Cell-phone traces reveal infection-associated behavioral change

Ymir Vigfusson[a,b,1,2], Thorgeir A. Karlsson[b,1], Derek Onken[a,1], Congzheng Song[c], Atli F. Einarsson[b], Nishant Kishore[d], Rebecca M. Mitchell[a], Ellen Brooks-Pollock[e], Gudrun Sigmundsdottir[f,g], and Leon Danon[h,i]

[a]Simbiosys Lab, Department of Computer Science, Emory University, Atlanta, GA 30322; [b]School of Computer Science, Reykjavik University, 101 Reykjavik, Iceland; [c]Department of Computer Science, Cornell University, Ithaca, NY 14853; [d]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115; [e]Department of Veterinary Medicine and Population Health Sciences, University of Bristol, Oakfield Grove, Bristol BS8 2BN, United Kingdom; [f]Landspitali University Hospital, 101 Reykjavik, Iceland; [g]Centre for Health Security and Communicable Disease Control, 101 Reykjavik, Iceland; [h]Department of Engineering Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom; and [i]The Alan Turing Institute, British Library, London NW1 2DB, United Kingdom.

Epidemic preparedness depends on our ability to predict the trajectory of an epidemic and the human behavior that drives spread in the event of an outbreak. Changes to behavior during an outbreak limit the reliability of syndromic surveillance using large-scale data sources, such as online social media or search behavior, which could otherwise supplement healthcare-based outbreak-prediction methods. Here, we measure behavior change reflected in mobile-phone call-detail records (CDRs), a source of passively collected real-time behavioral information, using an anonymously linked dataset of cell-phone users and their date of influenza-like illness diagnosis during the 2009 H1N1v pandemic. We demonstrate that mobile-phone use during illness differs measurably from routine behavior: Diagnosed individuals exhibit less movement than normal (1.1 to 1.4 fewer unique tower locations; $P < 3.2 \times 10^{-3}$), on average, in the 2 to 4 d around diagnosis and place fewer calls (2.3 to 3.3 fewer calls; $P < 5.6 \times 10^{-4}$) while spending longer on the phone (41- to 66-s average increase; $P < 4.6 \times 10^{-10}$) than usual on the day following diagnosis. The results suggest that anonymously linked CDRs and health data may be sufficiently granular to augment epidemic surveillance efforts and that infectious disease-modeling efforts lacking explicit behavior-change mechanisms need to be revisited.

disease | surveillance | call detail records | influenza | outbreak

Infectious disease outbreaks remain a major threat to humanity in the 21st century, as evidenced by the ongoing pandemic of COVID-19 (1) and 5 of 10 threats to global health identified by the World Health Organization being related to infectious disease (2). Estimating the current and future burden of disease through surveillance and predictive modeling is essential for appropriate allocation of resources aimed at reducing impact, especially in the early stages of an outbreak.

Traditional influenza healthcare-based surveillance methods rely on data gathered from symptomatic individuals seeking medical treatment from doctors. These approaches suffer from delays in reporting that differ from setting to setting and difficulty in identifying unusual activity (3). Such issues led to the development of alternative syndromic surveillance methods (4) that combine a broad range of data sources on behavioral markers; some were developed, used, and assessed during the H1N1v pandemic (5). These surveillance methods include analyzing patterns in social media such as Twitter (6, 7), search-engine queries (8–10), over-the-counter medication sales (11), airport traffic patterns (12), city traffic patterns (13), cell-phone surveys (14), or ensemble methods that incorporate survey data (15). Directly inferring disease incidence from these sources also assumes that the cause of behavior change is known and usually associated with influenza. Yet, studies indicate that individuals alter behavior for various reasons, even when not symptomatic, e.g., to avoid infection (16) or due to anxiety (17), complicating estimation of infectious disease burden (18).

