Epidemiology / Épidémiologie

The perpetuation and epidemic recurrence of communicable diseases in human populations

Pierre-Yves Boëlle a,b,c

a UMR-S 707, université Pierre-et-Marie-Curie (Paris-6), faculté de médecine Saint-Antoine, 27, rue Chaligny, 75571 Paris cedex 12, France
b INSERM, UMR-S 707, 27, rue Chaligny, 75571 Paris cedex 12, France
c Assistance publique–Hôpitaux de Paris (APHP), hôpital Saint-Antoine, 27, rue Chaligny, 75571 Paris cedex 12, France

Received 23 January 2007; accepted after revision 15 February 2007
Available online 23 April 2007
Presented by Alain-Jacques Valleron

Abstract

Recurrence of communicable diseases is a looming threat for human populations. Factors explaining the recurrences are partially known, involving demographics, biology, and complex relationships with the environment, but no comprehensive theory exists today. Here, we review some recent results obtained in modelling studies with a view to understanding better the mechanisms of perpetuation. Factors intrinsic to the interaction of pathogen and host have regained interest in this respect, especially with multiple pathogen and multiple population interactions. Extrinsic factors, including pure demography and environmental forcing are also strong predictors. With increasingly detailed data available, large-scale integrated models will help sorting out the multiple influences on recurrence. To cite this article: P.-Y. Boëlle, C. R. Biologies 330 (2007).

Résumé


Keywords: Epidemiology; Models; Theoretical models; Communicable diseases

Mots-clés : Épidemiologie ; Maladies transmissibles ; Modèles théoriques

E-mail address: boelle@u707.jussieu.fr.
**Introduction**

Le mot « épidémie », utilisé par Hippocrate dans son acception médicale, signifiait « de retour dans son pays », avant de qualifier la propagation dans une population. Ceci illustre la menace permanente que constituent les maladies transmissibles, « chez elles » dans les populations humaines qu’elles visitent de temps en temps. En mai 2003, cette perspective était clairement rendue à propos du SRAS, lorsqu’un éditorial du New England Journal of Medicine concluait : « si nous sommes vraiment chanceux, l’épidémie sera maîtrisée, et deviendra saisonnière, ce qui améliorera les chances de contrôle ». Cette possibilité ne pouvait être négligée, mais la fin de l’histoire montre que notre compréhension de ce qui rend une maladie récurrente, les relations multiples et multi-niveaux nécessaires, est encore approximative.

Autant de maladies, autant de cycles : certaines sont franchement saisonnières (grippe, maladies infantiles), d’autres alternent en taille entre saisons (rougeole), ou créent des épidémies avec des fréquences variables (quatre ans pour la dengue et le chikungunya ; une dizaine d’années pour la syphilis et les méningites).

On dit qu’une maladie se perpétue lorsqu’elle cause des cas dans la population de manière ininterrompue. La perpétuation est compatible avec des dynamiques saisonnières ou plus aléatoires. L’importance des facteurs intrinsèques et extrinsèques dans la perpétuation est encore débattue.

**Les facteurs intrinsèques**

Dans le modèle mathématique SIR, qui décrit la dissémination d’une maladie par contact direct dans une population, il est possible d’obtenir une dynamique cyclique lorsque l’on renouvelle continûment les individus susceptibles. Cette dynamique est cependant amortie vers un point d’équilibre stable. L’inclusion de distributions réalistes pour les durées de latence ou infectieuse, comme celle de composantes stochastiques, fragilise cette propension à l’oscillation.

En revanche, dès lors que l’on étend la modélisation pour incorporer plusieurs populations en interaction, des dynamiques source–puits peuvent voir le jour et expliquer la survenue d’épidémies. La prise en compte des pathogènes en compétition ou en interaction, par exemple plusieurs souches ou une diversité virale, permet également de voir apparaître des cycles plus ou moins réguliers, voire des trajectoires chaotiques.

Dans des réseaux complexes, l’existence de cycles est encore peu abordée, bien qu’il ait été décrit que, dans certaines conditions de croissance du réseau, des cycles dans l’incidence pourraient être présents.

**Les facteurs extrinsèques**

L’évolution démographique propre peut être motrice dans l’existence de cycles, par exemple en imposant un remplacement de la population susceptible lent par rapport à la dynamique de la maladie.

