the imaginary part of the relevant eigenvalues. Since the fluctuations of the strains associated with this natural frequency are anti-phase (Kamo & Sasaki 2002), the period between one strain’s peak and the other strain’s peak, i.e. the inter-epidemic period, is simply $\pi/\omega_2$. We plot this inter-epidemic period of the damped oscillations in figure 2b, as a function of $R_0$ and the degree of cross-immunity, $\gamma$. For a given value of $R_0$, increasingly higher cross-immunity levels lead to serotype cycles of longer periods. As expected, when cross-immunity between strains is complete ($\gamma = 1$), the cycle length is infinite, effectively resulting in the competitive exclusion of one strain. Given the value of $R_0$ estimated from our calculated $s^*$ in figure 2a, the 15-year period of oscillation observed in our serotype dynamics (figure 1b) is consistent with the strain dynamics of the model when the degree of cross-immunity value is $\gamma = 0.956$.

We simulated the model given by equations (3.1a–e) in figure 2c with our refined parameter values for $\gamma$ and $R_0$. As expected, the transient dynamics show the 15-year anti-phase behaviour for the proportion of infected individuals, $y_I$ and $y_O$. Also visible in this simulation is another frequency, associated with a shorter period. This frequency ($\omega_2$) corresponds to the previously mentioned synchronous fluctuations of the two strains, with a period of 4.5 years, calculated directly from the relevant eigenvalues. This period is not evident in cholera’s ACF (figure 1b). This may be because the second natural frequency mode, corresponding to $\omega_2$, contributes more to the amplitude response of the linearized system when perturbed by demographic or environmental noise. Indeed, when we calculate the friction coefficients of the two modes (Kamo & Sasaki 2002), we find that $|\omega_2| < |\omega_1|$.

Seasonal fluctuations in the transmission rate, $\beta$, have been shown capable of generating interannual disease cycles in single-strain deterministic models (Schwartz & Smith 1983; Schwartz 1992). Likewise, Kamo & Sasaki (2002) showed that when seasonality is included in simulations of


2882 K. Koelle and others Serotype cycles in cholera dynamics
this two-strain model, the system can sustain interannual cycles of various periods, potentially a result of resonance with the natural frequencies of oscillation. Since cholera is highly seasonal (Glass et al. 1982), we simulated the deterministic cholera two-serotype system with realistic seasonality patterns derived from empirical monthly case distributions (electronic supplementary material). However, in our system, seasonality alone simply generated annual cycles with no interannual cycles evident (results not shown). Since the model we use is one that has been shown capable of generating interannual cycles when deterministic seasonal forcing is present (Kamo & Sasaki 2002), the inability of seasonal forcing to sustain these cycles in our system must be due to cholera’s specific parameter values. We suspect that either the very high degree of cross-immunity between the serotypes or the specific seasonal forcing function of cholera does not generate sufficient resonance with the natural oscillation frequencies to sustain interannual cycles in a purely deterministic framework.

It is also well known that demographic noise can sustain oscillations that would otherwise decay in time (Bartlett 1957; Bailey 1975). We therefore simulated a stochastic event-driven implementation of the two-serotype model. Simulations both with and without seasonality in transmission rates sustained the oscillations in case numbers (results not shown for without-seasonality case and figure 3 for with-seasonality case), with long inter-epidemic periods of approximately 15 years. Variability in the duration of serotype dominance was present, with the inter-epidemic period ranging between 10 and 20 years.

4. Effective Reproductive Rates
If we assume that this cross-immunity model adequately describes the interaction between the cholera serotypes, we can use Matlab’s demographic data and observed case data to reconstruct time-varying effective reproductive rates (figure 4), an indicator of serotype fitness levels in the population. The effective reproductive rate over time for Inaba is given by $R_I(t) = (\beta (x_N(t) + (1 - \gamma)x_O(t))) / (\sigma + \mu)$, and similarly for Ogawa, $R_O(t) = (\beta (x_N(t) + (1 - \gamma)x_I(t))) / (\sigma + \mu)$. With known parameters $\beta$, $\sigma$, $\mu$ and $\gamma$, and initial values for $x_N$, $x_O$, and $x_I$ that are consistent with the serological data (figure 1c), these time-varying effective reproductive rates can be reconstructed through an iterative approach (electronic supplementary material). The relative values of the effective reproductive rates quantify the magnitude of selection against the less-fit serotype (figure 4b), and can be computed by standardizing the serotypes’ effective reproductive rates to the one with higher fitness. Relative fitness levels are thereby computed as $r_I(t) = (R_I(t)/\max(R_I(t), R_O(t)))$ for Inaba and $r_O(t) = (R_O(t)/\max(R_I(t), R_O(t)))$ for Ogawa.

