Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017

Joacim Rocklöv, Yesim Tozan, Aditya Ramadona, Maquines O. Sewe, Bertrand Sudre, Jon Garrido, Chiara Bellegarde de Saint Lary, Wolfgang Lohr, Jan C. Semenza

With regard to fully harvesting the potential of big data, public health lags behind other fields. To determine this potential, we applied big data (air passenger volume from international areas with active chikungunya transmission, Twitter data, and vectorial capacity estimates of *Aedes albopictus* mosquitoes) to the 2017 chikungunya outbreaks in Europe to assess the risks for virus transmission, virus importation, and short-range dispersion from the outbreak foci. We found that indicators based on voluminous and velocious data can help identify virus dispersion from outbreak foci and that vector abundance and vectorial capacity estimates can provide information on local climate suitability for mosquitoborne outbreaks. In contrast, more established indicators based on Wikipedia and Google Trends search strings were less timely. We found that a combination of novel and disparate datasets can be used in real time to prevent and control emerging and reemerging infectious diseases.

Many sectors of society have taken full advantage of new opportunities provided by big data, but public health has not (*1*). Although electronic health records have long been used in surveillance, novel applications of big data are rare. Internet search query data from Google or Wikipedia have been applied to anticipate influenza epidemics but are hampered by several limitations, including specificity and granularity (*2–4*). More recently, crowdsourcing of symptoms through emails, text messages, or tweets has been explored, and outbreaks have been tracked by scanning high-volume surveillance systems (*5*,*6*). However, when it comes to fully harvesting the potential of big data, public health still lags behind other fields. Using chikungunya as a case study, we illustrate how big data can help tackle emerging infectious diseases through prevention, detection, and response.

Author affiliations: Umeå University, Umeå, Sweden (J. Rocklöv, A. Ramadona, M.O. Sewe, W. Lohr); New York University, New York, New York, USA (Y. Tozan); European Centre for Disease Prevention and Control, Stockholm, Sweden (B. Sudre, J. Garrido, C.B. de Saint Lary, J.C. Semenza)

DOI: https://doi.org/10.3201/eid2506.180138

A key driver of the emergence and spread of vectorborne diseases is human mobility (*7–10*), yet little is known about the epidemiologic consequences of mobility patterns at different spatial scales within the context of vectorborne diseases. A main obstacle to studying the complex interactions between human hosts, pathogens, and vectors has been the limited availability of spatiotemporal datasets for analyzing human mobility patterns. Prior research relied on low-resolution mobile phone records, such as call and messaging logs from mobile phone networks (*11–13*), for which biases were notable (*14*,*15*). Furthermore, use of mobile phone data for tracking human mobility is likely to be fraught with privacy concerns and data access restrictions (*15*).

Recently, social media has emerged as an alternative source of real-time, high-resolution geospatial data on a large scale (*1*,*15*). Use of this unique aspect of publicly available social media data to study the human dimensions of the introduction and spread of emerging infectious diseases has not been explored to its fullest extent. In areas where risk for virus importation and onward transmission is heightened, such knowledge can inform outbreak preparedness and response planning by pinpointing receptive areas where proactive countermeasures should be implemented in a timely fashion (*16*,*17*).

The impediments to using big data in public health are not only the size of the databases but also the complexity of their processing. The challenges include 3 main dimensions: volume, velocity, and variety (*18–20*). Volume calls for statistical sampling; velocity, for instant access to near real-time transaction data; and variety, for management of nonaligned data structures. We illustrate how big data can be used to monitor the introduction and spread of the 2017 chikungunya outbreak in Europe by tackling these challenges (*18–20*).

To assess risk for virus importation from international areas with active chikungunya transmission, we extracted air passenger volume from large-scale aviation data. To quantify the risk for short-range dispersion (defined as the potential for onward transmission and spread of chikungunya virus from

the initial outbreak foci to other areas during transmission season), we used a mining algorithm to process quasi–real-time, geolocated Twitter activity data and computed mobility patterns of users. We have previously shown that mobility data from Twitter users is predictive of disease spread (*21*). We then estimated the seasonal vectorial capacity of *Aedes albopictus* mosquitoes to transmit chikungunya virus and linked it with human mobility patterns. We further complemented these data with Internet and information search activities related to chikungunya infection, vectors, and clinical signs and symptoms collected from Wikipedia and Google Trends. Last, we estimated the empirical basic reproduction number (R_0) from the outbreaks and compared these numbers with our model predictions of epidemic potential based on climate conditions. More detail on our methods is provided in Appendix 1 (https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App1.pdf).

Climate Suitability: Vectorial Capacity

The vectorial capacity of *Ae. albopictus* mosquitoes to transmit chikungunya virus in areas of Europe where the vector is established (*17*), such as the outbreak zones in France and Italy, was estimated to be high in July and August but lower in September and October. Estimates of suitability were low in October for most areas, except those in southern Italy and Greece and southeastern Spain (Figure 1). Overall, warmer than average temperatures led to a substantial increase in vectorial capacity during the study period (June–October 2017) (Appendix 2 Figure 1, https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App2.pdf). Using empirical data from the outbreaks in Italy (*22*), we estimated R_0 to be 2.28 (95% CI 2.01–2.59) for the Anzio region, 3.54 (95% CI 2.62–4.97) for the Rome region, and 3.11 (95% CI 2.16–4.79) for the Calabria region (Figure 2).

Long-Range Importation: Air Passenger Volume

On average, ≈50,000 air passenger-journeys (1 passenger flight, including all legs of travel) were taken each month from areas with active chikungunya transmission worldwide to the outbreak zones (Figure 3). Specifically, in August, 56,300 passengers from outbreak zones were estimated to arrive in Rome, 6,484 in Nice, and 5,629 in Marseille. The passenger-journey volume into Europe when the outbreak started in June is shown in Appendix 2 Figure 2. The countries with the highest number of departing passengers in August were Thailand (352,332 passengers), Brazil (255,439 passengers), and India (301,298 passengers). According to molecular epidemiology, the genome sequence of a chikungunya virus isolate from the Lazio region of Italy revealed the East/Central/South African lineage, Indian Ocean sublineage, which is similar to that of recent sequences from Pakistan and India (*23*). We also extracted air passenger-journey data for flights from the outbreak zones in southeastern France and central Italy to other areas in Europe (Figure 3). The top 5 destinations with the highest volume were the larger metropolitan areas of Europe, most of which were outside

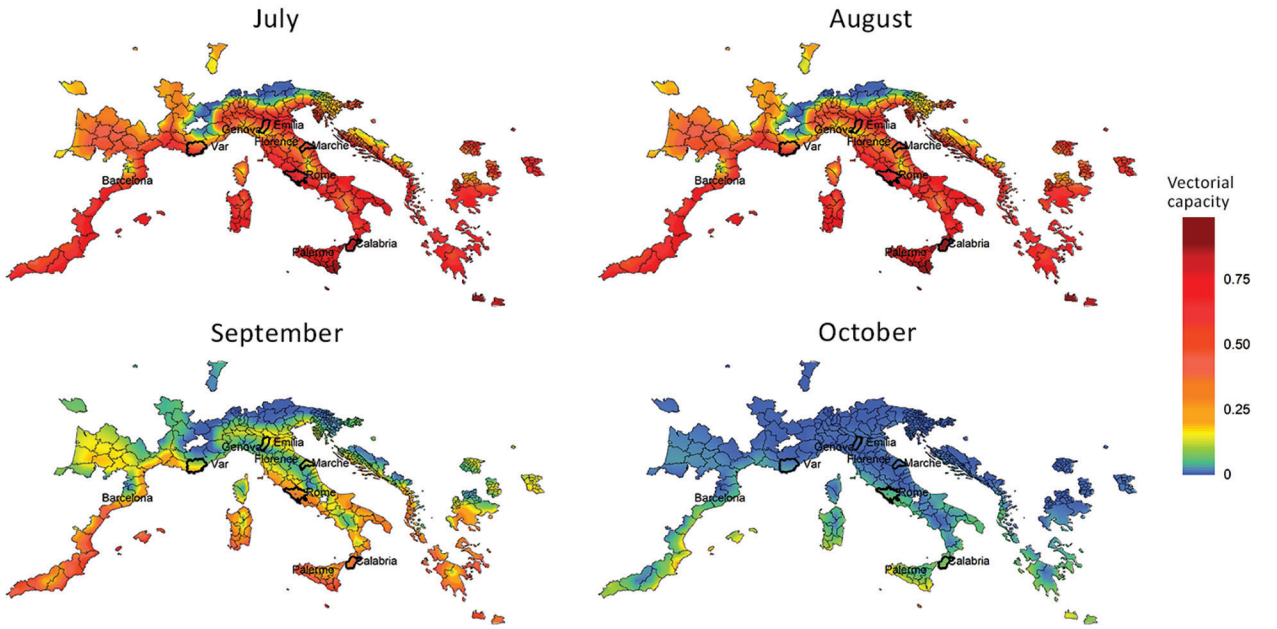


Figure 1. Vectorial capacity estimates based on average temperature conditions in Europe with stable populations of *Aedes albopictus* mosquitoes around chikungunya outbreak zones, Italy and France, July–October 2017. Heavy outlines indicate the outbreak areas. The vectorial capacity translates to an average basic reproduction number in the range of 2–3 in Anzio and Rome and in the range of 3–4 in Calabria during the months of July and August for an infectious period of 4 days.

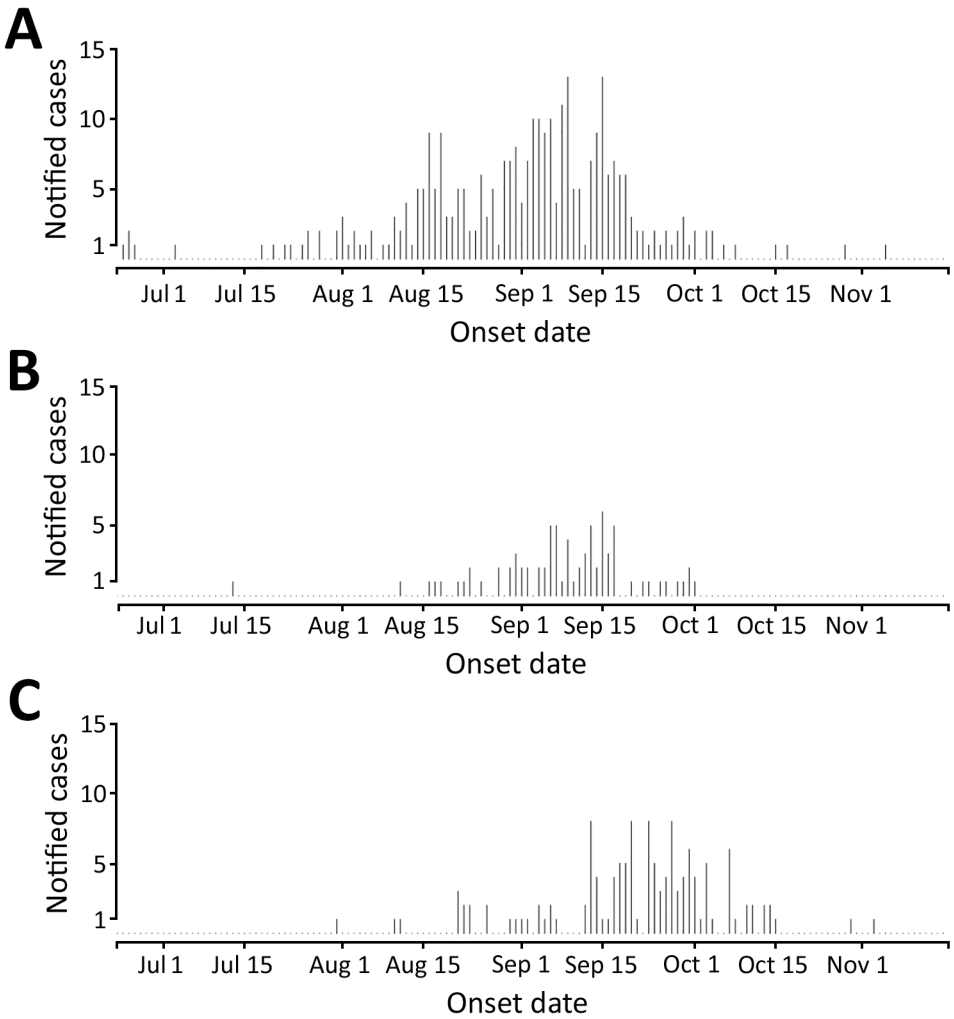


Figure 2. Notified chikungunya cases in the Anzio (A), Rome (B), and Calabria (C) regions and basic reproduction number (R_0) estimates of outbreaks, June–October 2017, Italy.

the boundaries of areas where the vector is known to be present (Figure 4). However, high flight connectivity was observed from the outbreak zones to Barcelona (Spain) and Catania and Palermo (Italy).

