



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## Chapter 18

# Viral Epidemiology

## Tracking Viruses with Smartphones and Social Media

Kaitlin Rainwater-Lovett<sup>1</sup>, Isabel Rodriguez-Barraquer<sup>2</sup>, William J. Moss<sup>3</sup>

<sup>1</sup>Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD, USA; <sup>2</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; <sup>3</sup>Departments of Epidemiology, International Health and Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

### Chapter Outline

<b>1. History</b>	<b>241</b>	4.1 Movement Loggers	248
1.1 Traditional Epidemiological Methods	242	<b>5. Transmission Networks</b>	<b>248</b>
1.2 Methods in Infectious Disease Epidemiology	242	5.1 Contact Patterns	249
<b>2. Surveillance</b>	<b>243</b>	<b>6. Molecular Epidemiology</b>	<b>249</b>
2.1 Novel Data Streams	243	<b>7. Phylodynamics</b>	<b>249</b>
2.2 Communicating Surveillance Data	244	<b>8. Mathematical Models of Transmission Dynamics</b>	<b>251</b>
<b>3. Viral Infections in Space and Time</b>	<b>245</b>	8.1 Agent-Based Models	251
3.1 Remotely Sensed Data	246	<b>9. Synthesis</b>	<b>251</b>
<b>4. Population Movement</b>	<b>248</b>	<b>Further Reading</b>	<b>252</b>

The science of epidemiology has been developed over the last 200 years, using traditional methods to describe the distribution of diseases by person, place, and time. However, in the last several decades, a new set of technologies has become available, based on the methods of computer sciences, systems biology, and the extraordinary powers of the Internet. Technological and analytical advances can enhance traditional epidemiological methods to study the emergence, epidemiology, and transmission dynamics of viruses and associated diseases. Social media are increasingly used to detect the emergence and geographic spread of viral disease outbreaks. Large-scale population movement can be estimated using satellite imagery and mobile phone use, and fine-scale population movement can be tracked using global positioning system (GPS) loggers, allowing estimation of transmission pathways and contact patterns at different spatial scales. Advances in genomic sequencing and bioinformatics permit more accurate determination of viral evolution and the construction of transmission networks, also at different spatial and temporal scales. Phylodynamics links evolutionary and epidemiological processes to better understand viral transmission patterns. More complex and realistic

mathematical models of virus transmission within human and animal populations, including detailed agent-based models, are increasingly used to predict transmission patterns and the impact of control interventions such as vaccination and quarantine. In this chapter, we will briefly review traditional epidemiological methods and then describe the new technologies with some examples of their application.

### 1. HISTORY

Insight into the epidemiology of viral infections long preceded the recognition and characterization of viruses as communicable agents of disease in humans and animals, extending at least as far back as the treatise of Abu Becri (Rhazes) on measles and smallpox in the tenth century. Successful efforts to alter the epidemiology of viral infections can be traced to the practice of variolation, the deliberate inoculation of infectious material from persons with smallpox (see chapter on the history of viral pathogenesis). Documented use of variolation dates to the fifteenth century in China. Edward Jenner greatly improved the practice of variolation in 1796 using the less-virulent cowpox virus,

establishing the field of vaccinology. An early example of rigorous epidemiological study prior to the discovery of viruses was the work of the Danish physician Peter Panum who investigated an outbreak of measles on the Faroe Islands in 1846. Through careful documentation of clinical cases and contact histories, Panum provided evidence of the contagious nature of measles, accurate measurement of the incubation period, and demonstration of the long-term protective immunity conferred by measles. The discovery of viruses as “filterable agents” in the late-nineteenth and early twentieth centuries greatly enhanced the study of viral epidemiology, allowing the characterization of infected individuals, risk factors for infection and disease, and transmission pathways.

### 1.1 Traditional Epidemiological Methods

Traditional epidemiological methods measure the distribution of viral infections, diseases, and associated risk factors in populations in terms of person, place, and time using standard measures of disease frequency, study designs, and approaches to causal inference. Populations are often defined in terms of target and study populations, and individuals within study populations in terms of exposure and outcome status. The purpose of much traditional epidemiological research is to quantify the strength of association between exposures and outcomes by comparing characteristics of groups of individuals. Exposures or risk factors include demographic, social, genetic, and environmental factors, and outcomes include infection or disease. In viral epidemiology, infection status is determined using diagnostic methods to detect viral proteins or nucleic acids, and serologic assays to measure immunologic markers of exposure to viral antigens. Infection status can be defined as acute, chronic, or latent.

Standard measures of disease frequency include incidence, the number of new cases per period of observation (e.g., 1000 person-years), and prevalence, the number of all cases in a defined population and time period. Prevalence is a function of both incidence and duration of infection and can increase despite declining incidence, as observed with the introduction of antiretroviral therapy for human immunodeficiency virus (HIV) infection in the United States. Although the number of new cases of HIV infection declined, the prevalence of HIV infection increased as treated individuals survived longer. Commonly used study designs include,

- cross-sectional studies in which individuals are sampled or surveyed for exposure and disease status within a narrow time frame,
- cohort studies in which exposed and unexposed individuals are observed over time for the onset of specified outcomes,

- case-control studies in which those with and without the outcome (infection or disease) are compared on exposure status, and
- clinical trials in which individuals are randomized to an exposure such as a vaccine or drug and observed for the onset of specified outcomes

Appropriate study design, rigorous adherence to study protocols, and statistical methods are used to address threats to causal inference (i.e., whether observed associations between exposure and outcome are causal), such as bias and confounding.

Much can be learned about the epidemiology of viral infections using such traditional methods and many examples could be cited to establish the importance of these approaches, including demonstration of the mode of transmission of viruses by mosquitoes (e.g., yellow fever and West Nile viruses), the causal relationship between maternal viral infection and fetal abnormalities (e.g., rubella virus and cytomegalovirus), and the role of viruses in the etiology of cancer (e.g., Epstein–Barr and human papilloma viruses).

### 1.2 Methods in Infectious Disease Epidemiology

The epidemiology of communicable infectious diseases is distinguishable from the epidemiology of noncommunicable diseases in that the former must account for “dependent happenings.” This term was introduced by Ronald Ross to capture the fact that infectious agents are transmitted between individuals or from a common source. Traditional epidemiological and statistical methods often assume disease events in a population are independent of one another. In infectious disease epidemiology, individuals are defined in terms of susceptible, exposed, infectious, and recovered or immune. Key characteristics of viral infections that determine the frequency and timing of transmission, and thus the epidemiology, include the mode of transmission (e.g., respiratory, gastrointestinal, sexual, blood-borne, and vector-borne), whether infection is transient or persistent, and whether immunity is short or long lasting. Temporal changes in the transmission dynamics of viral infection can be displayed with epidemic curves, by plotting the number or incidence of new infections over time to demonstrate outbreaks, seasonality, and the response to interventions.

Key metrics in infectious disease epidemiology that capture the dependent nature of communicable diseases include: (1) the latent period, the average time from infection to the onset of infectiousness; (2) the infectious period, the average duration of infectiousness; (3) the generation time, the average period between infection in one individual and transmission to another; and (4) the basic reproductive number ( $R_0$ ), the average number of new infections initiated by a single infectious individual in a completely susceptible population over the course of that individual’s infectious period.