La régularité du forçage saisonnier suggère également que la relation hôte–pathogène ait pu en être affectée. Effectivement, nombre de maladies transmissibles ont un cycle saisonnier, que celui-ci soit induit par les vacances scolaires ou par d’autres modifications, même ténues, des comportements. La résistance maximale face à l’invasion par un autre pathogène mènerait à cependant à exacerber les variations saisonnières, voire à perdre la synchronisation pour adopter un cycle pluri-annuel. L’influence du climat est également démontrée, qu’il crée les conditions d’existence par un vecteur ou par des modalités encore indéterminées.

**Conclusion**

Il n’existe pas encore de théorie unique pour expliquer la totalité des manifestations cycliques et récurrentes des épidémies. La prise en compte des multiples acteurs dans des systèmes réalistes sera source de travaux dans les prochaines années.

1. Introduction

When it was first used by Hippocrates in a medical context, the Greek word ‘epidemic’ had already changed meanings from an initial ‘back in his country’ for the more dynamical ‘propagating in the country’, adequately qualifying the spread of a disease [1]. Still, the very etymology of the word made it clear that ‘epidemics’ were ‘at home’ in human populations that they visited from times to times [2]. Indeed, while outbreaks of communicable diseases may be limited in space and time, their propensity to recur place a looming threat on human populations. In May 2003, an editorial in the New England Journal of Medicine about SARS concluded: “if we are extremely lucky, the epidemic will be curtailed, develop a seasonal pattern that will improve prospects for regional containment” [3], showing that recurrence of the disease was taken seriously by experts in the field. This prediction did not realize, illustrating that our current understanding of epidemics allows
beting rather than logic. Indeed, the factors making recurrence more likely remain unclear, as multiple relationships, acting at various levels, may be required [4].

Many infectious diseases present a seasonal pattern, especially those caused by respiratory pathogens [5]. Interpandemic influenza epidemics occur during fall and winter in temperate countries of the northern hemisphere [6]. Respiratory syncitial virus activity peaks in the same period, although seemingly independently [7]. Rotavirus infection displays also a very seasonal pattern, especially among the young [8], as do other childhood diseases [9]. In the latter, seasonal patterns vary with location, with yearly epidemics (for example, in France [10]) or epidemics of varying size every other year, or do not exist at all in the absence of imported cases [11]. Other diseases present a pseudo periodic behaviour that is more loosely related with seasonal variation. Chikungunya fever is reportedly recurring every four years in Senegal [12]. Smallpox outbreaks occurred with a frequency of five to seven years [13,14], while meningococcal meningitis presented cycles of 10 years [15]. Syphilis also shows recrudescence periods separated by approximately 11 years [16], while gonorrhea or HIV infection, although transmitted mainly by the same route, do not show evidence of periodic behaviour.

These various situations illustrate diseases that perpetuate in human populations: cases may be found without interruption, even if incidence varies [17]. Seasonal outbreaks may rhythm the perpetuation, or seemingly haphazard outbreaks interspersed with large silent periods. In all cases, perpetuation requires that a portion of the population is susceptible to the disease, through either demographic replacement, or loss of immunity. Recurrence may require a reservoir population for the pathogen, which may be the human population itself or some interacting species. In the following, we will not discuss the case of zoonotic diseases unable to perpetuate in humans, but introduced from time to time by direct exposure: plague, for example, but also the current form of avian influenza.

The relative importance of intrinsic and extrinsic factors in the perpetuation and recurrence of infectious diseases is debated [18,19]. Intrinsic factors are those coming directly into play in the interaction between a pathogen population and its host population. Extrinsic factors are those that may modulate this interaction, but not be affected by it, like climate for instance.

In this article, we review some theoretical results regarding perpetuation and recurrence of communicable diseases. In a first part, intrinsic factors are explored, and extrinsic factors are then reviewed.

2. Intrinsic factors associated with perpetuation and recurrence

A number of determinants for the perpetuation of transmissible diseases in human populations have been listed [17], highlighting the importance of both demographic and biological parameters. Demographical parameters include the size of the population, the rate of population turnover, the rate and pattern of contacts; biological parameters include the duration of immunity, the susceptibility and transmissibility, the duration of infectiousness and the generation period (the duration between an index case and a secondary case). The importance of these factors is illustrated in the following paragraphs.