Short-Range Dispersion: Geocoded Tweets

The spatiotemporal analysis of geocoded Twitter data showed strong human mobility from Lazio (Figure 4) and the Var department in France (Appendix 2 Figure 3) toward several larger cities where *Ae. albopictus* mosquitoes are present. The top 10 estimates of mobility out of the 2 outbreak zones of Var and Lazio showed the strongest pattern for potential dispersion of chikungunya virus not only into the areas geographically close to the outbreak zones but also to several relatively large cities in Italy, France, and Spain (Table). The monthly mobility patterns during the study period varied between months; for example, the vacation month of August showed a stronger mobility pattern out of Var to areas not in direct connectivity, most notably to Rome (Appendix 2 Figure 4). When we contrasted the mobility proximities between the 2 outbreak zones, we observed the highest proximities within countries (Figure 4; Appendix 2 Figure 3). Although the Var and Lazio outbreak zones experienced high mobility proximity to Barcelona, Lazio was also highly connected to southern Italy (e.g., Catania and Palermo), in close proximity to the chikungunya outbreak in the Calabria area, which was also observed in the International Air Transport Association (IATA) flight passenger data (Figures 3, 4). In Italy, cases were first notified in Anzio at the end of June, followed by notifications in Rome later in July, and in Calabria in early August in order of temporal appearance (Figure 2). In our mobility analysis, we identified the mobility links to all outbreak regions (Figure 4), with the exception of the Emilia-Romagna region, although the region neighboring Emilia-Romagna was positive in our analysis. The mobility patterns correlated more strongly to the outbreak regions in July and August.

A closer look at the Lazio outbreak zone in Italy revealed strong connectivity between Anzio (where the first

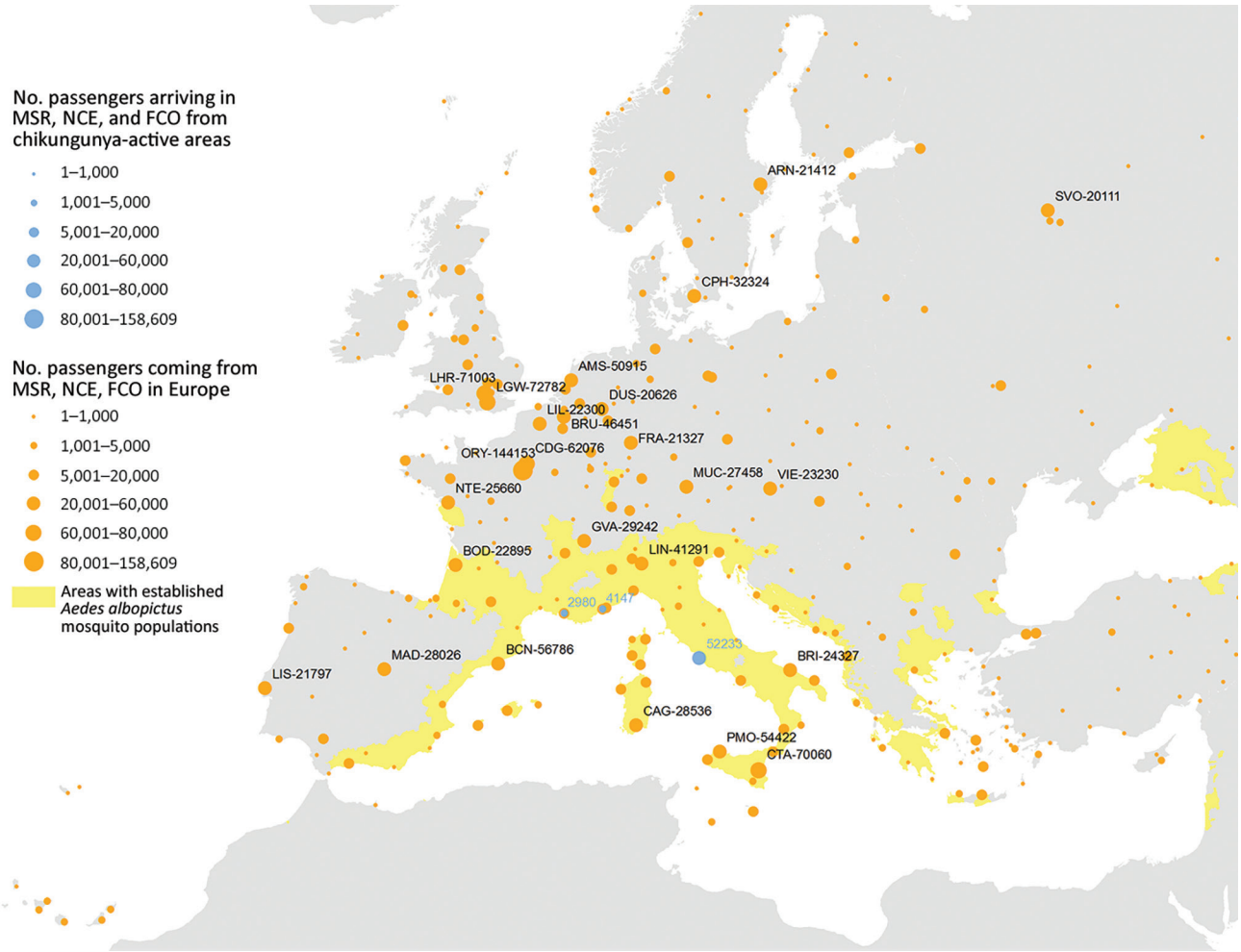


Figure 3. Incoming passengers from chikungunya active transmission areas and outgoing passengers to other airports in Europe from Rome (FCO), Marseille (MRS), and Nice (NCE) airports, August 2017. The stable vector presence area is highlighted in yellow.

cases in Italy were confirmed) and Rome (where a higher number of cases were notified) (Figure 5). We compiled the top 10 mobility proximity areas from the outbreak zones of Anzio and Rome in August and September (Table). Although the highest mobility proximity from Anzio was to Rome in August and September, the mobility proximity from Rome to Anzio was also found among the top 10 destinations. Overall, Rome had higher connectivity to many more areas than Anzio.

Synergistic Effects: Human Mobility and Transmission Suitability

We derived risk maps for autochthonous chikungunya transmission by combining the vectorial capacity and mobility proximity estimates for the Lazio region in Italy and Var department in France for August–October 2017 (Appendix 2 Figure 4). The areas at risk because of the outbreak in Var were identified to be located along the French and northern Spanish Mediterranean coastlines, Mallorca, and Rome in August (Appendix 2 Figure 4); the risk regions for the Lazio outbreak in August included large parts of Italy as well as areas in France, Spain, and Greece (Figure 6). In general, the size of the area at risk contracted in September and more so in October because of less favorable climate conditions, except in the most southern region of Italy (Figure 6), such as the Calabria region, where the outbreak also empirically continued longer in the fall (Figure 2).

In the Lazio region, an analysis of the combination of vectorial capacity (Appendix 2 Figure 5) and mobility proximity revealed a higher transmission potential in August (Appendix 2 Figure 6), with implications for targeting surveillance and outbreak control activities to this region. The largest area of risk for spread from Anzio was Rome, but the risk for spread from Rome was more widespread in the region (Appendix 2 Figures 6, 7). The areas at risk for spread in the Lazio region differed during August compared with September and October.

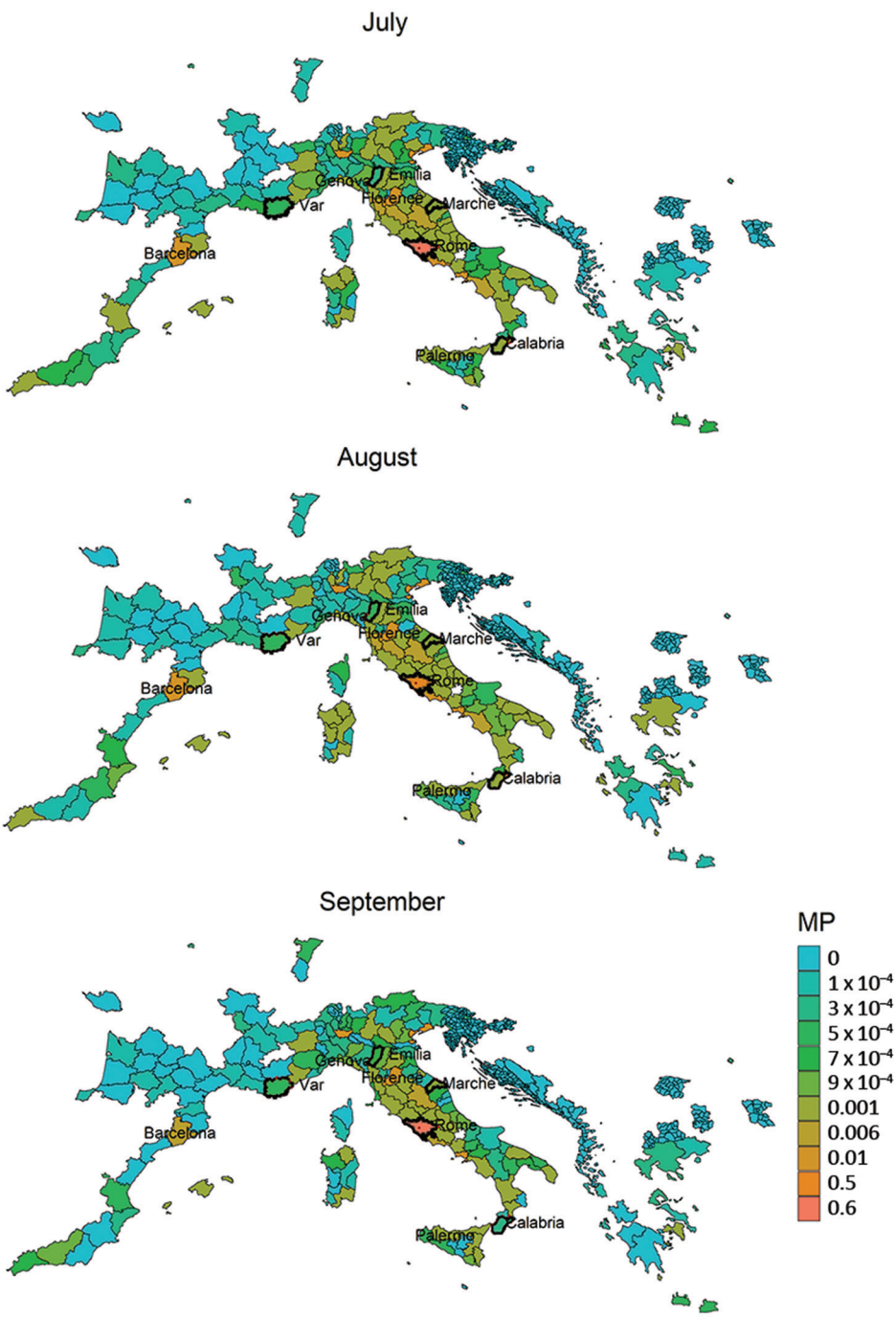


Figure 4. MP estimates from the Lazio region, Italy, to areas in Europe with stable populations of *Aedes albopictus* mosquitoes, July–September 2017. Heavy outlines indicate the chikungunya outbreak areas. MP, mobility proximity.