If  $R_0$  is larger than one, the number of infected individuals and hence the size of the outbreak will increase. If  $R_0$  is smaller than one, each infectious individual infects on average less than one other individual and the number of infected individuals will decrease and the outbreak ceases. The reproductive number ( $R$ ) is a function not only of characteristics of the viral pathogen (e.g., mode of transmission), but also the social contact network within which it is transmitted and changes over time in response to a decreasing number of susceptible individuals and control interventions. An important concept related to the interdependence of transmission events is herd immunity, the protection of susceptible individuals against infection in populations with a high proportion of immune individuals because of the low probability of an infectious individual coming in contact with a susceptible individual.

The concepts and methods of infectious disease epidemiology provide the tools to understand changes in temporal and spatial patterns of viral infections and the impact of interventions. Traditional epidemiological methods provide powerful analytical approaches to measure associations between exposures (risk factors) and outcomes (infection or disease). Recent technological advances enhance these methods and permit novel approaches to investigate the emergence, epidemiology, and transmission dynamics of viruses and associated diseases.

## 2. SURVEILLANCE

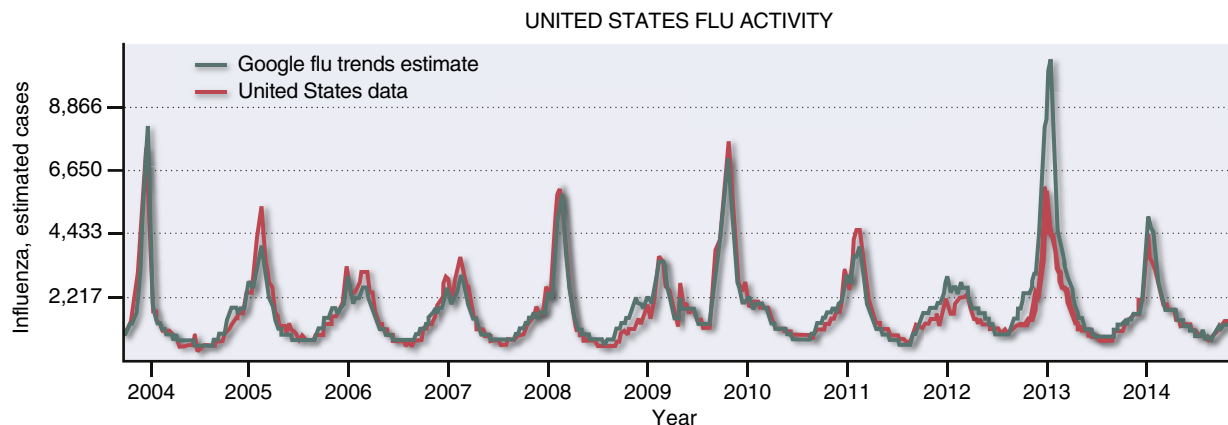
Expanded access to the Internet and social media has revolutionized outbreak detection and viral disease surveillance by providing novel sources of data in real time (Chunara, 2012). Traditional epidemiologic surveillance systems rely on standardized case definitions, with individual cases typically classified as suspected, probable, or confirmed based on the level of evidence. Confirmed cases require laboratory evidence of viral infection. Surveillance systems are either

active or passive. Active surveillance involves the purposeful search for cases within populations whereas passive surveillance relies on routine reporting of cases, typically by health care workers, health care facilities, and laboratories. Data acquired through active surveillance are often of higher quality because of better adherence to standardized case definitions and completeness of case ascertainment but are more expensive and resource intensive. However, both active and passive surveillance are prone to delays in data reporting.

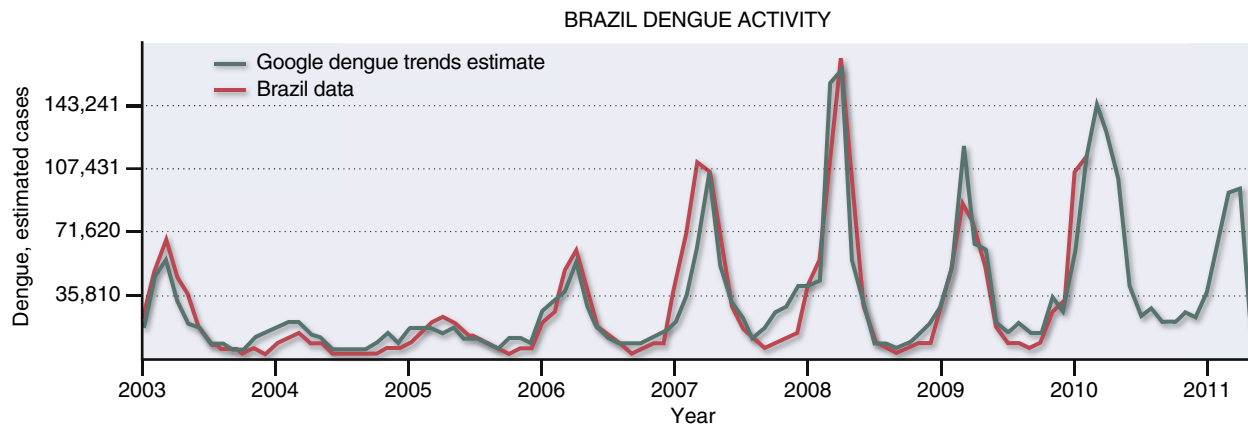
The major advantage of using the Internet and social media to monitor disease activity is that the signal can be detected without the lag associated with traditional surveillance systems. Influenza is the most common viral infection for which the Internet and social media have been used for disease surveillance because of its high incidence, wide geographic distribution, discrete seasonality, short symptomatic period, and relatively specific set of signs and symptoms. However, the Internet and social media have several limitations compared to traditional active and passive surveillance systems and complement rather than replace these methods. These limitations include lack of specificity in the “diagnosis,” and waxing and waning interest and attention in social media independent of disease frequency.

### 2.1 Novel Data Streams

In 2008, the Internet company Google developed a web-based tool called Google Flu Trends, for early detection of influenza outbreaks. Google Flu Trends is based on the fact that millions of people use the Google search engine each day to obtain health-related information (Ginsberg, 2009). Logs of user key words for pathogens, diseases, symptoms, and treatments, as well as information on user location contained in computer Internet Protocol (IP) addresses, allow temporal and spatial analyses of trends in search terms (Figure 1). Early results suggested that Google Flu Trends



**FIGURE 1** Number of cases of influenza in the United States from 2004 to 2013 estimated by Google Flu Trends and the Centers for Disease Control and Prevention. The number of cases of influenza estimated by Google Flu Trends closely approximated the number of cases estimated by the Centers for Disease Control and Prevention. Redrawn from <http://www.google.org/flutrends/about/how.html>.



**FIGURE 2** Number of cases of dengue in Brazil from 2003 to 2011 estimated by Google Dengue Trends and the Ministry of Health, Brazil. The number of cases of dengue estimated by Google Dengue Trends closely approximated the number of cases estimated by the Ministry of Health of Brazil. *Redrawn from <http://www.google.org/denguetrends/about/how.html>.*

detected regional outbreaks of influenza 7–10 days before conventional surveillance by the Centers for Disease Control and Prevention (Carneiro, 2009). However, accurate prediction was not as reliable as initially thought, and Google estimates did not closely match measured activity during the 2012–2013 influenza season. Google now reevaluates estimates using data from traditional surveillance systems (specifically those of the Centers for Disease Control and Prevention) to refine model and parameter estimates. These refinements more accurately capture the start of the influenza season, the time of peak influenza virus transmission, and the severity of the influenza season.