2.1. One population, one pathogen: the SIR model

In the famous SIR model, the human population is split in three subgroups: (S)usceptibles to infection, (I)nfectious and (R)emoved by cure or death, with the following differential equations describing contamination, cure and removal [20]:

\[
\begin{align*}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{align*}
\]

This basic model aims at describing the circulation of a pathogen in a single host population, leaving those infected immune or dead. These equations do not allow for perpetuation or recurrence unless a demographic birth and death process is also implemented or a loss of immunity with time taken into account in those already recovered. In either case, an equilibrium state where infectious perpetuate may be obtained, with quasi-periodic oscillations around the equilibrium. The approximate period of the oscillations is \( T = 2\pi \sqrt{DA} \), where \( D = 1/\gamma \) is the average duration of the infectious period and \( A = \mu(1 - \frac{1}{R_0}) \) is the average age at infection, with \( \mu \) the birth (or loss of immunity) rate and \( R_0 = \beta/(\mu + \gamma) \) the basic reproduction number of the disease (i.e. the number of secondary cases borne from one initial index case in a completely susceptible population) [21]. The period of oscillations is therefore increased by a low birth rate (or low loss of immunity) and a short duration of the disease.

The occurrence of oscillations is a direct consequence of the form of the interaction between susceptible and infected individuals (i.e. the law of mass action) in a finite population, leading to a critical threshold in the number of susceptible individuals below which the
incidence of new cases fails to compensate those removed; the dynamics of births, which replenishes the susceptible compartment with time and allow for renewed increase in prevalence. This intrinsic periodic behaviour was first thought to account fully for recurrent epidemics [22], all the more because almost yearly periods may be obtained with reasonable parameter values. However, the oscillations damp with time and fail to explain sustained seasonal cycles [23]. Inclusion of realistic distributions for the duration of the infectious period (in the SIR model, an exponential distribution is assumed) shows that, while damping may take considerable time, it is nevertheless present [24]. Introducing stochastic components changes the behaviour of models with respect to perpetuation in an essential way, since there is a possibility of disease extinction [25]. A stochastic treatment to the SIR model leads to undamped oscillations in incidence, but the disease goes extinct with probability 1 [26]. Furthermore, taking into account realistic distributions for the epidemic period even hasten this process, as the troughs are more pronounced [24]. Therefore, although the simple SIR model provides for both perpetuation and recurrence, it fails short to provide a satisfying explanation to observed infectious disease dynamics.

2.2. Several populations, one pathogen

Having multiple coupled populations somewhat alleviates the problem of extinction. Measles in the UK has provided ample evidence of source dynamics explaining recurrences in isolated populations. In London, for example, measles never goes extinct, since the size of the children population is large enough to perpetuate the disease, with a simple seasonal pattern (see also extrinsic factors below). On the contrary, small places show scattered epidemics, interspersed by irregular time intervals [27]. These ‘meta-population’ dynamics may have vastly diverging effects depending on the disease. For example, measles dynamics tends to be synchronized through these interactions, while whooping cough, with an infectious period approximately three times longer, is desynchronized [28]; this is explained by a subtle interplay between the duration of the infectious period, the coupling between populations and seasonal forcing. Perpetuation may be more likely with desynchronized epidemics in a meta-population approach.

A very different meta-population dynamics is seen in vector-borne diseases. In dengue and malaria, the pathogen is transmitted by mosquitoes. The disease dynamics is therefore subject to the population dynamics of both species. Interestingly, this translated into an almost three-year interepidemic period for both diseases, in very good agreement with the prediction of a SEIR model [29]. More generally, multi-host infectious diseases may be associated with original perpetuation and recurrence characteristics. This is so because multi-host pathogens have generally reservoir hosts in which they persist with little clinical manifestation, and cause outbreaks outside this reservoir [30]. Even if malaria was one of the first diseases submitted to a mathematical treatment, there is still considerable work to gain further insight into the emerging zoonotic threat [31].

2.3. One population, several pathogens

In the preceding, a limited view of the interaction between pathogens and humans was adopted, taking no notice of diversity in the pathogen population. This may be true for diseases like measles, where little viral diversity exist. However, the influenza virus effectively escapes immunity by constant changes at key antigenic sites [32], causing an effective influx of susceptible individuals. The joint study of pathogen diversity and epidemic transmission is the subject of much current research [33].

When several strains of a pathogen co-circulate, we may expect independent behaviour and coexistence at levels determined by the relative fitness of each strain. However, this does not account for potential interference between strains: competing for susceptible hosts [34], reducing susceptibility to infection by genetically ‘close’ strains [35] or transmissibility after an initial infection [36]. With these refinements, unconstrained coexistence may be the rule at low levels of cross immunity, but only non-interfering strains may coexist at very high levels of cross immunity. Furthermore, in a range of intermediate levels of cross immunity, periodic or chaotic cycling between strains is possible [35–38]. In this case, the secular variation seen in the prevalence of the strains of some pathogens may result from complex interplay due to cross immunity. Examples conforming to these predictions may include lineages of N. meningitidis and serotypes of group-A streptococcus [39]. If the co-circulating strains have different pathogenicity, periodic or recurrent outbreaks will occur [40].