Wikipedia and Google Trend Indicators

For the outbreaks in Italy, several pathogen and vector-related Wikipedia and Google Trend search pattern anomalies are illustrated (Appendix 2 Figure 8). The peaks in these abnormalities coincided with the peak of the outbreak and therefore are not useful for early detection and response activities. Detailed information about Wikipedia and Google Trend indicators are provided in Appendix 3 (https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App3.pdf).

Big Data and Emerging Infectious Diseases

In light of the arrival and explosive expansion of chikungunya in the Americas in 2013 through *Ae. aegypti* moquitoes (*24*), big data offer the opportunity to monitor the introduction and spread of chikungunya in Europe. An

Table. Top 10 areas where mobility proximity to the 2 chikungunya outbreak zones was highest, Europe, August 2017

	Southern Europe		Lazio region	
Rank	From Var department, France	From Lazio region, Italy	From Anzio	From Rome
1	Alpes-Maritimes	Florence	Roma	Vatican
2	Bouches-du-Rhône	Milano	Nettuno	Fiumicino
3	Torino	Napoli	Sabaudia	Sabaudia
4	Paris	Venezia	Ardea	Civitavecchia
5	Alpes-de-Haute-Provence	Paris	Civitavecchia	Santa Marinella
6	Rhone	Barcelona	Pomezia	Tivoli
7	Hérault	Perugia	Aprilia	Anzio
8	Vaucluse	Latina	Cisterna Di Latina	Ladispoli
9	Barcelona	Siena	Fondi	Pomezia
10	Baeleares	Salerno	Amatrice	Valmontone

outbreak can be divided, broadly speaking, into 2 distinct phases. The first phase is importation of the virus via a viremic person into a virus-naive population. For this phase, we used big data (volume) to estimate air passenger-journeys from areas with active chikungunya transmission as a measure of the force of introduction of the virus into the outbreak zones in Europe. To identify areas with onward transmission risk, we also considered the volume of air passengers leaving these outbreak zones. For the second phase, the establishment of autochthonous transmission in Europe is a function of virus importation, population density, vector activity, climate conditions, exposure patterns, and several other factors that are more difficult to quantify (*17*). Our study addressed some of these epidemiologic challenges by using big data. Rather than a Twitter content analysis, which has been performed for several outbreaks (*25–28*), we used near–real-time geocoded Twitter data (velocity) to quantify human mobility patterns and disentangled connectivity between populations. Mobility estimates also reflect population density and indirectly take into account exposure patterns because such populations on the move are occasionally susceptible to exposure and are also a source of exposure. The ecology of the virus and the human-vector transmission cycle were captured by vectorial capacity (variety), which quantified transmission risk on the basis of climate conditions. Thus, we were able to quantify the trajectory of an arbovirus outbreak by dissecting and better understanding its phases.

Our analysis of big data revealed distinct mobility patterns between the outbreak zones in France and Italy, between Rome and Anzio, and between Rome and most of the local outbreak clusters in Italy. However, the potential effects of these mobility patterns on local spread need to be confirmed epidemiologically by phylogenetic analyses. Although the sensitivity of our risk maps based on mobility and climate data to identify areas at risk for virus spread was good, the specificity needs to be further improved, for example, by including local contextual factors such as land use and vector activity. Wikipedia page hits and Google Trends have been proposed as resources for disease surveillance and outbreak detection. However, our analysis demonstrates that these sources seemed to mainly indicate public awareness of the chikungunya outbreaks as they peaked. For such reasons, they seem to be of little use for early response.

The combination of short-distance air passenger-journeys (within Europe, as opposed to overseas) and geocoded Twitter data lends itself to cross-validation. We found that the 2 approaches consistently identified several cities with established vector populations at a heightened risk for virus importation, reflecting the potential for spread between countries and cities in Europe. Some of these regions had previously encountered autochthonous transmission (*29*).

The R_0 estimates, which were derived by using epidemiologic data, were in accordance with the vectorial capacity predictions for the outbreak zones based on local climate conditions. Based on the vectorial capacity, R_0 can be derived by multiplication with the infectious period. For chikungunya, an infectious period of 3–7 days was reported (*30*). The vectorial capacity of ≈0.7 would give rise to an R_0 of ≈2–3. This range is within that which we observed in the Rome and Anzio regions in July and August, but the vectorial capacity was estimated to be higher (≈0.8) in the Calabria region, translating into an R_0 of just over 3–4, which is in agreement with the epidemiologic analysis of the outbreak data (Figure 2).

Although our mobility analysis showed that the local mobility from Var was considerable, no autochthonous chikungunya cases were reported from other identified risk regions along the Mediterranean coast of France and in northern Spain. However, the vectorial capacity of *Ae. albopictus* mosquitoes to transmit the virus is lower in Var than in Lazio, which may explain this discrepancy. Previous studies assessing the risk for local outbreaks after outbreaks outside of Europe found that inbound flight traveler frequencies correlated strikingly well with local reports of virus importation frequencies into Europe (*9*). However, most of these studies evaluated these risks independently and did not attempt to estimate the combined risk for virus importation and climate suitability (*31*,*32*). Moreover, they did not assess local dispersion patterns from airports or outbreak areas. We analyzed big data for long- and short-distance mobility. A major strength of this big data approach is the near real-time availability of mobility patterns based

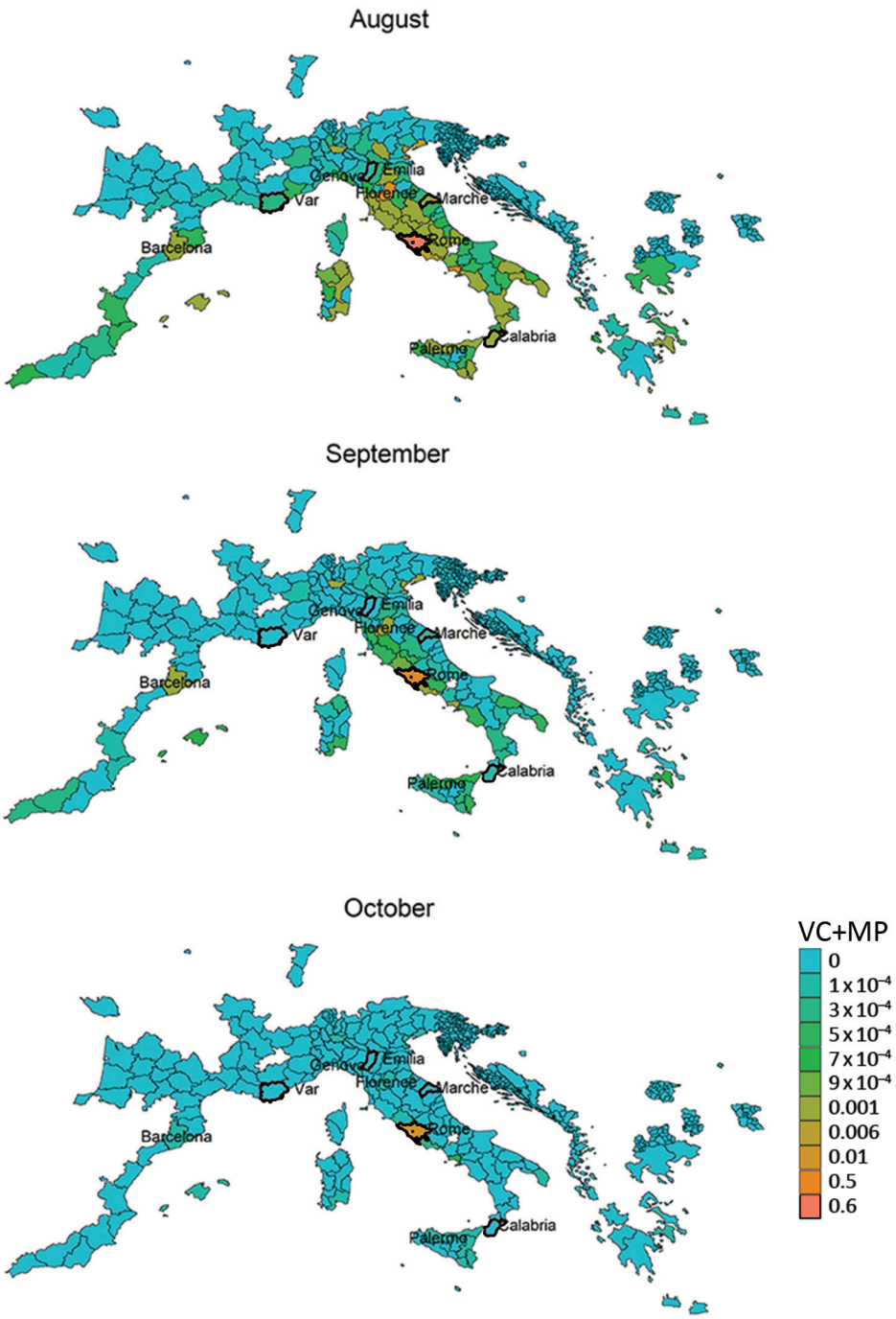


Figure 5. Estimated areas of risk for chikungunya spread from the outbreak areas of Anzio and Rome in the Lazio region, Italy, based on combined VC and MP estimates, August–October 2017. Heavy outlines indicate the outbreak areas. MP, mobility proximity; VC, vectorial capacity.

on social media, which are timelier and more accessible and less costly than air passenger data available from commercial providers, such as the IATA. This approach can identify areas of heightened mobility that are potentially at risk for onward transmission, as we have shown in this analysis. Geocoded Twitter data can be a good proxy for human mobility (*15*), but prior research did not explore how such data can be a timely resource for preparedness and response to infectious disease outbreaks.

Similar to others who have used IATA and Twitter data in their studies, we found these novel data sources to be reliable and useful. However, we note that Twitter data can potentially be biased because Twitter users may represent a select population whose mobility patterns differ from those of the general population; more specifically, they represent a population of Twitter users who have allowed Twitter to follow their geolocations. Future studies need to validate the use of social media data in such applications.