A similar approach, called Google Dengue Trends, is used to track dengue virus infections by aggregating historical logs of anonymous online Google search queries associated with dengue, using the methods developed for Google Flu Trends. Early observations suggest Google queries are correlated with national-level dengue surveillance data, and this novel data source may have the potential to provide information faster than traditional surveillance systems (Figure 2).

Other Internet sources are being explored to enhance viral surveillance. Wikipedia is a free, online encyclopedia written collaboratively by users and is one of the most commonly used Internet resources since it was started in 2001. As with Google searches, the use of disease-specific queries to Wikipedia are expected to correlate with disease activity. The number of times specific influenza-related Wikipedia sites were accessed provided accurate estimates of influenza-like illnesses in the United States 2 weeks earlier than standard surveillance systems and performed better than Google Flu Trends (McIver, 2014).

Similarly, social media data are being evaluated for surveillance purposes. Twitter is a free social networking service that enables users to exchange text-based messages of up to 140 characters known as tweets. As with Google Flu Trends, the number of tweets related to influenza activity

is correlated with the number of symptomatic individuals. Several published studies reported correlations between Twitter activity and reported influenza-like illnesses (Chew, 2010; Signorini, 2011; Figure 3).

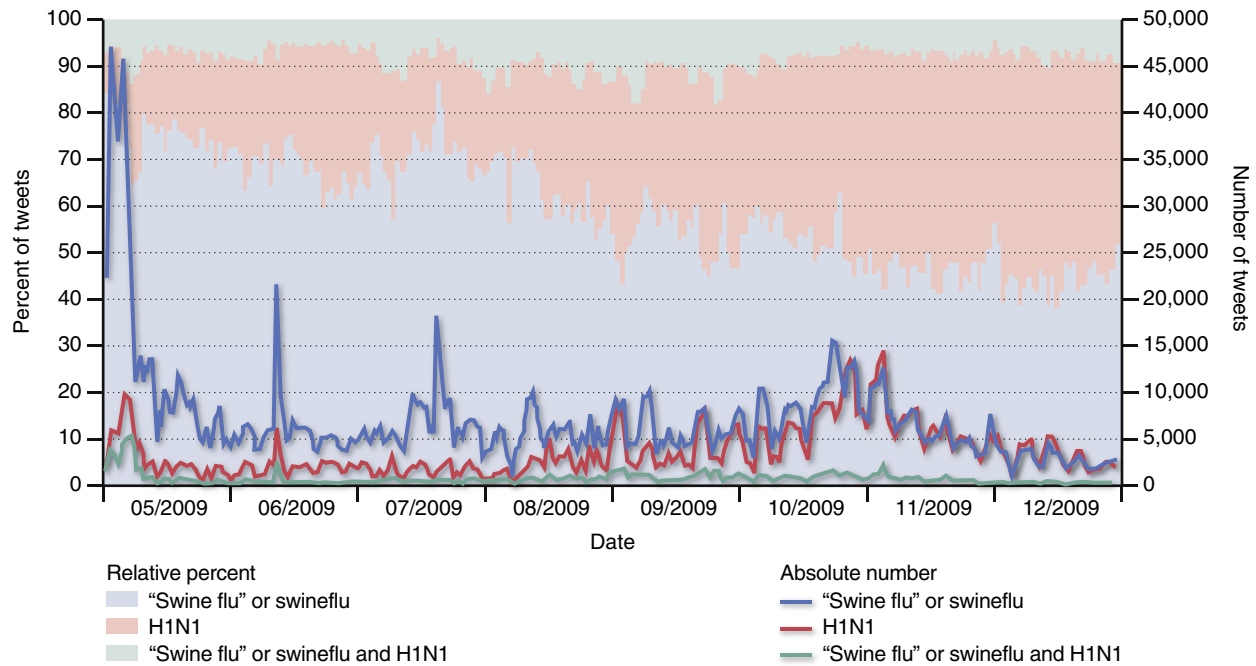
Limitations to using social media, such as Twitter, to monitor disease activity are illustrated by the Ebola virus outbreak in West Africa in early 2014. Despite the fact that Ebola had not yet occurred in the United States, posts to Twitter on Ebola rose dramatically, likely in response to intense media coverage and fear. Clearly, such tweets could not be interpreted to indicate Ebola disease activity in the United States. Studies reporting misleading associations, or the lack of correlation between social media and disease activity, are rarely published, providing a cautionary note.

While initial efforts using data from the Internet for viral disease surveillance offer promising results, concerns have been raised regarding the utility and robustness of these approaches (Lazer, 2014). Integration into existing surveillance frameworks will be necessary to maximize the utility of these data streams.

## 2.2 Communicating Surveillance Data

The Internet allows rapid processing and communication of health-related information, including the aggregation and display of surveillance data for viral infections. Traditional surveillance networks can be linked through the Internet to allow rapid integration and dissemination of information. Information on viral disease outbreaks available through Internet postings of health care agencies such as the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC), as well as press reports and blogs, can provide data that are more current than traditional surveillance systems. Information from these online sources can be made available to a large, global audience. Several of the most





**FIGURE 3** Absolute number and relative percent of tweets containing H1N1, swine flu, or both from May to December 2009. The lines show the absolute number of tweets and the shading shows the relative percent of tweets. The blue lines and shading indicate tweets specific for swine flu, the red lines and shading indicate tweets specific for H1N1 and the green lines and shading indicate tweets containing both terms. The data highlight monthly variation in tweets and the increase in the relative percent of tweets specific for H1N1 over the study period. *Redrawn from Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010; 5: e14118.*

#### Sidebar 1 Sources of information on outbreaks of viral diseases

ProMED, the Program for Monitoring Emerging Diseases, is an Internet-based reporting system established in 1994 that compiles information on outbreaks of infectious diseases affecting humans, animals, and food plants. ProMED relies on official announcements, media reports, and local observers, including the network of subscribers. A team of experts screen, review, and investigate reports before posting and often provide commentary. Reports are distributed by email to direct subscribers and posted on the ProMED-mail Web site. ProMED-mail currently reaches over 60,000 subscribers in at least 185 countries.

Started in 2006 by epidemiologists and software developers at Boston Children's Hospital, HealthMap monitors disease

outbreaks and provides real-time surveillance of emerging public health threats, including viral infections (Figure 4). HealthMap organizes and displays data on disease outbreaks and surveillance using an automated process. Data sources include online news aggregators, eyewitness reports, expert-curated discussions and validated official reports.

#### Internet Resources

Google flu trends	<a href="http://www.google.org/flutrends">http://www.google.org/flutrends</a>
Google dengue trends	<a href="http://www.google.org/denguetrends">http://www.google.org/denguetrends</a>
HealthMap	<a href="http://healthmap.org">http://healthmap.org</a>
ProMED	<a href="http://www.promedmail.org">http://www.promedmail.org</a>

commonly used surveillance sites report animal as well as human diseases (see Sidebar 1 and Figure 4).