Whereas data sources that depend on active, conscious user participation may produce unreliable estimates (14, 20), call-detail records (CDRs) can act as a passive pattern sensor (21). Mobile networks pervade most nations: In raw numbers, 2019 cell-phone subscriptions in developed and developing countries exceeded 100% of their populations (22), although mobile use invariably skews away from underresourced groups (23). CDRs, collected in real-time, contain spatiotemporal information that captures mobility. Past analyses have used cell-phone data to study human-movement scaling (13), social-network structure inference (24), poverty and wealth prediction (25), and risk and spread of multiple diseases, including malaria (26, 27), cholera (28), and influenza (29). Furthermore, smartphone apps have been used to track behavior change in relation to influenza onset (30) or as contact trackers during the COVID-19 pandemic (31, 32). These methods are all limited by either unreliable health data (self-diagnosed symptoms), aggregate-level data to model the population (33), or fraught with privacy concerns (34). Until now, the link with verified health data at the individual level has been missing.

Here, we explicitly combine CDRs with information from the 2009 H1N1v pandemic collected by the national

Significance

Infectious disease control critically depends on surveillance and predictive modeling of outbreaks. We argue that routine mobile-phone use can provide a source of infectious disease information via the measurements of behavioral changes in call-detail records (CDRs) collected for billing. In anonymous CDR metadata linked with individual health information from the A(H1N1)pdm09 outbreak in Iceland, we observe that people moved significantly less and placed fewer, but longer, calls in the few days around diagnosis than normal. These results suggest that disease-transmission models should explicitly consider behavior changes during outbreaks and advance mobile-phone traces as a potential universal data source for such efforts.

POPULATION BIOLOGY

healthcare-based surveillance system used by all health providers in Iceland through a protocol that maintains reasonable expectations of individual privacy from government surveillance. The influenza pandemic reached Iceland in May 2009 (19), with a shallow peak before the school holidays in May/June 2009, followed by a dip over the summer and a strong peak in October 2009 (Fig. 1). The outbreak started in the capital of Reykjavík, home to 37% of the population of 318,499, approximately 1 wk ahead of the rest of the country (19). Health officials recorded the date of diagnosis (DoD) of 10,175 clinically diagnosed cases of influenza-like illness (ILI) around the country between June 4, 2009, and February 11, 2010. Of 3,011 samples taken, 700 were confirmed by a real-time (PCR) protocol to be H1N1v influenza infections (19); we assume that other patients diagnosed with ILI were infected with the same strain, which displaced other strains until February 2010 (35).

We analyzed behavioral patterns in Iceland extracted from the CDRs, provided in a deidentified format by a major mobile-network operator (MNO). The CDR logs span a broad time period around the 2009 outbreak. Mobile-phone owners were anonymously matched to records of ILI diagnosis, yielding DoD and CDR traces for 1,434 diagnosed individuals after data processing. We measured and identified behavioral traits that showed significant changes in the diagnosed group around the DoD compared to a control group.

Methods

Data Collection. The original dataset joins individual CDRs that MNOs routinely gather for billing purposes with individual-level ILI diagnosis data from Iceland's Centre for Health Security and Communicable Disease Control (CHS-CDC), which collects and stores all records of ILI diagnoses in Iceland. We developed and used a privacy-preserving data hand-off and merging protocol approved by Iceland's Bioethics Committee (Vísindasiðanefnd): A large MNO sent encrypted phone identifiers (IDs) and national ID numbers (NINs, which are public information in Iceland) to the CHS-CDC. The CHS-CDC supplied dates of ILI diagnoses for NINs and then replaced NINs with an anonymous encrypted identifier before providing the data to us (*SI Appendix, Data Linking and Privacy*). The MNO provided us with CDR data (*SI Appendix, Mobile Network Data*) containing the encrypted IDs of the phones on either side of a call, the timestamp, the length of the call (in seconds), and the geographical coordinates of the cell-phone towers that interacted with the phones (*SI Appendix, Table S1*). No demographic or private data, such as age, gender, or contents of calls or texts, were included. The cell tower accessed during normal phone use provides a proxy for the device's location. The granularity of location varies with locality—regional tower density increases proportionally with regional population (Fig. 1). At the time, MNOs provided cell coverage for virtually all residences in Iceland, either directly through their network or through a roaming service. We filtered out individuals with multiple subscriptions (*SI Appendix, Data Preprocessing*). Using phone-ownership information, each