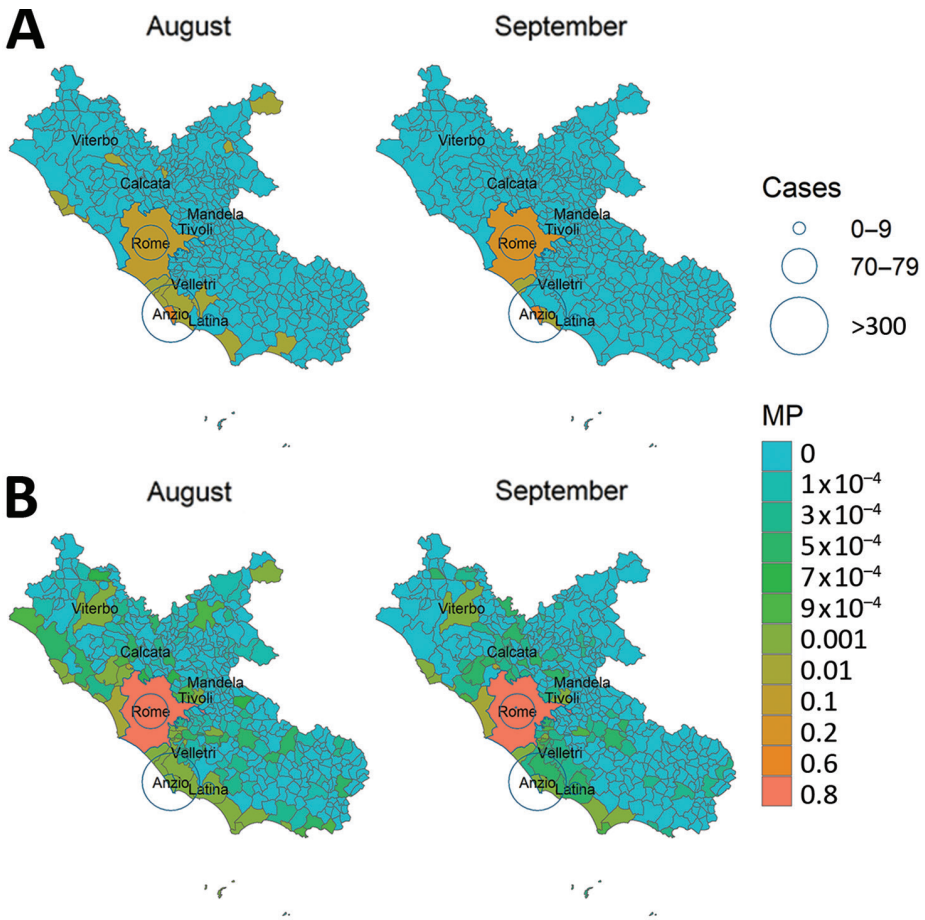


Figure 6. Estimated areas of risk for chikungunya spread from the outbreak areas in Lazio region, Italy, based on MP estimates, August–September 2017. A) Anzio; B) Rome. Circles indicate number of reported cases. MP, mobility proximity.

These methods are an improvement over mobile telephone tracking data because they do not rely on a single provider network and are a less costly data source to acquire.

Seasonal weather forecasts may have provided better input into the assessment of vectorial capacity, specifically for the fall of 2017. Moreover, autochthonous transmission risk may also be related to local proliferation of vectors and local environmental, social, and behavioral characteristics, such as awareness about the symptoms of chikungunya (Appendix 3). Such factors have been found to be associated with the local transmission risk for dengue (*33*). Last, because of the paucity and underreporting of chikungunya cases, we may have potentially underestimated the passenger volume from active transmission areas in Africa.

This study illustrates the potential value of using big data (*18–20*) to pinpoint areas at risk for the introduction and dispersion of emerging infectious diseases. The analysis identified that the areas at greatest risk were those in close proximity to the original outbreaks and several larger metropolitan areas. The trajectory and sustained spread of emerging infectious diseases can be anticipated with predictive modeling in real time. This study suggests that big data can be an indispensable tool for the prevention and control of emerging infectious diseases.

J.R. received partial funding from the Swedish Research Council for Sustainable Development (FORMAS) (no. 2017-01300). The funder had no influence on the research conducted.

About the Author

Dr. Rocklöv is professor of epidemiology and public health at Umeå University. His research interests focus on how climate, environmental, biological, medical, and social information can benefit preparedness and control of infectious diseases and be applied to early warning and response systems.

References

1. Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: towards faster and locally relevant systems. J Infect Dis. 2016;214(suppl 4):S380–5.
2. Butler D. When Google got flu wrong. Nature. 2013;494:155–6. http://dx.doi.org/10.1038/494155a
3. 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. http://dx.doi.org/10.1038/nature07634
4. 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. http://dx.doi.org/10.1371/journal.pcbi.1003581

5. Health HSoP. HealthMap. 2016 [cited 2017 Jan 5]. http://www.healthmap.org
6. World Health Organization [cited 2017 Jan 5]. http://www.who.int/csr/alertresponse/epidemicintelligence
7. Semenza JC, Lindgren E, Balkanyi L, Espinosa L, Almqvist MS, Penttinen P, et al. Determinants and drivers of infectious disease threat events in Europe. Emerg Infect Dis. 2016;22:581–9. http://dx.doi.org/10.3201/eid2204.151073
8. Semenza JC, Rocklöv J, Penttinen P, Lindgren E. Observed and projected drivers of emerging infectious diseases in Europe. Ann N Y Acad Sci. 2016;1382:73–83. http://dx.doi.org/10.1111/nyas.13132
9. Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International dispersal of dengue through air travel: importation risk for Europe. PLoS Negl Trop Dis. 2014;8:e3278. http://dx.doi.org/10.1371/journal.pntd.0003278
10. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, Paz Soldan V, Kochel TJ, Kitron U, et al. The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis. 2009;3:e481. http://dx.doi.org/10.1371/journal.pntd.0000481
11. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, et al. Using mobile phone data to predict the spatial spread of cholera. Sci Rep. 2015;5:8923. http://dx.doi.org/10.1038/srep08923
12. Finger F, Genolet T, Mari L, de Magny GC, Manga NM, Rinaldo A, et al. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proc Natl Acad Sci U S A. 2016;113:6421–6. http://dx.doi.org/10.1073/pnas.1522305113
13. Wesolowski A, Metcalf CJ, Eagle N, Kombich J, Grenfell BT, Bjørnstad ON, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proc Natl Acad Sci U S A. 2015;112:11114–9. http://dx.doi.org/10.1073/pnas.1423542112
14. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. J Infect Dis. 2016 ;214(suppl 4):S375–9.
15. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. Understanding human mobility from Twitter. PLoS One. 2015;10:e0131469. http://dx.doi.org/10.1371/journal.pone.0131469
16. Semenza JC, Zeller H. Integrated surveillance for prevention and control of emerging vector-borne diseases in Europe. Euro Surveill. 2014 ;19:pii:20757.
17. Semenza JC, Suk JE. Vector-borne diseases and climate change: a European perspective. FEMS Microbiol Lett. 2018;365. http://dx.doi.org/10.1093/femsle/fnx244
18. Laney D. 3D data management: controlling data volume, velocity, and variety [cited 2019 Apr 3]. https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
19. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. Health Aff (Millwood). 2014;33:1523–30. http://dx.doi.org/10.1377/hlthaff.2014.0771
20. Gandomi AHM, Haider M. Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manage. 2015;35:137–44. http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007
21. Ramadona A, Tozan Y, Lazuardi L, Rocklöv J. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. Plos Negl Trop Dis. In press 2019.
22. Istituto Superiore di Sanita. Italy: autochtonous cases of chikungunya virus [cited 2019 Mar 21] http://www.salute.gov.it/portale/temi/documenti/chikungunya/bollettino_chikungunya_20171221.pdf
23. Carletti F, Marsella P, Colavita F, Meschi S, Lalle E, Bordi L, et al. Full-length genome sequence of a chikungunya virus isolate from the 2017 autochthonous outbreak, Lazio region, Italy. Genome Announc. 2017;5:e01306-17. http://dx.doi.org/10.1128/genomeA.01306-17
24. Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. Chikungunya in the Americas. Lancet. 2014;383:514. http://dx.doi.org/10.1016/S0140-6736(14)60185-9
25. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One. 2010;5:e14118. http://dx.doi.org/10.1371/journal.pone.0014118
26. 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. http://dx.doi.org/10.1371/journal.pone.0019467
27. Kim EK, Seok JH, Oh JS, Lee HW, Kim KH. Use of Hangeul Twitter to track and predict human influenza infection. PLoS One. 2013;8:e69305. http://dx.doi.org/10.1371/journal.pone.0069305
28. Broniatowski DA, Dredze M, Paul MJ, Dugas A. Using social media to perform local influenza surveillance in an inner-city hospital: a retrospective observational study. JMIR Public Health Surveill. 2015 Jan–Jun;1:e5.
29. Italy Ministry of Health. National plan of surveillance and response to arbovirus transmitted by mosquitoes (*aedes* sp.), with particular reference to chikungunya, dengue and zikaviruses–2017 [cited 2019 Apr 3]. http://www.salute.gov.it/portale/temi/documenti/chikungunya/bollettino_chikungunya_20171221.pdf
30. Centers for Disease Control and Prevention. Travelers health. Chapter 3. Infectious diseases related to travel: chikungunya [cited 2018 Oct 16]. https://wwwnc.cdc.gov/travel/yellowbook/2018/infectious-diseases-related-to-travel/chikungunya
31. Rocklöv J, Quam MB, Sudre B, German M, Kraemer MUG, Brady O, et al. Assessing seasonal risks for the introduction and mosquito-borne spread of Zika virus in Europe. EBioMedicine. 2016;9:250–6. http://dx.doi.org/10.1016/j.ebiom.2016.06.009
32. Faria NR, Azevedo RDSDS, Kraemer MUG, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: early epidemiological and genetic findings. Science. 2016;352:345–9. http://dx.doi.org/10.1126/science.aaf5036
33. Reiter P, Lathrop S, Bunning M, Biggerstaff B, Singer D, Tiwari T, et al. Texas lifestyle limits transmission of dengue virus. Emerg Infect Dis. 2003;9:86–9. http://dx.doi.org/10.3201/eid0901.020220

Address for correspondence: Joacim Rocklöv, Umeå University, Department of Public Health and Clinical Medicine, 901 87 Umeå, Sweden; email: joacim.rocklov@umu.se

Article DOI: https://doi.org/10.3201/eid2506.180138

Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017

Appendix 1

Epidemic Intelligence Data

We analyzed the 2 outbreak zones in the Var department of France (15 confirmed and 2 probable cases) and around the cities of Anzio and Rome in the Lazio region of central Italy (206 confirmed cases) and 74 confirmed cases in the Calabria region in south Italy (Appendix 3, https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App3.pdf) (*1*–*4*). The disease vector *Ae. albopictus* mosquito is well established in all outbreak regions (*5*). Worldwide monthly chikungunya outbreak reports were compiled by the Epidemic Intelligence team at the European Centre for Disease Prevention and Control (Appendix 3) (*6*). We mapped and visualized the passenger volume of outbound flights to Europe from areas with chikungunya activity by month for March, April, May, and June 2017.

Air Passenger Volume

We analyzed anonymized flight itinerary data obtained from the IATA Market Intelligence Services and calculated the monthly volume of air passenger-journeys in 2016 (latest data available; presumed to be similar to 2017) from worldwide airports in areas with chikungunya virus active transmission to a final destination in Europe, by using a previously described method (*7*) (Appendix 3). The distribution of the number of passenger-journeys arriving into Europe from airports located in areas with active chikungunya virus transmission was then overlaid with European vector surveillance data compiled by the European Centre for Disease Prevention and Control (VectorNet, https://vectornet.ecdc.europa.eu) for *Ae. albopictus* mosquitoes by using ESRI ArcGIS (*5*).