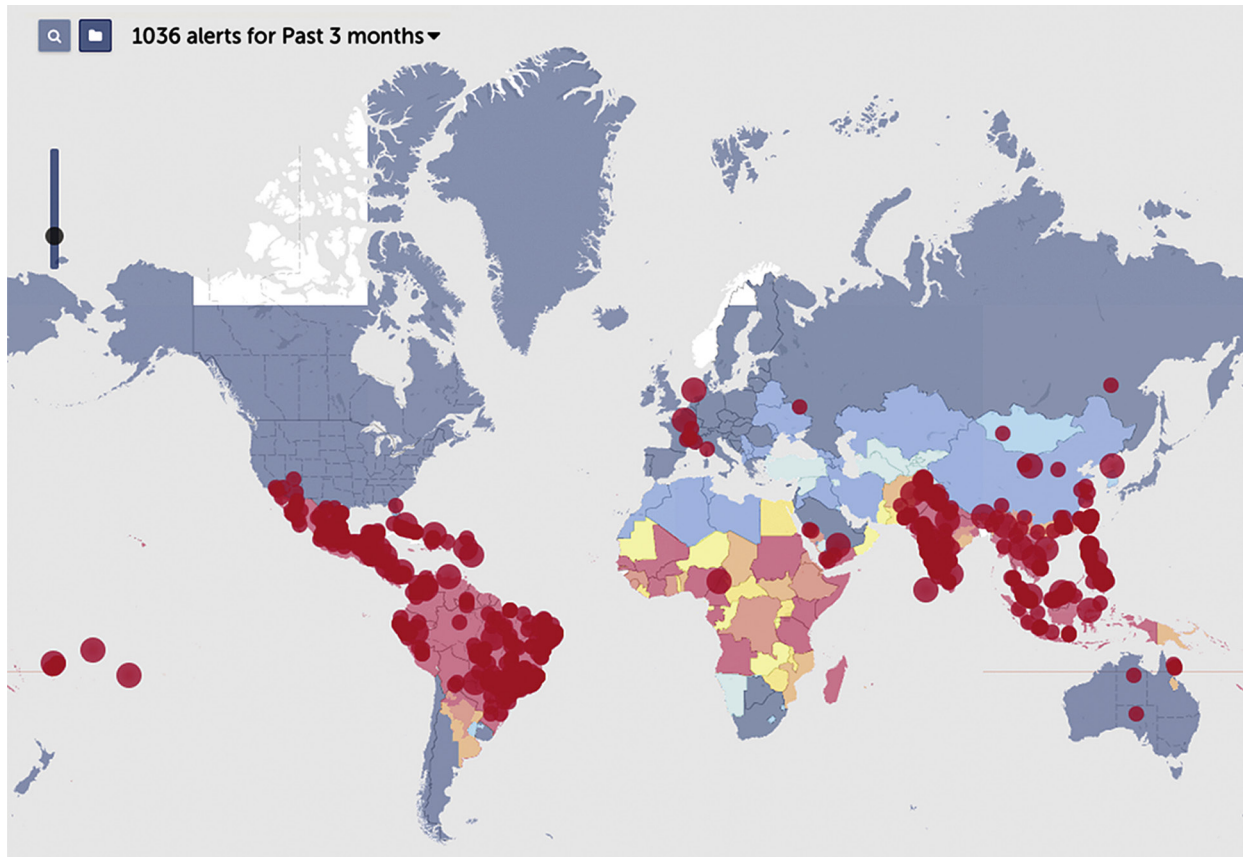
### 3. VIRAL INFECTIONS IN SPACE AND TIME

Mapping spatial patterns of disease and relationships with environmental variables preceded the development of modern epidemiology. The classic example is John Snow's hand-drawn map of London cholera cases of 1854. However, routine mapping of health data only became commonplace in the 1990s after desktop geographic information systems became widely available. Combined with satellite imagery and remotely sensed environmental and ecological

data, spatial mapping of viral infections is a powerful tool for surveillance and epidemiological research.

Spatial epidemiology is typically used to identify and monitor areas of differential risk. An early example was a large outbreak of St. Louis encephalitis virus infection in Houston, Texas in 1964. Spatial analysis showed that the outbreak was concentrated in the city center, with lower incidence at the outskirts. Further investigation revealed that the city center was associated with the lowest economic strata, unscreened windows, lack of air-conditioning and pools of standing water, factors facilitating virus transmission.

Investigation into the spatiotemporal dynamics of viral diseases at smaller spatial scales has become



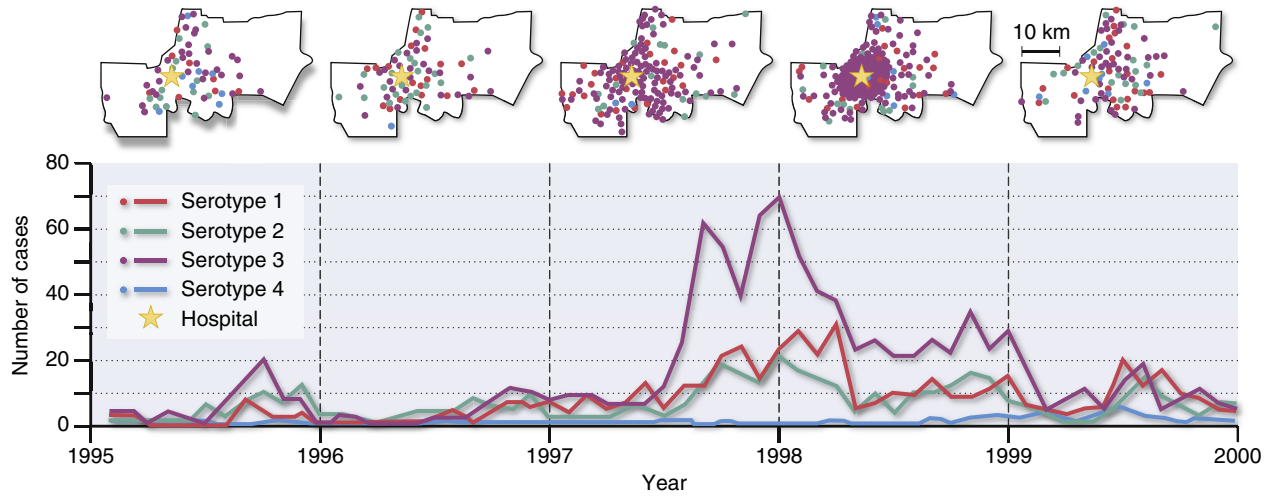
**FIGURE 4** HealthMap display showing reports of global dengue cases from official sources and media reports. The large red circles indicate reports at the country level and the small red circles indicate reports at the local level. The risk of dengue at a country level is indicated by the color: blue absent or unlikely; yellow uncertain; orange likely; purple known. From <http://healthmap.org/dengue/en/>.

possible with increasing availability of global positioning systems devices and geocoding algorithms. Such studies have revealed spatial heterogeneity in the local transmission of some directly (e.g., HIV and influenza) and indirectly (e.g., dengue and chikungunya) transmitted viruses. For example, clustering analyses of the residential locations of people with dengue in Bangkok over a 5-year period showed evidence of localized transmission at distances less than 1 km (Salje, 2012; [Figure 5](#)). Analyses of data from a large population-based cohort of HIV-infected persons in Rakai District, Uganda revealed strong within-household clustering of prevalent and incident HIV cases as well clustering of prevalent cases up to 500 m (Grabowski, 2014).