phone was matched to the DoD of its owner for the subset of users that pay only for one phone. This postprocessed subset, referred to as the dataset below, accounted for 25 to 30% of the MNO's users and encompassed all data analyzed in our paper. We defined the home tower of each individual as the tower that picks up more calls and texts between midnight and 8 a.m. than other towers. The distribution of home-tower locations was strongly correlated with residential census counts for the corresponding postal codes for our dataset ($r = 0.86$, $P < 8 \times 10^{-49}$) and among those with ILI diagnosis ($r = 0.88$, $P < 2 \times 10^{-43}$). The home towers were thus spatially representative for the entire Icelandic population. We focused our analysis on the 1,434 diagnosed users who generated sufficient CDR data to establish a home-tower location in the 4-wk period centered on their DoD.

Feature Extraction. To characterize user behavior, we extracted 36 features (independent variables) from both incoming and outgoing CDR data encompassing movement, activity, and social-network behavioral patterns (*SI Appendix, Feature Extraction*). Most features exhibited a right-skewed distribution (*SI Appendix, Fig. S2*) and shared general characteristics across control and diagnosis groups. They include the following (boldface in Table 1).

Number of towers visited measures the number of unique tower coordinates connected to by the cell phone within a time interval (bin). This feature helps describe movement during the time period, but can inflate in areas where multiple towers can provide cellular signal.

Mean call duration (incoming and outgoing) measures call activity by dividing the total duration of calls by the number of calls the user placed or received in the time interval.

Number of calls (outgoing) measures the number of calls placed by the device in the time interval.

Departure from Routine Behavior. We use $x_{fid} = E_f(i, d)$ to denote the raw feature value for a feature f, extracted from the CDR by function E, for individual i, and on day d. Extraction is performed for all features f in Table 1.

To control for the weekly behavioral routine of individual i, each feature value is detrended through linear regression over values of the same weekday in the past W weeks. Specifically, let

$$p_j = x_{fi,(d-7 \cdot (W-j))} \quad \text{for } j = 0, 1, \ldots, W,$$

and denote by J those indices $j \in \{0, 1, \ldots, W\}$, where p_j is defined. Then, $(p_j)_{j \in J}$ is the measured behavior on the same day of the week from the previous W weeks before day d for feature f and individual i, with p_W denoting the behavior in week W.

We used $W = 10$ weeks of past data to correct for seasonality in our experiments, which gave comparable results to an alternative approach to detrending based on ranking features and normalizing them (*SI Appendix, Seasonality*).

Based on the data, we used a linear model to capture the change in values over time $p_j = \beta j + \alpha + \varepsilon_j$ with errors ε_j for each $j \in J$; we fit parameter values for $\hat{\alpha}$ and $\hat{\beta}$ to minimize the squares regression error
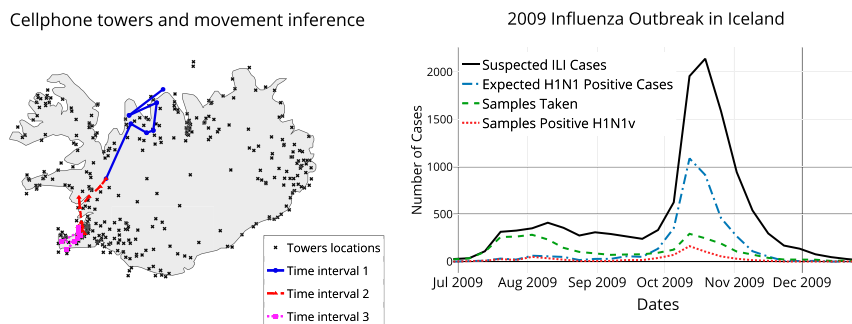