Effective reproductive rates, $R_I(t)$ and $R_O(t)$, fluctuate in response to changes in immunity levels, providing evidence for immunity-driven cholera case dynamics. Note that the number of Inaba cases generally grows until 1979 when the reconstructed effective reproductive values become less than 1 (for both serotypes), and likewise that the number of Ogawa cases grows during the period 1990–1993, then decreases in the rest of the 1990s and stabilizes at very low levels until the present time, in agreement with the dynamics expected from the effective reproductive rates.

Changes in the relative magnitude of the effective reproductive rates arise as a response to the population dynamics of the cholera cases; Ogawa becomes the fitter serotype after the long period in the 1970s of predominantly Inaba cases, while Inaba becomes the fitter serotype again in the early 1990s, after a period of predominantly Ogawa cases. This observation indicates that partial cross-immunity between Inaba and Ogawa serotypes generates conditions for frequency-dependent selection pressures. Moreover, switches in the identity of the fitter serotype (figure 4b) are seen to occur near the peaks of the outbreaks in 1978 and 1993 (figure 1a), with the highest relative fitness differences occurring when effective reproductive rates are near their minima, in 1983 and 2000. Given that cases are declining when effective reproductive rates are less than 1, a subsequent increase in cases when these rates...
infectious, s
understand how epidemiological factors (such as serotype dominance changes in cholera; it also helps us
dominance. This research provides insight into the reason
nation for the observed long-term changes in serotype
between cholera’s serotypes is a well-supported expla-
incomplete cross-immunity provides an explanation for
the increase in the mean age of infected individuals, which
become greater than 1 are likely to occur predomi-
nantly with the fitter serotype. Hence, fluctuations in
relative effective reproductive rates can explain the
previously noted pattern of serotype re-emergences
occurring in the time-intervals between epidemic out-
breaks (Venkatraman 1947).

5. DISCUSSION
Our analysis investigated whether patterns in cholera’s
serotype fluctuations provided quantitative empirical
evidence for theoretically expected strain-cycling
behaviour. A mathematical model was able to capture
the observed cyclical serotype dynamics seen in the
cholera data when all four of its parameter values were
either independently estimated from demographic and
edemiological data (μ and σ) or had reasonable ranges
independently estimated (R0 and γ). Combined with the
reconstructed effective reproductive rates, these analyses
indicate that high, but incomplete, cross-immunity
between cholera’s serotypes is a well-supported expla-
nation for the observed long-term changes in serotype
dominance. This research provides insight into the reason
for serotype dominance changes in cholera; it also helps us
understand how epidemiological factors (such as R0 and
the degree of cross-immunity) affect the period of these
serotype cycles. Moreover, the two-serotype model with
incomplete cross-immunity provides an explanation for
the increase in the mean age of infected individuals, which
is observed when one cholera serotype re-emerges
(ProMed Digest 2005), a pattern we also see in our data
(figure 1e,f). When a serotype re-emerges, older individ-
uals who are only immune to the previously dominant
serotype may come in contact with the re-emerging
serotype, thereby becoming infected and raising the
mean age of cholera cases.

Previous research of ours using nonlinear time-series
models to fit cholera incidence data has shown that instead
of conferring permanent immunity to a previously infected
host, cholera may confer temporary immunity of the order
of about a decade in duration (Koelle & Pascual 2004;
Koelle et al. 2005b). These analyses did not consider
cholera cases at the level of the specific serotypes, and may
therefore reflect phenomenological waning of immunity
caused by serotype shifts at the level of the population.
However, if we consider that cholera of a given serotype
may induce less than lifelong immunity, we would expect
serotype cycles to be faster than cycles generated with the
assumption of permanent immunity, for given values of
R0 and cross-immunity. Temporary immunity is still
capable of generating serotype cycles of long periods when
cross-immunity levels between the two serotypes is
sufficiently high (results not shown).

Insights gained from considering cholera dynamics at
the serotype level also have predictive value, suggesting
general qualitative trends. Given the calculated
trajectories in the effective reproductive rates, we would
expect the coming decade to show an increase in absolute
case numbers, with Inaba re-emerging as the predominant
serotype. In the early period of Inaba re-emerging, adults
should also become infected, but the mean age of cholera
patients should decrease thereafter as Inaba becomes
increasingly established in the population. More quanti-
tative predictions require further developments. Of
particular relevance is the demonstrated high sensitivity
of the model dynamics to the cross-immunity between
Inaba and Ogawa. It is therefore critical to obtain better
estimates of this parameter for predicting the timing of
serotype switching. Complementary approaches include
statistical modelling of cholera dynamics based on this
serotype model, along with stronger surveillance programs