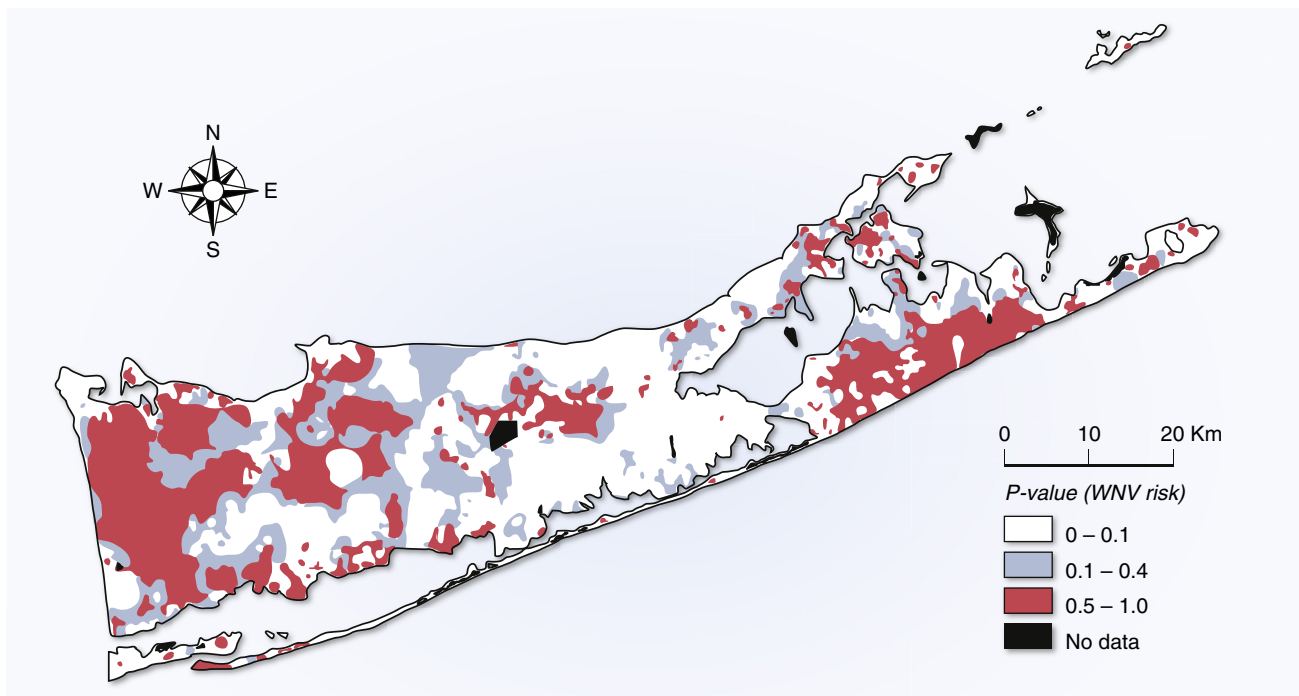
Beyond descriptive applications, mapping spatiotemporal patterns of viral infections can provide fundamental insights into transmission dynamics at different spatial scales. Traveling waves from large cities to small towns were shown to drive the spatiotemporal dynamics of measles in England and Wales (Xia, 2004). The incidence of dengue hemorrhagic fever across Thailand manifested as a traveling wave emanating from Bangkok and moving radially at a speed of 148 km/month (Cummins, 2004).

### 3.1 Remotely Sensed Data

Insight into the spatial epidemiology of viral infections and associations with environmental risk factors can be greatly enhanced when information on the spatial location of cases is combined with remotely sensed environmental data (Rodgers, 2003). The spatial coordinates of cases can be overlaid on satellite imagery to demonstrate relationships with environmental features—such as bodies of water—and formally analyzed using spatial statistical techniques. Satellite sensors that detect reflected visible or infrared radiation provide additional information on temperature, rainfall, humidity, and vegetation among other variables, which are particularly important for the transmission dynamics of vector-borne viral infections. Satellite data for epidemiologic analyses are provided by a number of sources such as: (1) earth-observing satellites with high spatial resolution (1–4 m) but low repeat frequencies such as Ikonos and Landsat satellites; (2) oceanographic and atmospheric satellites such as MODIS and ASTER with lower spatial resolution (0.25–1 km) that provide images of the Earth surface twice a day; and (3) geostationary weather satellites such as GEOS with large spatial resolution (1–8 km).



**FIGURE 5** Spatial and temporal distribution of clinical cases of dengue disease by month at Queen Sirikit Hospital between 1995 and 2000 showing spatial heterogeneity in local transmission. The border in each map represents the Bangkok provincial boundary. Redrawn from Salje, H, Lessler J, Endy TP, Curriero FC, Gibbons RV, Nisalak A, Nimmannitya S, Kalayanaroj S, Jarman RG, Thomas SJ, Burke DS, Cummings DA. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. Proc Natl Acad Sci U S A 2012; 109: 9535–8.



**FIGURE 6** Risk map for West Nile virus in Suffolk County, New York. The risk of West Nile virus is displayed as a probability ranging from low (white), medium (blue) to high (red) based on data on vector habitat, landscape, virus activity, and socioeconomic variables derived from publicly available data sets. Redrawn from Rochlin I, Turbow D, Gomez F, Ninivaggi DV, Campbell SR. Predictive mapping of human risk for West Nile virus (WNV) based on environmental and socioeconomic factors. PLoS One 2011; 6: e23280.

The statistical relationships between cases and environmental risk factors can be used to construct risk maps. Risk maps display the similarity of environmental features in unsampled locations to environmental features in locations where the disease is measured to be present or absent. Spatial analysis of the initial cases of West Nile virus infection in New York City in 1999 identified a significant spatial

cluster (Brownstein, 2002). Using models incorporating measures of vegetation cover from satellite imagery, the risk of West Nile Virus could be estimated throughout the city. A more recent risk map for West Nile virus in Suffolk County, New York, was generated with data on vector habitat, landscape, virus activity, and socioeconomic variables derived from publicly available data sets (Rochlin, 2011; Figure 6).



## 4. POPULATION MOVEMENT

Population movement plays a crucial role in the spread of viral infections. In the past, quantifying the contribution of movement to viral transmission dynamics at different spatial scales was challenging, due to limited data. As an early example, the impact of restrictions of animal movement on transmission of foot-and-mouth disease in 2001 was estimated, using detailed contact-tracing data from farms in the United Kingdom (Shirley, 2005). However, such detailed data are rarely available for patterns of human movement. Studies have attempted to model the impact of long-range human movement on the spread of viral diseases using measures such as distance between cities, commuting rates, and data on air travel. This approach has been used to explain regional and interregional spread of influenza viruses. Data on air traffic volume, distance between areas, and population sizes have been invoked to describe and predict local and regional spread of chikungunya virus in the Americas (Tatem, 2012).