Fig. 1. Combining health records with call-data records. (*Left*) Cell towers act as a proxy for location, which, when coupled with the timestamp, allow movement inference. Different colors show inferred movements of a typical cell-phone user at different time periods over a period of 3 d. (*Right*) The epidemic curve for the 2009 H1N1 outbreak in Iceland, showing a single pronounced peak. The green dotted line shows the number of laboratory samples taken, the red line shows the number of those testing positive for H1N1v, and the black line shows the estimate of suspected H1N1v cases per week from the recorded ILI incidence (19). The expected H1N1v positive cases (blue dotted line) are extrapolated from the suspected ILI cases and the percentage of samples found positive each week.

Vigfusson et al.
Cell-phone traces reveal infection-associated behavioral change
www.manaraa.com

Table 1. Feature characteristics from the 29-d period around each individual's DoD (additional characteristics are in *SI Appendix*, Table S2)

Feature	Control Mean	Diagnosed Mean	Anomalous days
Number of towers visited	**3.04**	**2.74**	**−12, −2, −1, 0, 1, 2, 3, 4**
Number of new locations visited	0.5	0.43	1, 3
Unique contact count			
Incoming	2.25	2.02	∅
Outgoing	2.50	2.28	1, 2, 3, 6
Both	4.04	3.67	2
New contacts			
Incoming	0.61	0.50	∅
Outgoing	0.65	0.55	1, 3, 6
Both	1.19	1	0, 1, 3
Call duration, total, s			
Incoming	190	480.51	0, 1, 2
Outgoing	162	435.09	0
Both	479.5	915.60	∅
Calls count			
Incoming	3.10	2.84	∅
Outgoing	**3.60**	**3.37**	**1, 2**
Both	6.66	6.22	1
Texts count			
Incoming	2.71	2.87	−10
Outgoing	1.74	1.93	1, 2
Both	4.46	4.79	−10, 11
Calls and texts count			
Incoming	5.78	5.71	∅
Outgoing	5.34	5.30	1, 2
Both	11.12	11.01	2
Mean call duration			
Incoming	133.35	140.74	−1, 0, 1, 2, 4, 11
Outgoing	107.96	106.74	∅
Both	**143.14**	**149.84**	**−8, −1, 0, 1, 2, 3, 4, 11, 13, 14**
Top 3 contacts by duration (*SI Appendix, Sensitivity Analysis*)			
Incoming	0.68	0.69	∅
Outgoing	0.70	0.69	∅
Both	1.37	1.38	1
Remaining contacts by duration			
Incoming	1.57	1.33	∅
Outgoing	1.80	1.58	1, 2, 4, 6
Both	3.37	2.92	1, 4, 6
Top 3 contacts by frequency			
Incoming	0.58	0.56	∅
Outgoing	0.63	0.61	1, 2
Both	1.21	1.17	1
Remaining contacts by frequency			
Incoming	1.67	1.46	0
Outgoing	1.88	1.67	−9, 2, 6
Both	3.54	3.13	∅

The final column indicates which days relative to DoD show a statistically significant difference between the control and diagnosed groups. The boldface features are discussed in the main text.

$$\arg\min_{\hat{\alpha},\hat{\beta}} \sum_{j\in J} \varepsilon_j^2 = \arg\min_{\hat{\alpha},\hat{\beta}} \sum_{j\in J} \left(p_j - \beta j - \alpha\right)^2.$$