Twitter Data

We developed a mining algorithm and collected Tweets by using the Twitter Streaming Application Programming Interface (https://developer.twitter.com). Although the tweets collected from the API represent only ≈1% of the total Tweeter feed, when geographic boundary boxes are used for data collection it provides a high representation of the overall geo-located activity on Twitter (*8*). We filtered the collected tweets based on location by using geocodes, and we extracted only those originating from the study area in July, August, and up to September 19, 2017. We longitudinally analyzed 8,120,417 Tweets. When Tweets from the same users could be followed by geographic coordinates, we obtained users' individual files. We analyzed unidirectional mobility of Twitter users by estimating the frequency of a user being observed in a specific geographic department within the study area and later being observed in any other department within the same month. To compute a rate, we aggregated the total number of movements in a month between any 2 departments and divided this by the total movement across all the departments. The range of all between-department mobility values was 0–1 and added up to 1 when summarized across the departments for inbound and outbound movements. We derived this quantity as a proxy for mobility proximity between any 2 departments and computed it for each month.

Vectorial Capacity

To estimate seasonal variability in the ability of *Ae. albopictus* mosquitoes to transmit chikungunya virus, we modified our previously established climate dependent vectorial capacity arbovirus models (*9*,*10*). The model uses temperature and diurnal temperature range to estimate the epidemic potential of an outbreak. Theoretically, vectorial capacity is related to R_0. More exactly, the R_0 is a function of vectorial capacity (VC) and duration of viremia in humans (T_h), that is $R_0 = VC \times T_h$. Vectorial capacity is a function of vector competence, vector lifespan, and extrinsic incubation period (*11*) and is defined mathematically in Appendix 3.

The 4 vector-related parameters in the vectorial capacity are 1) average vector biting rate, a; 2) the product of the probability of vector infection (b_{mi}) and transmission per bite (b_{mt}), b_m; 3) extrinsic incubation period, n (i.e., the interval between the acquisition of a pathogen by a vector

and the vector's ability to then transmit the pathogen to another susceptible host); and 4) vector mortality rate, μ_m ; and 4, female vector-to-human population ratio, *m.*

The effect of temperature on the ability of *Ae. albopictus* mosquitoes to transmit chikungunya virus has not been well studied. However, μ_m and *a* in relation to temperature have been described for *Ae. albopictus* mosquitoes. We assumed that *n*, b_m would have a dependence on temperature for chikungunya virus transmission similar to that for dengue virus, although we found evidence to support that it can be slightly lower at around 90% (*11–13*) and that *n* is shorter, peaking at around 8 instead of 10 days (*11–13*). Similar to a previous study (*9*), *m* was assumed to be proportional to its temperature-dependent survival curve. Parameter relationships used in the analysis are provided in Appendix 1 Figure.

Climate Data

We used data from the Climate Research Unit of East Anglia University (*14*) to estimate the average vectorial capacity for July, August, September, and October during 1996–2015. To describe the effect of warmer than usual temperature, we increased the average monthly temperature to its 75th percentile value for each month and recalculated the vectorial capacity. The Climate Research Unit data, originally provided in 0.5° × 0.5° grids by latitude and longitude, were resampled to fit into a grid of 0.01° to better align with the geographic departments of the study area.

References

1. Italy Ministry of Health. National plan of surveillance and response to arbovirus transmitted by mosquitoes (*aedes* sp.), with particular reference to chikungunya, dengue and zikaviruses—2017 [cited 2019 Apr 3]. http://www.salute.gov.it/portale/temi/documenti/chikungunya/bollettino_chikungunya_20171221.pdf

2. European Centre for Disease Prevention and Control. Cluster of autochthonous chikungunya cases in France–23 August 2017. Stockholm: The Centre; 2017.

3. Venturi G, Di Luca M, Fortuna C, Remoli ME, Riccardo F, Severini F, et al. Detection of a chikungunya outbreak in Central Italy, August to September 2017. Euro Surveill. 2017;22. http://dx.doi.org/10.2807/1560-7917.ES.2017.22.39.17-00646

4. Calba C, Guerbois-Galla M, Franke F, Jeannin C, Auzet-Caillaud M, Grard G, et al. Preliminary report of an autochthonous chikungunya outbreak in France, July to September 2017. Euro Surveill. 2017;22. http://dx.doi.org/10.2807/1560-7917.ES.2017.22.39.17-00647

5. European Centre for Disease Prevention and Control. *Aedes albopictus*—current known distribution in Europe, April 2017. Stockholm: The Centre; 2017.

6. European Centre for Disease Prevention and Control. Communicable disease threats report. Week 26, 25 June–1 July 2017 [cited 2019 Apr 3]. https://ecdc.europa.eu/en/threats-and-outbreaks/reports-and-data/weekly-threats

7. Semenza JC, Sudre B, Miniota J, Rossi M, Hu W, Kossowsky D, et al. International dispersal of dengue through air travel: importation risk for Europe. PLoS Negl Trop Dis. 2014;8:e3278. PubMed http://dx.doi.org/10.1371/journal.pntd.0003278

8. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. In: International Conference on Weblogs and Social Media: Association for the Advancement of Artificial Intelligence; 2013. p. 400–8.

9. Rocklöv J, Quam MB, Sudre B, German M, Kraemer MUG, Brady O, et al. Assessing seasonal risks for the introduction and mosquito-borne spread of Zika virus in Europe. EBioMedicine. 2016;9:250–6. PubMed http://dx.doi.org/10.1016/j.ebiom.2016.06.009

10. Liu-Helmersson J. Quam M, Wilder-Smith A, Stenlund H, Ebi K, Massad E, et al. Climate change and *Aedes* vectors: 21st century projections for dengue transmission in Europe. 2016;7:267–77.

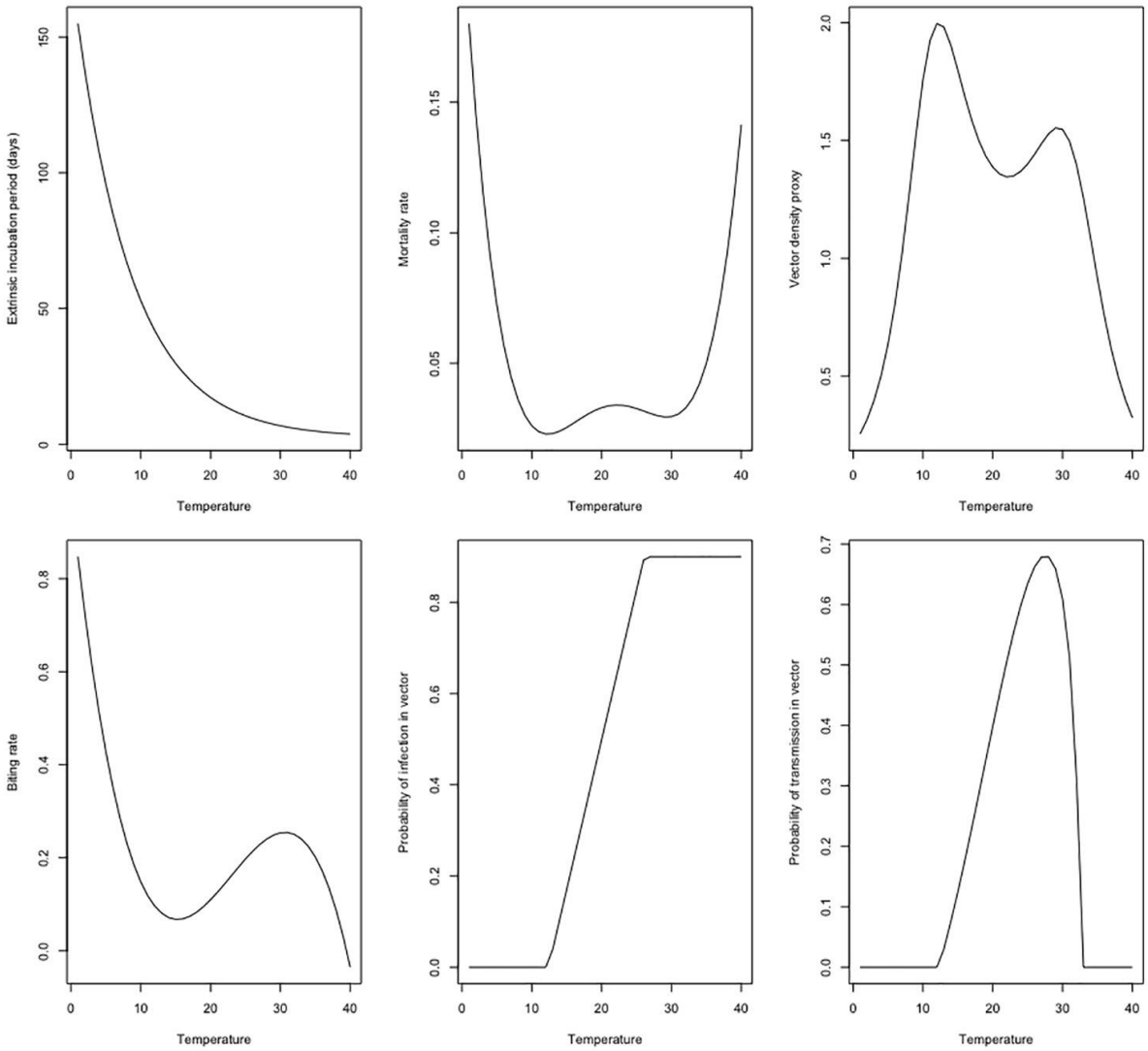
11. Christofferson RC, Chisenhall DM, Wearing HJ, Mores CN. Chikungunya viral fitness measures within the vector and subsequent transmission potential. PLoS One. 2014;9:e110538. PubMed http://dx.doi.org/10.1371/journal.pone.0110538

12. Vega-Rua A, Zouache K, Caro V, Diancourt L, Delaunay P, Grandadam M, et al. High efficiency of temperate *Aedes albopictus* to transmit chikungunya and dengue viruses in the southeast of France. PloS One. 2013; 8:e59716. http://dx.doi.org/10.1371/journal.pone.0059716

13. Vega-Rúa A, Zouache K, Girod R, Failloux AB, Lourenço-de-Oliveira R. High level of vector competence of *Aedes aegypti* and *Aedes albopictus* from ten American countries as a crucial factor in the spread of chikungunya virus. J Virol. 2014;88:6294–306. PubMed http://dx.doi.org/10.1128/JVI.00370-14

14. University of East Anglia Climatic Research Unit; Harris IC, Jones PD. CRU TS4.00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high resolution gridded data of month-by-

month variation in climate (Jan. 1901- Dec. 2015). 2017; Centre for Environmental Data Analysis. http://dx.doi.org/10.5285/edf8febfdaad48abb2cbaf7d7e846a86

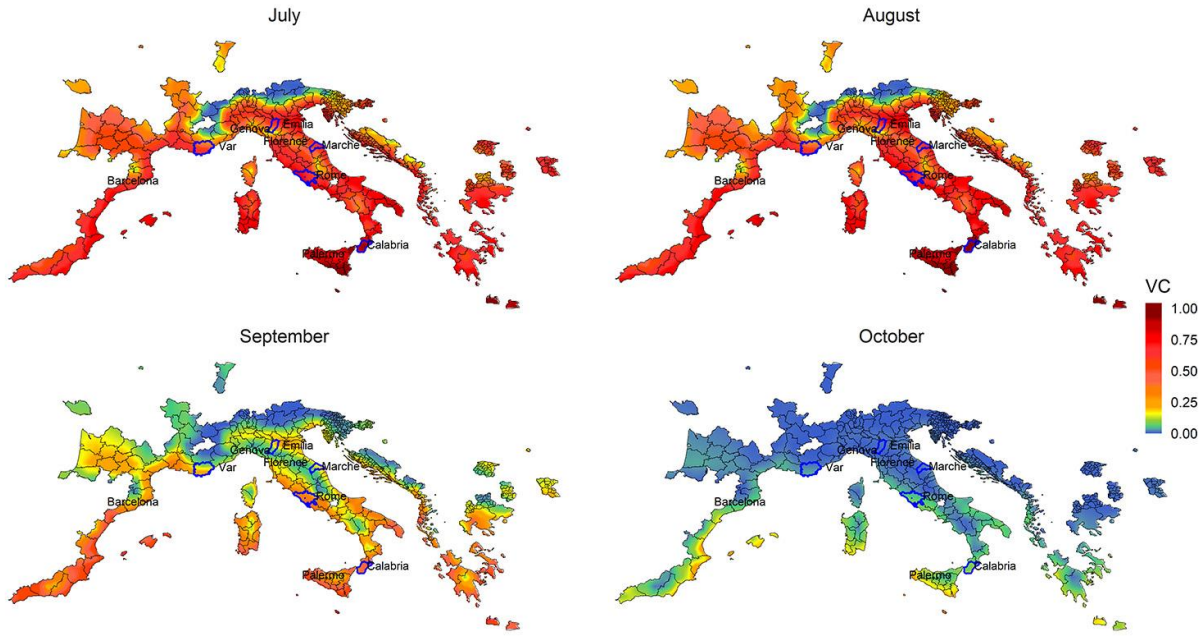


Appendix 1 Figure. The relationship of vector-related parameters to temperature (°C), describing the ability of *Aedes albopictus* to transmit chikungunya virus.