Figure 4. Reconstructed effective reproductive rates. (a) Time-varying effective reproductive rates for Inaba and Ogawa. (b) Time-varying relative effective reproductive rates. Both El Tor and classical cases were included in the reconstruction of reproductive rates, with full immunity between classical and El Tor biotypes of a given serotype. Initial conditions for the reconstruction were $s_N = 0.33$, $s_t = 0.19$, $s_C = 0.32$ and $s_B = 0.15$, consistent with the average fraction susceptible to becoming infectious, $s^*$, of 0.34, and consistent with the serological evidence that the fraction susceptible to Inaba infection is greater than that susceptible to Ogawa infection in 1965 (figure 1c).
and more comprehensive re-challenge studies. The extreme sensitivity to the cross-immunity parameter also some of the variability in immunity cycle lengths of the stochastic model (figure 3), and therefore why trends in serotype dominance may seem irregular or not readily apparent (Bhaskaran & Gorrill 1957; ProMed Digest 2005).

Demographic stochasticity was considered here for a much larger population than the one for the surveillance region. The scope of the stochastic simulations was to verify that in the presence of noise, the decaying oscillations of the deterministic model become persistent. For small populations, extinctions are inevitable. While these extinctions are realistic, they require the explicit treatment in the models of additional processes associated with strain re-introductions or ‘rescue’ effects. In the case of cholera, the most important process to be considered is primary transmission from an environmental, aquatic reservoir (Hartley et al. 2006), a parameter for which independent estimates are not currently available. Our model formulation with only a direct transmission pathway should now be extended to include primary transmission, and fitted to the data in the presence of stochasticity, to consider smaller population sizes, evaluate the different transmission pathways and the predictability of the dynamics.

Mathematical models for cholera early warning systems will also need to examine how to couple the dynamics of serotype-specific immunity levels to variable transmission rates driven by environmental factors. Extrinsic factors, such as the El Niño-Southern Oscillation (ENSO) and long-term variability in northeast India rainfall, have previously been shown to play a role in the interannual variability of cholera at shorter periods (Pascual et al. 2000; Koelle et al. 2005b). Although the two-serotype model demonstrates that general trends in cholera cases and serotype fluctuations can be explained through intrinsic immunity mechanisms, several pieces of evidence also point here towards extrinsic factors as additional players in driving disease dynamics. In this model, variability in the reconstructed effective reproductive rates was generated solely by fluctuations in population-level susceptibility levels (figure 4), while the transmission rate, \( \beta \), was assumed a constant. It is this transmission term that is modulated by environmental drivers. For example, the resurgence of cases in the time-interval 1982–1984 by the classical biotype (of both Inaba and Ogawa serotypes) occurs when effective reproductive rates, only influenced by immunity levels, are negative. This period follows the strong 1982–1983 El Niño event, which may have transiently raised the transmission rate of cholera, thereby effectively reducing the transmission rate of cholera, thereby effectively reducing the effective reproductive rate greater than 1 for the period 1982–1984. Similarly, the negative ENSO anomalies present in the early 1970s, indicative of La Niña conditions, may have transiently lowered cholera transmission rates, thereby explaining the low number of cholera cases during this time, despite the high effective reproductive rates shown in figure 4. How environmental forcing at characteristic time-scales influences the period of the serotype cycles remains to be addressed more fully in future research. The variability observed in the stochastic simulations around a mean period cycle of 13 years suggests that cycle length may indeed vary, although these long-term cycles are generally expected to be sustained.

Major advances in understanding and predicting disease dynamics have been made for childhood diseases, such as measles (Earn et al. 2000; Finkenstädt & Grenfell 2000), where immunity is permanent and strain transmission. However, among the most ubiquitous diseases are influenza, malaria and diarrhoeal diseases, many of which have underlying strain structure with only partial cross-immunity between strains. With advances in molecular epidemiology and phylogenetic approaches, the dynamics of these multi-strain diseases are starting to be examined, with implications for early warning systems and vaccine development. Here, we have demonstrated that high cross-immunity between cholera’s serotypes is a well-supported mechanism that sets these strains in motion.

We thank Kim Sneathfield at the ICDDR,B for support with the demographic data and the NCEAS Seasonality of Infectious Diseases working group, in particular Otmar Bjornstad, David Alonso and Bruce Kendall, for helpful comments, and the Centre for the Study of Complex Systems at UM for cluster access. K.K. thanks the Rackham Graduate School for travel funds and G. S. S. for technical advice and feedback. This work was supported by EEB graduate fellowship support to K.K. and by the joint support of the NSF-NIH (Ecology of Infectious Diseases) and NOAA (Oceans and Health) to M.P.

REFERENCES


Serotype cycles in cholera dynamics