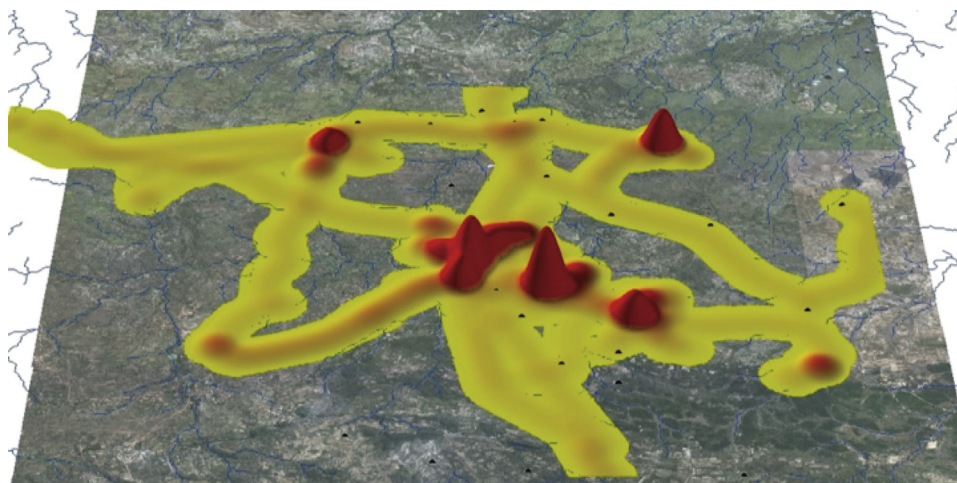
### 4.1 Movement Loggers

New technologies have greatly enhanced the capacity to study the impact of human movement on transmission dynamics of infectious diseases. Data from mobile phones and GPS loggers can be used to characterize individual movement patterns and the time spent in different locations (Figure 7). Individual movement patterns can be overlaid on risk maps to quantify movement to and from areas of high (sources) and low risk (sinks) as well as to estimate potential contact patterns. GPS data loggers generated 2.3 million GPS data points to track the fine-scale mobility patterns of 582 residents from two neighborhoods in Iquitos, Peru, to better understand the epidemiology of

viral infections (Vazquez-Prokopec, 2013). Most movement occurred within 1 km of an individual's home. However, potential contacts between individuals were irregular and temporally unstructured, with fewer than half of the tracked participants having a regular, predictable routine. The investigators explored the potential impact of these temporally unstructured daily routines and contact patterns on the simulated spread of influenza virus. The projected outbreak size was 20% larger as a consequence of these unstructured contact patterns, in comparison to scenarios modeling temporally structured contacts.

## 5. TRANSMISSION NETWORKS

In addition to identifying individual and environmental characteristics associated with temporal and spatial patterns of viral infections, transmission networks are critical drivers of the dynamics of viral infections. Analysis of transmission networks defines the host contact structure within which directly transmitted viral infections spread. Network theory and analysis are complex subjects with a long history in mathematics and sociology, but have recently been adapted by infectious disease epidemiologists. The epidemiologic study of social networks is facilitated by unique study designs, including snowball sampling or respondent-driven sampling, in which study participants are asked to recruit additional participants among their social contacts. Differing sexual contact patterns serve as an example of the importance of contact networks to the understanding of viral epidemiology. Concurrent sexual partnerships amplify the spread of HIV compared with serial monogamy. This could partially explain the dramatic differences in the prevalence of HIV in different countries. Social networks were shown to affect transmission of the 2009 H1N1 influenza virus, and were responsible for cyclical



**FIGURE 7** Movement density map of residents of Choma District, Zambia who carried a GPS logger for 1 month, overlaid on a satellite image of the area. The yellow shading shows the movement paths and the height in red indicates the time spent at location, with the taller peaks indicating more time spent in that location. By overlaying the movement density map on the satellite image as shown, investigators can explore relationships between movement, time and duration of stay, and environmental risk factors for disease. Image courtesy of Kelly Searle.

patterns of transmission between schools, communities, and households. Technological advances in quantifying contact patterns, with wearable sensors and the use of viral genetic signatures, have greatly enhanced the ability to understand complex transmission networks.

## 5.1 Contact Patterns

Self-reported contact histories and contact tracing are the traditional epidemiological methods to define transmission networks. Contact tracing has a long history in public health, particularly in the control of sexually transmitted diseases and tuberculosis, and is critical to the control of outbreaks of viral infections such as the Middle East respiratory syndrome coronavirus (MERS-CoV) and Ebola virus. To better understand the nature of human contact patterns, sensor nodes or motes have been used to characterize the frequency and duration of contacts between individuals in settings such as schools and health-care facilities. These technologies offer opportunities to validate and complement data collected using questionnaires and contact diaries.

As an example, investigators used wireless sensor network technology to obtain data on social contacts within 3 m for 788 high school students in the United States, enabling construction of the social network within which a respiratory pathogen could be transmitted (Salathe, 2010). The data revealed a high-density network with typical small-world properties, in which a small number of steps link any two individuals. Computer simulations of the spread of an influenza-like virus on the weighted contact graph were in good agreement with absentee data collected during the influenza season. Analysis of targeted immunization strategies suggested that contact network data can be employed to design targeted vaccination strategies that are significantly more effective than random vaccination.

## 6. MOLECULAR EPIDEMIOLOGY

Advances in nucleic acid sequencing and bioinformatics have led to major advances in viral epidemiology. Population (Sanger) sequencing has been the standard method for DNA sequencing but is increasingly replaced by deep sequencing in which variants within a viral swarm are distinguished. Sequencing allows for the detection of single nucleotide polymorphisms (SNPs) and nucleotide insertions or deletions (“indels”), analysis of synonymous and nonsynonymous mutations, and phylogenetic analysis (see chapter on virus evolution).

Sequencing techniques can be applied to both viral and host genomes. SNPs may be associated with changes in viral pathogenesis, virulence, or drug resistance. Molecular techniques applied to pathogens also have been fundamental to the study of the animal origins of many viral infections including HIV and MERS. Phylogeographic

approaches were used to trace the origins of the HIV pandemic to spillover events in central Africa (Sharp, 2010). More recently, sequence data were used to track the animal reservoirs of MERS-CoV associated with the 2014 outbreaks (Haagmans, 2014), and to compare the Ebola virus strain circulating in the 2014 West Africa outbreak to strains from prior outbreaks (Gire, 2014).