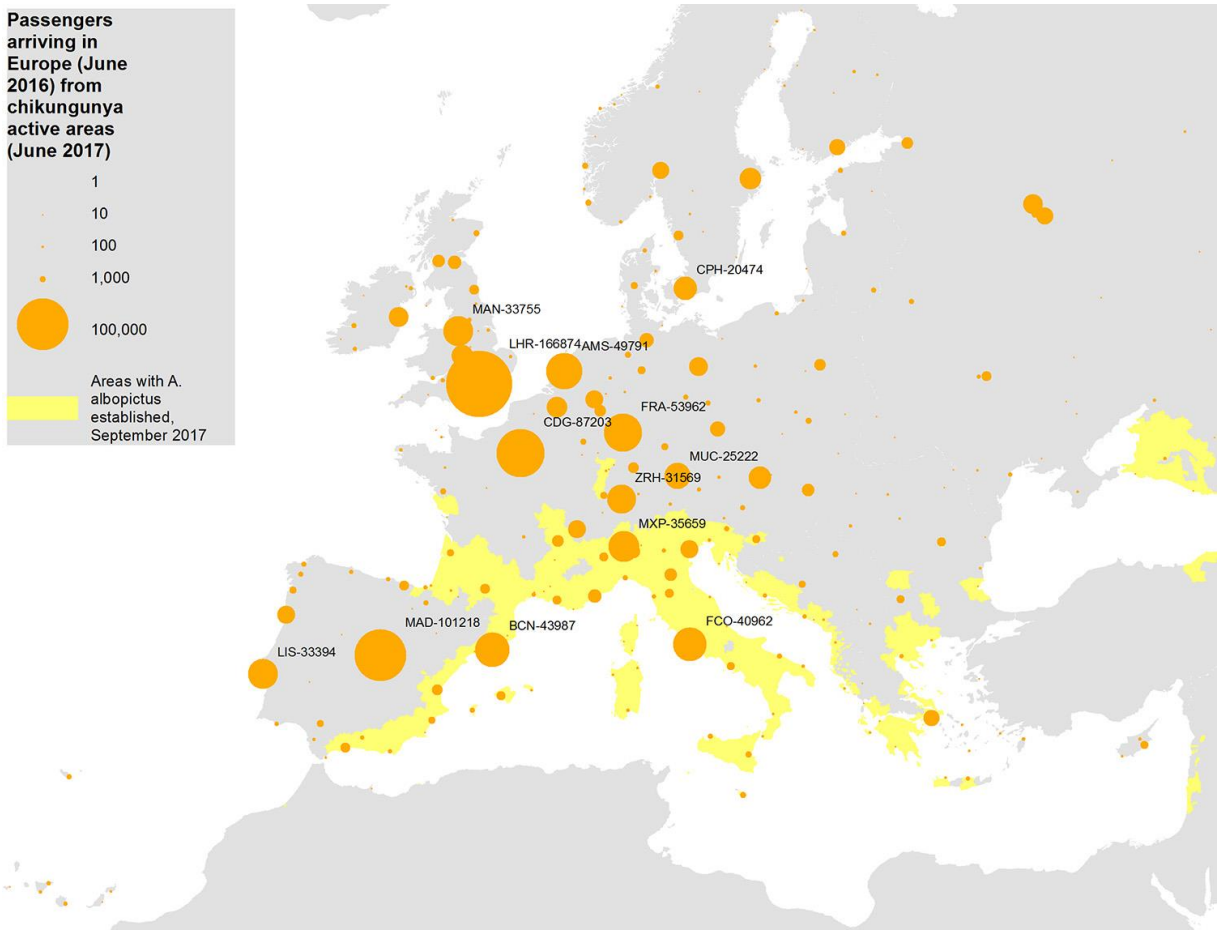
Article DOI: https://doi.org/10.3201/eid2506.180138

Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017

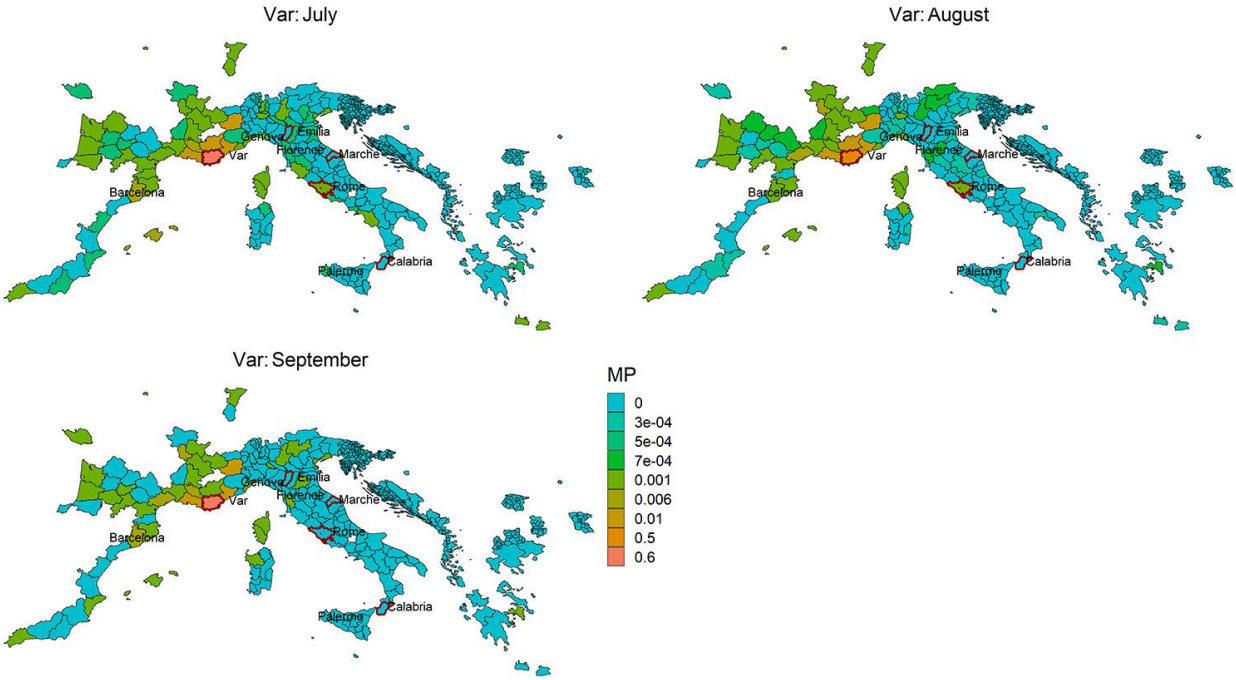
Appendix 2



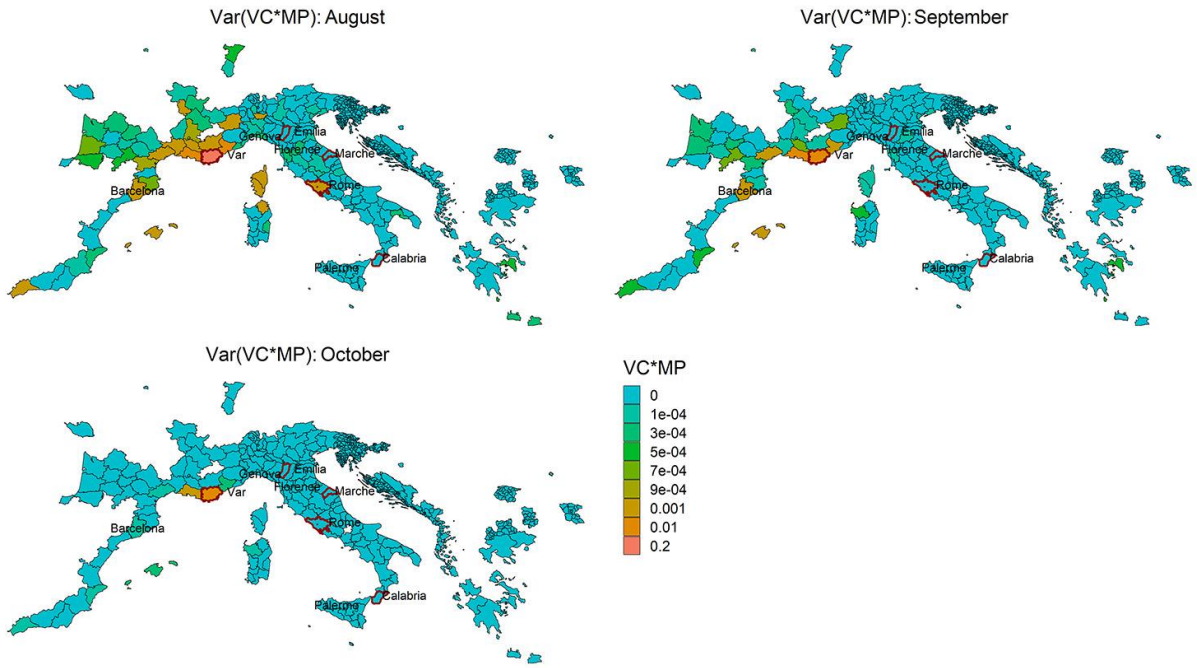
Appendix 2 Figure 1. Average monthly vectorial capacity (VC) estimates derived on the basis of temperature averaging to the 75th percentile of monthly distribution, July-October, 2017. Areas with autochthonous transmission are indicated by colored polygons.