It is noteworthy that such rich dynamics may also be observed in a very particular case of interference known as ‘antibody-dependent enhancement’ [41,42]. In this case, previous infection does not act to reduce susceptibility or transmission to a new infection, but on the contrary may increase transmissibility during a second infection and infectiousness. This has been proposed in
the case of dengue (although it is debated [40]), and for other viruses as well (among others coronaviruses of the SARS type [43]). Modelling shows that antibody enhancement may lead to undamped cyclical epidemics with different strains, much like more classical interactions.

Diversity in pathogens is, like transmission, a dynamic process, and should be considered in a ‘phylogenetics’ approach [33]. Models were proposed in this respect where strain diversity changes with time given a constant mutation rate, accompanied by a decrease in cross immunity with increasing genetic distance [44,45]. In a detailed analysis of the behaviour of this model, it appears that combining short infectious duration with long-lived immunity induced by infection (total immunity against the same strain, and partial immunity to other strains) lead to the occurrence of renewed outbreaks with time, where at each time the dominant strain is different from that in earlier epidemics [45]. Calibrating the model to reflect influenza, it was found that the simulated phylodynamics were reminiscent of that observed in real influenza [44]: several closely related genetic strains co-circulating at any given time, with a continuous drift [46].

2.4. Networks and heterogeneous mixing

An assumption of the basic SIR model is that of ‘homogeneous mixing’, whereby all infectious may contact all the susceptible individuals. This also means that little variation is expected in the number of secondary cases borne from two different infectious cases, since the effective contacts take the form of a completely random graph. Studying the impact of heterogeneity in transmission was first addressed by splitting compartments according to individual characteristics. The impact of such structuring was manifest in the study of sexually transmitted infections (STI), and led to the influential concept of ‘core groups’ having more numerous partners than the rest of the population [47]. These groups were essential to explain the maintenance of STIs in a population where it would have otherwise disappeared.

The recent epidemic of SARS has put forward that in other diseases as well, the number of secondary cases could be highly variable. In fact, most cases were without descent, while few cases had a disproportionately high number of descendants [48]. The so-called ‘superspreaders’ have since then been recognized as fairly common in many communicable diseases, appearing as the norm rather than the exception [49]. While superspreaders may have a definite role in introducing a disease in a population, and be targeted first to prevent its spread, it is not clear as of today how their role would extend in perpetuating a disease.

Superspreaders echo findings from complex network theoretical epidemiology. In those, the topology of contact networks has been changed from completely random to include small world or scale-free properties. A major finding of these studies has been the strong dependence of epidemic spread on network structure [50]. For example, contrary to random networks, infinite power-law networks do not display threshold behaviour for the persistence of a disease, even if a threshold-like phenomenon may be recovered in finite-size networks [51]. Provided that susceptible individuals are introduced or recovered ones lose their immunity, the perpetuation of any disease is theoretically granted in such networks. Indeed, any initial number of infected individuals may lead to large-scale epidemics. Oscillatory dynamics in prevalence may appear in growing networks [52], while in other cases perpetuation exists without periodicity.

3. Extrinsic properties leading to recurrence

A large number of extrinsic factors have been suggested as promoting recurrence or perpetuation of a communicable disease, especially those based on seasonal forcing. From an evolutionary point of view, the existence of a regular forcing imposed by seasons must have led to selecting the type of interactions that allows perpetuation of pathogens in the host populations. Other explanations involve seasonal demographic processes leading to changes in host behaviour, to environmental changes or to changes in pathogen presence.

3.1. Demographics

The mere size of the population has a profound effect on the perpetuation of diseases. Large populations enable perpetuation, whereas the disease generally disappears in small populations. The critical community size corresponds to the threshold below which perpetuation is not possible [53]. Changing birth rates may profoundly affect the pattern of recurrence of diseases. This is best exemplified for measles epidemics, where sustained annual epidemics require high birth rates. In the UK, decreasing birth rates have been associated with the occurrence of large epidemics every other year [54]. Importantly, external stimulation of the model through the changing birth rates was sufficient for the trajectory to hop between attractors having cycles of various lengths [54]. The same profound effect of birth rates was seen for smallpox. Indeed, the intrinsic period of
this disease, in a population of constant size, is 3 to 4 years; however, inspection of past time series show preferentially 5-year cycles [13], or even longer. The explanation of this period lengthening has been linked to extrinsic processes, involving for example the occurrence of famine and high wheat price [13].