The detrended feature value, measuring the deviation from weekly routine, is then defined as

$$z_{fid} = x_{fid} - \hat{\beta}W - \hat{\alpha}.$$

Control Group. Each diagnosed individual was matched with a control individual from the undiagnosed group, based on home location. All measurements thus far have applied to individuals diagnosed with ILI during the epidemic. To compare the diagnosed population against a control population, a subset was selected from the rest of the data—those not diagnosed for ILI were assumed to be uninfected, though they may show behavior consistent with symptoms but are well, or have ILI symptoms but did not use health services. Of 74,644 people, we were able to identify home towers for 36,140. Each diagnosed person's control was selected randomly from the undiagnosed individuals among the 36,140 who shared a home tower with the diagnosed individual. For this dataset, control selection exhibited no noticeable differences across three methods: selecting randomly, matching for home tower, or matching home tower and frequency of calls (36).

We analyzed the pattern differences between the means of the detrended feature values (z_{fid}) of the individuals in the two groups. The 29-d range (2 wk either side of DoD) centered around every diagnosed individual's DoD range [−14, 14], with DoD mapping to zero. Controls used the

www.manaraa.com

same days of data as their diagnosed match. The average deviation from weekly routine on all days in the range was compared (*SI Appendix*, Fig. S9) with original feature values (x_{fid}), shown in *SI Appendix*, Figs. S2, S3, and S8.

Statistical Comparison. We compared the behavior of the diagnosed and control groups across each detrended feature value z_{fid} and each day using the Kolmogorov–Smirnov (KS) statistic. To counteract the increase in type I errors caused by running multiple significance tests, we used the Benjamini–Hochberg (BH) procedure to control the false discovery rate (FDR), as it presents the most conservative FDR correction for this mix; the adjusted *P* values can then be used to assess the evidence for or against the null hypothesis. The BH procedure assumes independent tests. Some tests act on dependent, interacting samples—e.g., a value on a specific day is ranked against values from the same day of the week for several weeks prior—whereas others are independent tests. Confidence bands for the KS test were computed and plotted for each day of the primary three features deemed significant based on the *P* values with the FDR correction (*SI Appendix*, Fig. S9). The significance test and the CI calculations use $\alpha = 0.05$.

Results

Several features show significant change between the routine behavior of the control and diagnosed populations around their DoD. The actual time period and magnitude of the behavioral change varies by feature (Table 1, rightmost column), but the number of towers visited, mean call duration, and the number of outgoing calls show the most pronounced signals of behavior change.

Less Movement. The number-of-towers feature indicates that the diagnosed group tends to travel less than usual, even before diagnosis. Such lower travel patterns coincide with the typical symptomatic period of influenza (37). The maximum effect is observed on the day following diagnosis, when diagnosed individuals travel to 1.1 to 1.4 fewer locations than normal. Differences are observed between the diagnosed and control groups from 2 d prior to the DoD until 4 d after DoD ($KS > 0.084$, $P < 3.2 \times 10^{-3}$; Fig. 2 and *SI Appendix*, Fig. S10). Other days in the 4-wk period display the diagnosed and control groups acting similarly.

Longer Calls. Mean call duration shows that people tend to make longer calls on average on the day after the DoD (Fig. 2), when significant differences are observed between the diagnosed and control groups ($KS = 0.155$, $P < 4.6 \times 10^{-10}$; *SI Appendix*, Fig. S22). On the day following diagnosis, diagnosed individuals spend an average of 41 to 66 s longer on the phone than usual.

Fewer Calls Placed. Number of outgoing calls gives another perspective of behavior following diagnosis. Although call duration increases around DoD, the number of outgoing calls decreases on the day after the DoD, with an average of 2.3 to 3.3 fewer calls than is routine ($KS = 0.102$, $P = 5.6 \times 10^{-4}$; *SI Appendix*, Fig. S18). On the day of diagnosis, diagnosed individuals increase outgoing calls relative to their routine compared to the days before and after.

Statistical significance through FDR-corrected *P* values is supported by KS CIs for nearly all comparisons (Fig. 2 and *SI Appendix*, Fig. S9). Notably, the diagnosed group displays significant changes in mobility, even prior to seeking healthcare and receiving a diagnosis (*SI Appendix*, *Visualization*).