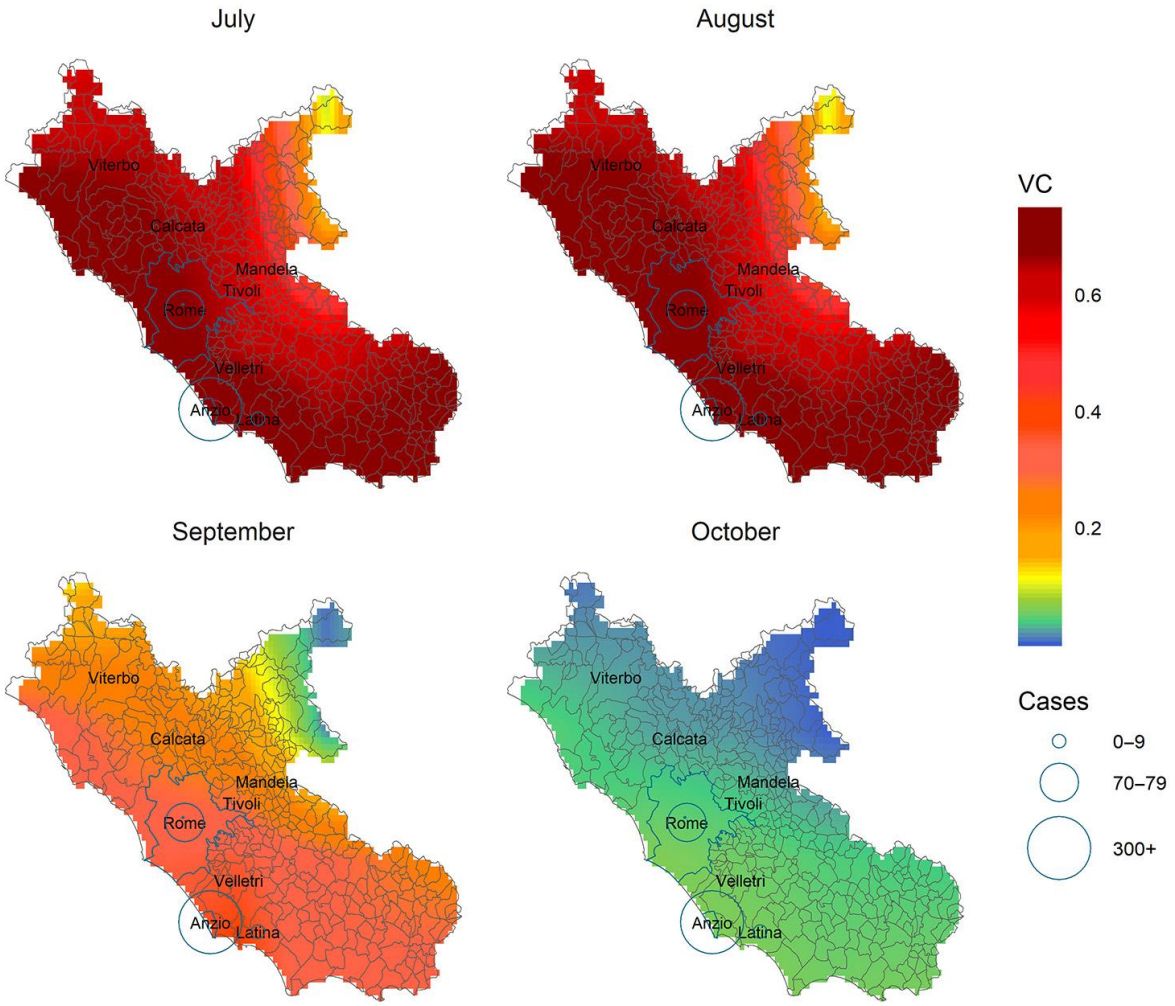
Appendix 2 Figure 2. Number of passengers arriving from chikungunya transmission active areas into Europe, August 2017.



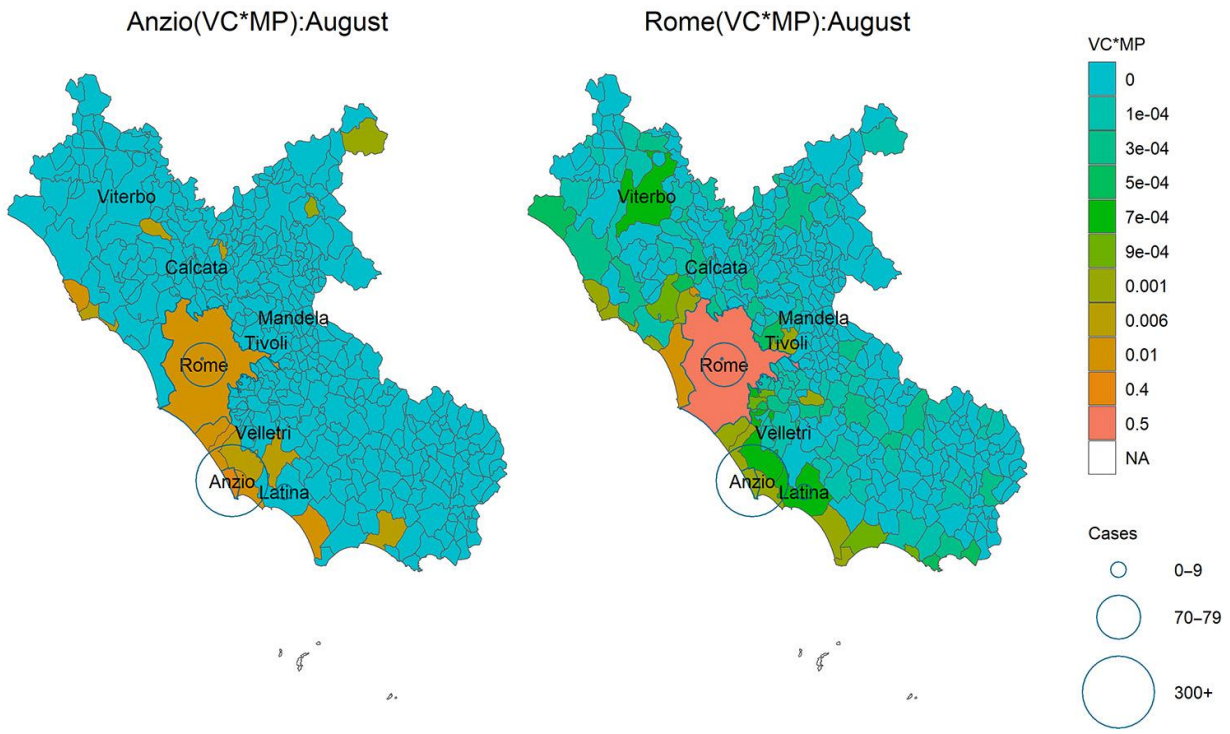
Appendix 2 Figure 3. Mobility proximity (MP) estimates from the Var department, France, to areas in Europe with stable *Ae. albopictus* populations, July-September 2017. The polygons mark the outbreak areas.



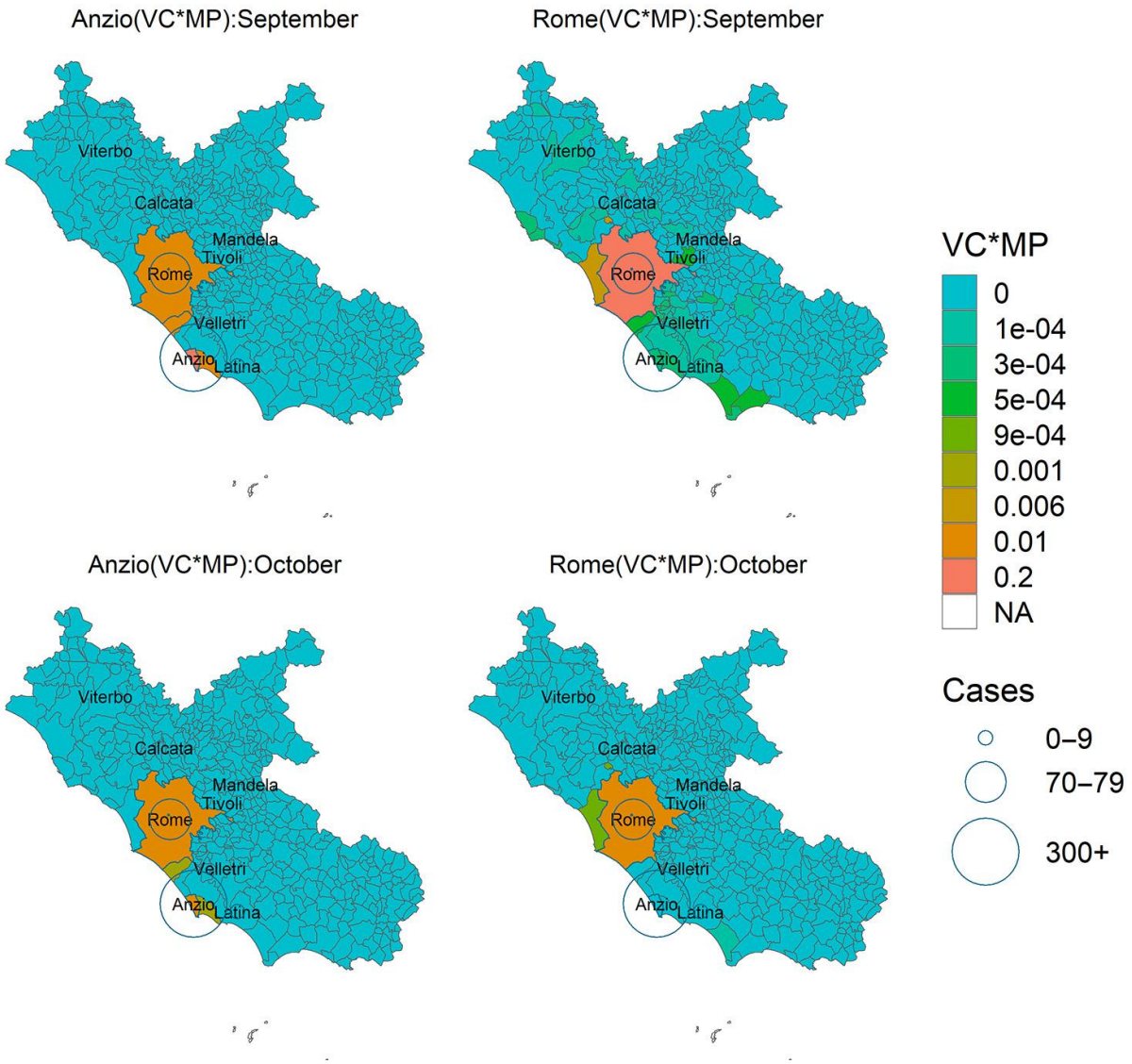
Appendix 2 Figure 4. Estimated risk areas of chikungunya spread from the outbreak areas in the Var department, France, based on combined vectorial capacity (VC) and mobility proximity (MP) estimates, August-October 2017. The polygons mark the outbreak areas.