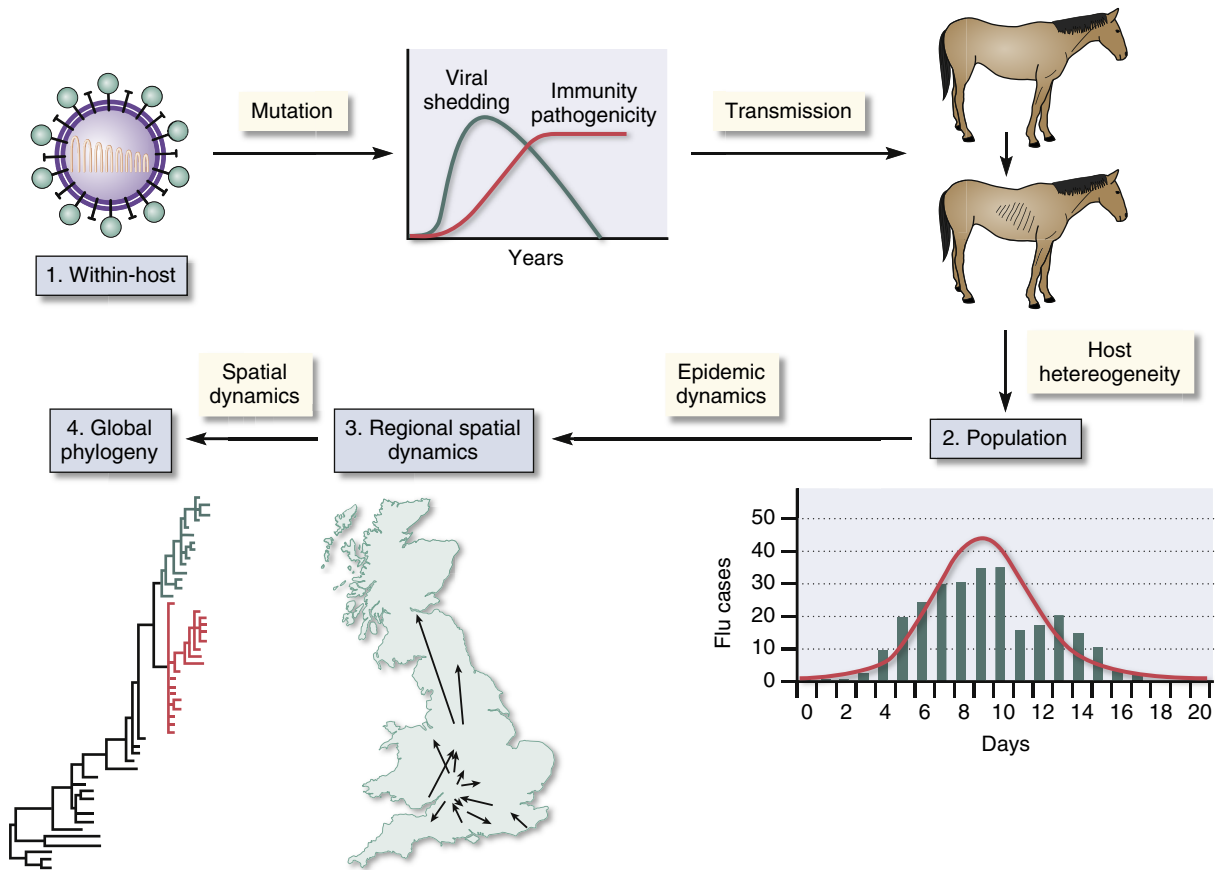
Epidemiologic studies that probe host genomes can be either candidate gene studies or genome-wide association studies. The goal of these studies is to link specific changes with an increased risk of infection or disease. As an example, a small subset of individuals who failed to acquire HIV infection despite exposure, prompted studies to determine how these individuals differed from those who acquired infection. A 32-base-pair deletion in the human CCR5 gene, now referred to as CCR5-delta 32, accounted for the resistance of these subjects. Individuals who are CCR5-delta 32 homozygotes are protected against HIV infection by CCR5-tropic HIV strains, while heterozygotes have decreased disease severity.

## 7. PHYLODYNAMICS

Infectious disease epidemiologists are increasingly linking evolutionary, immunologic, and epidemiological processes, a field referred to as phylodynamics (Grenfell, 2004; Voltz, 2013). Because of the high mutation rates of viral pathogens, particularly RNA viruses, evolutionary and epidemiological processes take place on a similar timescale (see chapter on virus evolution). According to this framework, phylodynamic processes that determine the degree of viral diversity are a function of host immune selective pressures and epidemiological patterns of transmission (Figure 8).

Intrahost phylodynamic processes begin with molecular characteristics of the virus as well as the host’s permissiveness and response to infection. For example, a single amino acid substitution in Epstein–Barr virus was shown to disrupt antigen presentation by specific human leukocyte antigen polymorphisms (Liu, 2014). This resulted in decreased T-cell receptor recognition and successful viral immune escape. The virus must also induce an “optimal” host immune response to maximize transmission to new hosts. If the virus induces a strong, proinflammatory immune response not balanced by the appropriate anti-inflammatory responses, the host may succumb to the overabundance of inflammation and cannot propagate viral transmission. Alternatively, a virus that fails to stimulate an immune response may also replicate uninhibited, overwhelming, and killing the host prior to transmission. Selective pressures maximize replication while sustaining transmission between hosts.

Interhost dynamics are affected by several factors including evolutionary pressures, timescales of infection, viral latent periods, and host population structures. Typically, only a small number of virions are transmitted



**FIGURE 8** Phylodynamics links evolutionary, immunologic, and epidemiologic processes to explain viral diversity, as shown here for equine influenza virus. For viral evolution, these processes take place on a similar timescale. Within host mutations (1) result from an interplay between optimization of viral shedding, immunologic selective pressures and host pathogenicity. Transmission bottlenecks and host heterogeneity (2) further determine the population genetic structure of the virus, which in turn influences and is determined by the epidemic dynamics. Larger scale spatial dynamics at local, regional, and global levels (3), determines the global viral phylogenetic structure. *Revised and redrawn from Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 2004; 303:327–32.*

between hosts, creating a genetic bottleneck that limits viral diversity. A virus that mutates to cause highly pathogenic disease but is not transmitted cannot propagate its pathogenicity. Cross-immunity between viral strains also precludes the replication of particular viral lineages. Influenza vaccine strains require annual changes due to new circulating influenza strains that have escaped immune pressures through high mutation rates and gene re-assortment. The strong selection pressure of cross-immunity is reflected in the short branch lengths in a phylogenetic tree of influenza viruses isolated from infected individuals. Thus, the selection of influenza strains for future vaccines is partly determined by cross-immunity to prior circulating strains, because influenza viral strains that circulated in the past may elicit immune protection against currently circulating strains.

At the population level, phylodynamic methods have been used to estimate  $R_0$  for HIV and hepatitis C virus, for which reporting and surveillance data are often incomplete (Volz, 2013). Phylodynamic and phylogeographic

models also have been useful in reconstructing the spatial spread of viruses to reveal hidden patterns of transmission. For example, epidemiological and molecular studies of influenza virus transmission were compared at different spatial scales to highlight the similarities and differences between these data sources (Viboud, 2013). The findings were broadly consistent with large-scale studies of interregional or inter-hemispheric spread in temperate regions with multiple viral introductions resulting in epidemics followed by interepidemic periods driven by seasonal bottlenecks. However, at smaller spatial scales—such as a country or community—epidemiological studies revealed spatially structured diffusion patterns that were not identified in molecular studies. Phylogenetic analyses of gag and env genes were used to assess the spatial dynamics of HIV transmission in rural Rakai District, Uganda, using data from a cohort of 14,594 individuals residing in 46 communities (Grabowski, 2014). Of the 95 phylogenetic clusters identified, almost half comprised two individuals sharing a household. Among

the remaining clusters, almost three-quarters involved individuals living in different communities, suggesting transmission chains frequently extend beyond local communities in rural Uganda.

The timescale of infection is also important for viral diversity and transmission dynamics. Some viruses are capable of initiating an acute infection that is cleared within days, while other viral infections are chronic and persist for a lifetime. The duration of infection impacts how quickly a virus must be transmitted and has implications for the infectious period and the potential to be transmitted to new hosts. Viruses with long latent periods create interhost phylogenetic trees with longer branch lengths. The long duration between infection and transmission permits accumulation of viral changes through many rounds of viral replication before transmission to the next host. Examples include hepatitis B virus, hepatitis C virus, and human immunodeficiency virus.

## 8. MATHEMATICAL MODELS OF TRANSMISSION DYNAMICS

Availability of computational resources allows widespread use and development of classic approaches to the mathematical modeling of viral transmission dynamics, such as compartmental, metapopulation and network models, to address epidemiologic questions (see chapter on mathematical methods). These models have been used extensively in the study of viral dynamics and to explore the potential impact of control interventions. New sources of high-resolution spatial, temporal, and genetic data create opportunities for models that integrate these data with traditional epidemiological data. Such analyses improve estimates of key transmission parameters and understanding of the mechanisms driving virus spread.