3.2. Seasonal changes in host behaviour

Many human activities exhibit a seasonal component, and this may have implications in disease transmission. For example, childhood diseases exhibit marked seasonality, although patterns may change between places. Incidence increases during the whole school year, and drops in the summer. A very good candidate to explain this seasonality pattern is the forcing imposed by school holidays, and it has been repeatedly found that this leads to improved fit to observed data [55]. For diseases with long infectious periods, school-term forcing leads to annual cycles. When the duration of the infectious period decreases, as well as the effective number of secondary cases, longer cycles may occur [56]. Large differences in contact rates also lead to longer cycles. On the contrary, most respiratory diseases exhibit a seasonal pattern with incidence increase during the cold season. Importantly, this is not a characteristic of the disease only, as it is well documented that the seasonal component increases with latitude: tropical countries experience similar attack rates as temperate countries, only the epidemic period is diffuse, instead of concentrated during a few weeks [5].

Up to now, no single theory has been sufficient to explain all seasonal patterns. First, there may be changes in contact rates as a consequence of crowding in colder seasons. It was recently proposed that even subtle changes in contact rate could lead to large amplitude changes due to resonance between the intrinsic cycle of the disease and that in contact rates [57]. In any case, seasonal forcing in the contact rate in SIR-type models introduces extremely rich dynamics [54]. It has also be suggested, and found in mice, that seasonal changes in immunity of the host were at stake [5].

Given that there is so much evidence for seasonal changes, one may wonder on the advantage of seasonal variation in incidence for perpetuation of a disease. We must recall first that the dynamics of the SIR model intrinsically leads to oscillations in the number of susceptible individuals. If there is a good correlation in the intrinsic cycle of the disease and seasonality in contact rates (i.e. if the pathogen has been selected so), the pathogen becomes more firmly established as the magnitude of seasonal changes increases. However, the same argument leads to favour an increase in duration between recurrences of the disease, and ultimately an escape from the seasonal behaviour [58]. Further theoretical results are required to resolve these apparent paradoxes.

3.3. Environment: weather, satellites and El Niño

It is well known that the environment plays a key role in the distribution of human diseases [59], through climatic factors. Could global processes be associated with the occurrence of disease in particular places? While not causal determinants, global processes could favour environmental conditions leading to recurrence or perpetuation. El Niño occurs irregularly (cycles between two and seven years). The El Niño cycle is associated with increased risks of some of the diseases transmitted by mosquitoes, such as malaria, dengue, and Rift Valley fever [60], but also of cholera [61]. Candidate mechanisms linked to the effect of environment on the vector, or on water movement have been proposed. More unexpectedly, the cold event of the El Niño oscillation (or La Niña) is also associated with higher influenza mortality [62].

Other environmental factors, like vegetation indices, have been linked to the occurrence of vector-borne diseases like the Rift Valley fever disease [63]. These indicators would be extremely helpful for predicting and organizing containment measures: this is a way forward for the coming years [64].

4. Conclusion

The natural movement of human populations is a key determinant to the perpetuation of communicable diseases. While it was laid down very early that “the recurrence of epidemics depends solely on two factors, the time of importation of the morbid poison and the number of persons susceptible to it” [65], mathematical models have been instrumental in suggesting which mechanisms ruled importation and susceptibility. For some diseases, where the virus is not to genetically diverse and does not exist in external reservoir, it seems that the determinants of the oscillatory-like dynamics of communicable diseases are well characterized: this is now to a large extent the case for measles. As the 20th century saw the solution to the one population/one pathogen problem, the 21st century must be the time when complex systems, involving multiple populations, multiple pathogens and the environment will be fully understood [66].
Current interest in influenza shows the complexity of this task, and how many domains will have to communicate: molecular techniques may help to precise how strains evolve and perpetuate in real epidemics [67], and large-scale computer simulations how epidemic spreads in realistic populations [68].

What will make all these developments possible is a renewed effort into observation, to correlate better observations now taken disparately: human, veterinarian, and the environment. The recent studies into cholera-recurrence dynamics, implicating first El Niño events [61] as likely to promote outbreaks, then more recently linked to the vibriophages’ population [69], illustrate perfectly the way to follow in the future: reasoning at several scales, using models to assemble detailed information gathered in the field in a comprehensive framework for analysis.

References