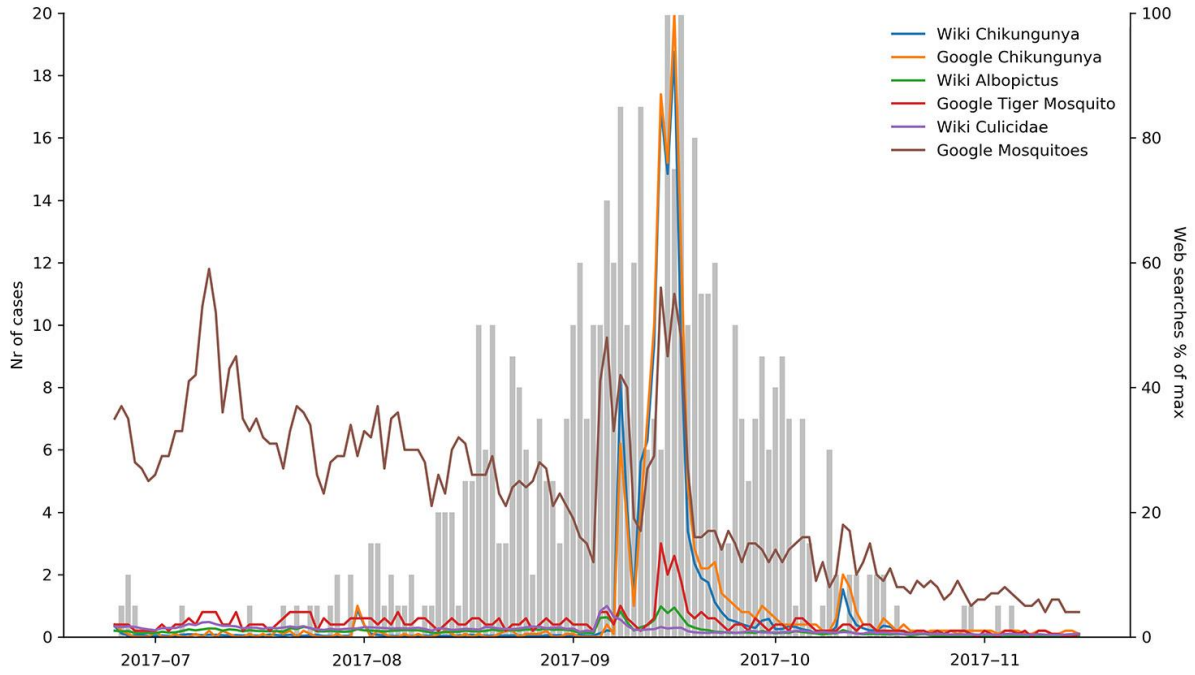
Appendix 2 Figure 5. Vectorial capacity (VC) estimates for Lazio region, July-October 2017, based on average climatic conditions during the period 1996–2015. The number of reported cases are overlaid as circles.



Appendix 2 Figure 6. Estimated risk areas of chikungunya spread from the outbreak areas of Anzio and Rome in Lazio region based on combined vectorial capacity (VC) and mobility proximity (MP) and estimates, August 2017. The number of reported cases are overlaid as circles.



Appendix 2 Figure 7. Estimated risk areas of chikungunya spread from the outbreak areas of Anzio and Rome in Lazio region based on combined vectorial capacity (VC) and mobility proximity (MP) and estimates, September and October 2017. The number of reported cases are overlaid as circles.



Appendix 2 Figure 8. Hits of Google and Wikipedia on search terms related to mosquito and chikungunya.

Article DOI: https://doi.org/10.3201/eid2506.180138

Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017

Appendix 3

Epidemic Intelligence Data

The analysis was conducted for a specified study area in reasonable proximity to the two outbreak zones in 15 confirmed and 2 probable cases reported in Var department (9 confirmed and 2 probable cases reported in Cannet-des-Maures and 6 confirmed cases in Taradeau) and around the cities of Anzio and Rome in Lazio region of central Italy (206 confirmed cases). Data on confirmed and suspected chikungunya cases were obtained from epidemic intelligence data and reports. The first reports of autochthonous transmission came from Var department, followed by Lazio region, and transmission was reported later on from Calabria (74 confirmed cases), Emilia-Romagna (1 confirmed case) and the Marche (1 confirmed case) regions in Italy (*1*–*4*). In all outbreak regions, the disease vector *Ae. albopictus* is known to be well established (*5*).

Worldwide monthly chikungunya outbreak reports were compiled by the Epidemic Intelligence team at the European Centre for Disease Prevention and Control (ECDC) based on data mining from the World Health Organization, Ministries of Health, and other official and non-official sources, such as media reports, to survey the current worldwide chikungunya situation (*6*). Rather than gauging chikungunya incidence qualitatively, our assessment was based on chikungunya events identified by the ECDC through web crawl searches and from confidential/official sources, such as Early Warning and Response Systems, Program for Monitoring Emerging Diseases, Medical Information System, and Global Public Health Intelligence Network. Weekly notifications from these sources were evaluated and geocoded by month. We mapped and visualized the passenger volume of outbound flights to Europe from areas with chikungunya activity by month, namely for March, April, May, and June 2017.

Air Passenger Volume

The International Air Transport Association (IATA) database has the most voluminous and comprehensive aviation data from over 80,000 travel and online agencies, 400 airlines, and 170 countries. Travelers on commercial, connecting and scheduled charter flights are captured. We analyzed anonymized flight itinerary data obtained from IATA Market Intelligence Services and calculated the monthly volume of air passenger-journeys in 2016 (latest data available; presumed to be similar to 2017) from airports worldwide located in areas with chikungunya active transmission with a final destination in Europe. We assumed that human passenger-journeys were the main vehicle of viral spread, rather than infected mosquitoes in airplanes, based on the index cases of past outbreaks that had a travel history to endemic areas. These large-scale IATA passenger data represent ≈93% of the world's commercial air traffic, while the remainder was estimated using market intelligence. The distribution of number of passenger-journeys arriving into Europe from airports located in areas with active chikungunya transmission was then overlaid with European vector surveillance data compiled by the ECDC (VectorNet) for *Ae. albopictus* using ESRI ArcGIS (*5*). Chikungunya continues to spread internationally due to several factors, most notably the adaptive mutations in the viral genome that enabled the virus to be more easily transmitted by *Ae. albopictus* (*7*). This vector has expanded its geographic range through increasing global trade of used tires and plants and established itself in areas with suitable climate and habitat in many parts of the world. However, *Ae. aegypti*, another competent vector for the Italian 2017 chikungunya virus strain (*8*), is largely not present in continental Europe with the exception of a small region around the eastern coast of the Black Sea. Once an outbreak occurs the disease can entrench itself in the local vector population and become endemic if climate allows vectors to be active around the year.

Vectorial Capacity

The vectorial capacity can be described by the following mathematical expression:

$VC = ma^2b_me^{-\mu mn}/\mu m$. See Appendix 1, https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App1.pdf, for a more detailed description.

Wikipedia and Google Trend Data

Wikipedia is a free, internet-based encyclopedia structured as an interconnected network of open-content articles and is considered one of the top Web sites visited globally (*9*). Internet users typically use Wikipedia to access background information on a specific topic and related subtopics. Although web searches, usually using Google, lead users to a Wikipedia article, the majority of users follow the links provided in the article to access other related articles. Therefore, it has been argued that Wikipedia access statistics may provide valuable insight into the emergence and shift of collective interests or activities of individuals, and sudden peaks in user access of specific Wikipedia pages may reflect extreme events in nature or society (*10*). Here we chose specific articles related, namely mosquitoes, albopictus and chikungunya, across the Italian, French, German (as control) and English (as reference) language editions of Wikipedia, and extracted daily article access logs, which provide a summary file listing the number of access requests for each article per day in each language during the period from July to November 2017. The Wikipedia data was downloaded 2018-10-13 using the mwviews.api/PageviewsClient (https://github.com/mediawiki-utilities/python-mwviews) Wikipedia articles "Aedes_albopictus" (redirected from "Zanzara tigre" = tiger mosquito), "Culicidae" (redirected from "Zanzara/e" = mosquitoes singular and plural), "Chikungunya) We also downloaded Google Trends data 2018-01-26 from https://trends.google.com/trends/explore?date = 2017-06-25%202017-11-15&geo = IT&q = %2Fm%2F09f96,%2Fm%2F01__7l,%2Fm%2F01yy_q using search topics which include several similar search terms (https://support.google.com/trends/answer/4359550). As Wikipedia gives absolute page hits and Google gives only proportions, the Wikipedia articles were added and calculated as percentages from the maximum number of page hits of the three search terms.

Our analysis of Wikipedia access logs and Google Trends shows clear peaks in terms of number of access requests for the articles on mosquitoes (Culicidae) and Albopictus first in June/July and then in mid-September in the Italian language version of Wikipedia (Appendix 2 Figure 8, https://wwwnc.cdc.gov/EID/article/25/6/18-0138-App2.pdf). A distinct peak in access requests was also observed for Tiger mosquitoes in mid-September in Italian language. We did not observe such peaks for these Wikipedia articles in 2016. For the articles on chikungunya, we

observed peaks in early August in Italian Wikipedia and in mid-August in French Wikipedia, which probably indicates an increasing awareness of the disease among the public. We observed a larger peak in access requests on chikungunya in Italian Wikipedia later in mid-September, followed by another small peak in mid-October, probably as a result of the continued exposure of the public through the media to the outbreak news because of its spread. We found a strong correlation between the number of notified chikungunya cases and the access requests for chikungunya in the Italian language version of Wikipedia (Figure 2). An overlay with the search data on chikungunya from Google yielded a similar pattern, probably because Wikipedia hits are typically preceded by Google searches (Appendix 2 Figure 8). These observations suggest that Wikipedia access logs to articles on specific health topics have the potential to supplement disease surveillance and outbreak prediction efforts when combined with disease incidence data, as demonstrated in (*11*).

References

1. European Centre for Disease Prevention and Control. Cluster of autochthonous chikungunya cases in France–23 August 2017 [cited 2019 Apr 3]. https://ecdc.europa.eu/sites/portal/files/documents/RRA-Chikungunya-France-revised-Aug-2017.pdf

2 Venturi G, Di Luca M, Fortuna C, Remoli ME, Riccardo F, Severini F, et al. Detection of a chikungunya outbreak in central Italy, August to September 2017. Euro Surveill. 2017;22(39):pii :17-00646. http://dx.doi.org/10.2807/1560-7917.ES.2017.22.39.17-00646

3. Italy Ministry of Health. National plan of surveillance and response to arbovirus transmitted by mosquitoes (*aedes* sp.), with particular reference to chikungunya, dengue and zikaviruses–2017 [cited 2019 Apr 3]. http://www.salute.gov.it/portale/temi/documenti/chikungunya/bollettino_chikungunya_20171221.pdf

4. Calba C, Guerbois-Galla M, Franke F, Jeannin C, Auzet-Caillaud M, Grard G, et al. Preliminary report of an autochthonous chikungunya outbreak in France, July to September 2017. Euro Surveill. 2017;22(39). http://dx.doi.org/10.2807/1560-7917.ES.2017.22.39.17-00647

5. European Centre for Disease Control and Prevention. *Aedes albopictus*–current known distribution in Europe, April 2017 [cited 2019 Apr 3]. https://ecdc.europa.eu/en/publications-data/aedes-albopictus-current-known-distribution-europe-april-2017

6. European Centre for Disease Prevention and Control. Communicable disease threats report. Week 26, 25 June–1 July 2017 [cited 2019 Apr 3]. https://ecdc.europa.eu/en/threats-and-outbreaks/reports-and-data/weekly-threats

7. Tsetsarkin KA, Weaver SC. Sequential adaptive mutations enhance efficient vector switching by chikungunya virus and its epidemic emergence. PLoS Pathog. 2011;7:e1002412. PubMed http://dx.doi.org/10.1371/journal.ppat.1002412

8. Carletti F, Marsella P, Colavita F, Meschi S, Lalle E, Bordi L, et al. Full-length genome sequence of a chikungunya virus isolate from the 2017 autochthonous outbreak, Lazio region, Italy. Genome Announc. 2017;5:e01306-17. PubMed http://dx.doi.org/10.1128/genomeA.01306-17

9. Schroeder R, Taylor L. Big data and Wikipedia research: social science knowledge across disciplinary divides. Inf Commun Soc. 2015;18:1039–56. http://dx.doi.org/10.1080/1369118X.2015.1008538

10. Kämpf M, Tismer S, Kantelhardt JW, Muchnik L. Fluctuations in Wikipedia access-rate and edit-event data. Physica A. 2012;391:6101–11. http://dx.doi.org/10.1016/j.physa.2012.07.004

11. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R. Global disease monitoring and forecasting with Wikipedia. PLOS Comput Biol. 2014;10:e1003892. PubMed http://dx.doi.org/10.1371/journal.pcbi.1003892