### 8.1 Agent-Based Models

Agent-based models (also known as individual-based models) can now be run using desktop computers, and offer advantages over more traditional mathematical models. Because each unit in a population is modeled explicitly in space and time and assigned specific attributes, agent-based models can reproduce the heterogeneity and complexity observed in the real world. More traditional compartmental, differential equation models often require simplifying assumptions that limit applicability. Agent-based models have been used to study the spread of viruses in populations as well as the evolution of viruses within and across populations. While agent-based models are intuitive and easy to formulate, these models are often difficult to construct due to the large number of parameters necessary to describe the behavior and interaction between individual units. Commercial frameworks that offer large

computational power and intuitive user interphases have also become increasingly available. The Global Epidemic and Mobility Model (GLEAM) on the gleamviz platform ([www.gleamviz.com](http://www.gleamviz.com)), for example, contains extensive data on populations and human mobility, and allows stochastic simulation of the global spread of infectious diseases using user-defined transmission models.

## 9. SYNTHESIS

Our understanding of the epidemiology of viral infections is being revolutionized by the integration of traditional epidemiological information with novel sources of data.

- Data streams from the Internet are promising sources to enhance traditional surveillance but have yet to be fully validated.
- Molecular data on viral genomic sequences provide unprecedented opportunities to characterize viral transmission pathways.
- Phylodynamic and phylogeographic models have been used to estimate  $R_0$ , and characterize the spatial spread of viruses.
- Network analysis reveals hidden patterns of transmission between population subgroups that are not easy to capture with traditional epidemiological methods.
- Novel analytical and computational resources are playing a key role in integrating information from multiple large data banks.

These more comprehensive methods improve our ability to estimate the impact of infection control measures. The combination of traditional and evolving methodologies is closing the gap between epidemiological studies and viral pathogenesis. These developments have laid the foundation for exciting future research that will complement other approaches to the pathogenesis of viral diseases.

With these evolving technologies in mind, it is timely to ask: Is the world able to control viral diseases more effectively? It is a mixed score card. On the one hand, smallpox has been eradicated and we are on the verge of elimination of wild polioviruses. Furthermore, deaths of children under the age of 5 years (which are mainly due to viral and other infectious diseases) have decreased by almost 50% in the last few decades. On the other hand, the AIDS pandemic continues to rage in low-income countries, with only a slight reduction in the annual incidence of new infections. The United States has not done any better in reducing HIV incidence which has been unchanged for at least 20 years. The 2014–15 Ebola pandemic in West Africa reflects the limited capacity for dealing with new and emerging viral diseases on a global basis. In conclusion, epidemiological science continues to advance with evolving new technologies, but their application to public health remains a future challenge and opportunity.



## FURTHER READING

### Review, Articles, and Books

- Chunara R, Freifeld CC, Brownstein JS. New technologies for reporting real-time emergent infections. *Parasitology* 2012;139:1843–51.
- Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 2009;49:1557–64.
- Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 2004;303:327–32.
- Rogers DJ, Randolph SE. Studying the global distribution of infectious diseases using GIS and RS. *Nat Rev Microbiol* 2003;1:231–7.
- Tufte ER. *The Visual Display of Quantitative Information*. 2<sup>nd</sup> edition. 2001. Graphics Press, Cheshire, Connecticut.

### Original Research Reports

- Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D. Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne Zoonotic Dis* 2002;2:157–64.
- Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;5:e14118.
- Cummings D, Irizarry RA, Huang NE, Endy TP. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature* 2004;427:344–7.
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;457:1012–4.
- Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang PP, Nekoui M, Colubri A, Coomber MR, Fonnies M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Kounuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JL, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheffelin JS, Lander ES, Happi C, Gevaio SM, Gnirke A, Rambaut A, Garry RF, Khan SH, Sabeti PC. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 2014;345:1369–72.
- Grabowski MK, Lessler J, Redd AD, Kagaayi J, Laeyendecker O, Ndy-anabo A, Nelson MI, Cummings DA, Bwanika JB, Mueller AC, Reynolds SJ, Munshaw S, Ray SC, Lutalo T, Manucci J, Tobian AA, Chang LW, Beyrer C, Jennings JM, Nalugoda F, Serwadda D, Wawer MJ, Quinn TC, Gray RH; Rakai Health Sciences Program. The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: evidence from spatial clustering, phylogenetics, and egocentric transmission models. *PLoS Medicine* 2014;11:e1001610.
- Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, Godeke GJ, Jonges M, Farag E, Diab A, Ghobashy H, Alhajri F, Al-Thani M, Al-Marri SA, Al Romaihi HE, Al Khal A, Bermingham A, Osterhaus AD, AlHajri MM, Koopmans MP. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect Dis* 2014;14:140–5.
- Lazer, DM, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science* 2014;343:1203–5.
- McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput Biol* 2014;10:e1003581.
- Liu YC, Chen Z, Neller MA, Miles JJ, Purcell AW, McCluskey J, Burrows SR, Rossjohn J, Gras S. A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism. *J Biol Chem* 2014;289:16688–98.
- Paz-Soldan VA, Stoddard ST, Vazquez-Prokopec G, Morrison AC, Elder JP, Kitron U, Kochel TJ, Scott TW. Assessing and maximizing the acceptability of global positioning system device use for studying the role of human movement in dengue virus transmission in Iquitos, Peru. *Am J Trop Med Hyg* 2010;82:723–30.
- Rochlin I, Turbow D, Gomez F, Ninivaggi DV, Campbell SR. Predictive mapping of human risk for West Nile virus (WNV) based on environmental and socioeconomic factors. *PLoS ONE* 2011;6:e23280.
- Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for infectious disease transmission. *Proc Natl Acad Sci U S A* 2010;107:22020–5.
- Salje, H, Lessler J, Endy TP, Curriero FC, Gibbons RV, Nisalak A, Nimmannitya S, Kalayanaroop S, Jarman RG, Thomas SJ, Burke DS, Cummings DA. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. *Proc Natl Acad Sci U S A* 2012;109:9535–8.
- Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Philos Trans R Soc Lond B Biol Sci* 2010;365:2487–94.
- Shirley MD1, Rushton SP. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. *Epidemiol Infect* 2005;133:1023–32.
- Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS One* 2011;6:e19467.
- Tatem AJ, Huang Z, Das A, Qi Q, Roth J, Qiu Y. Air travel and vector-borne disease movement. *Parasitology*. 2012;139:1816–30.
- Vazquez-Prokopec GM, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, Kochel TJ, Scott TW, Kitron U. Usefulness of commercially available GPS data-loggers for tracking human movement and exposure to dengue virus. *Int J Health Geogr* 2009;8:68.
- Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, Ramirez-Paredes J, Halsey ES, Kochel TJ, Scott TW, Kitron U. Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLoS One* 2013;8:e58802.
- Viboud C, Nelson MI, Tan Y, Holmes EC. Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Philos Trans R Soc Lond B Biol Sci* 2013;368:20120199.
- Volz EM, Koelle K, Bedford T. Viral phylodynamics. *PLoS Comput Biol* 2013;9:e1002947.
- Xia Y, Björnstad ON, Grenfell BT. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am Nat* 2004;164:267–81.