Limitations. The results depend on the metadata arising from mobile-phone use, presenting both advantages and drawbacks (21, 33). The increased data bandwidth provided by MNOs and rapid device and app development over the past decade have altered user behavior patterns to communicate more via internet-based applications and less via calls and text. In our dataset, cellular internet data access (denoted general packet radio service [GPRS]) provided additional location information to CDR records of calls and texts, a situation that has likely shifted since the H1N1 outbreak (*SI Appendix, Comparing CDR and GPRS Data*). At 3 y following the epidemic, the Icelandic CDR and GPRS data contained a stronger location proxy than in 2009 due to more smartphone apps periodically connecting to cellular towers for Internet access, but poorer information for features pertaining to call duration, frequency, and top contacts.

Since many nations experience limited Internet access [53.6% of the world population in 2019 (22)] and smartphone availability [39.4% worldwide (38)], it would be reasonable to assume that call and text usage in these locations may follow similar patterns as in our dataset, but we caution against assuming all cell-phone behavior to be universal (33). Further, mobile-phone ownership may bias against those in greatest need of public health intervention. The results report aggregate behavior changes, which are likely to include patterns caused by other illnesses or injuries. Our approach depends on maintaining individual-level behavioral histories, since the signal we identified concerns departure from routine behavior rather than the actual behavior itself, as seen by comparing the raw and detrended distributions 6 d prior to diagnosis (*SI Appendix*, Fig. S2) with the day following the DoD (*SI Appendix*, Fig. S3). Finally, Iceland contains a small, mostly homogeneous, and generally affluent population bound to an island, with idiosyncratic behavior, including unusually high mobile-phone usage. Seasonal effects may be exaggerated in Icelanders compared to other populations due to Iceland's proximity to the Arctic.

Discussion

The combination of mobile-phone traces with health records reveals behavior change associated with symptom onset for
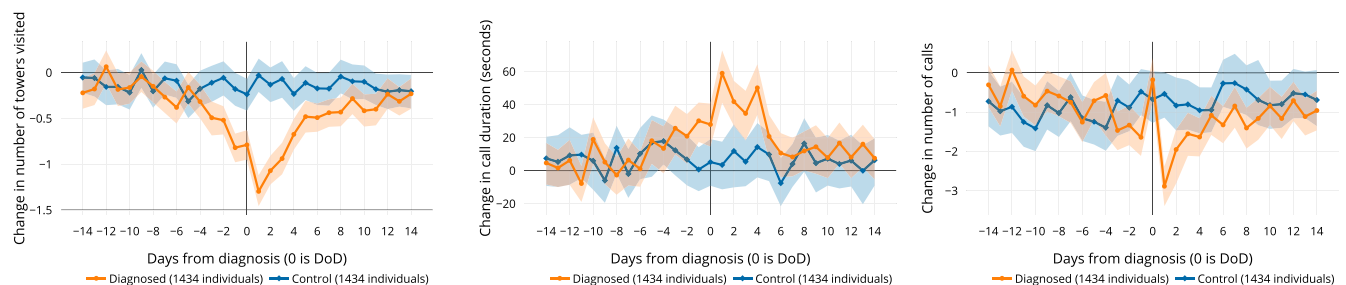

Fig. 2. Changes in average phone-use behavior associated with diagnosis. (*Left*) Users visit fewer locations on days around diagnosis. (*Center*) They make and receive longer phone calls on days near diagnosis. (*Right*) They initiate fewer calls on the days after diagnosis, with the exception of the day of diagnosis itself. Graphs display the mean deviation from "normal" routine behavior (z_{fid}) for each group on the relative day of illness determined by DoD (day 0). CIs (2.5 to 97.5%) are calculated using bootstrapping (*SI Appendix*, *Visualization*).
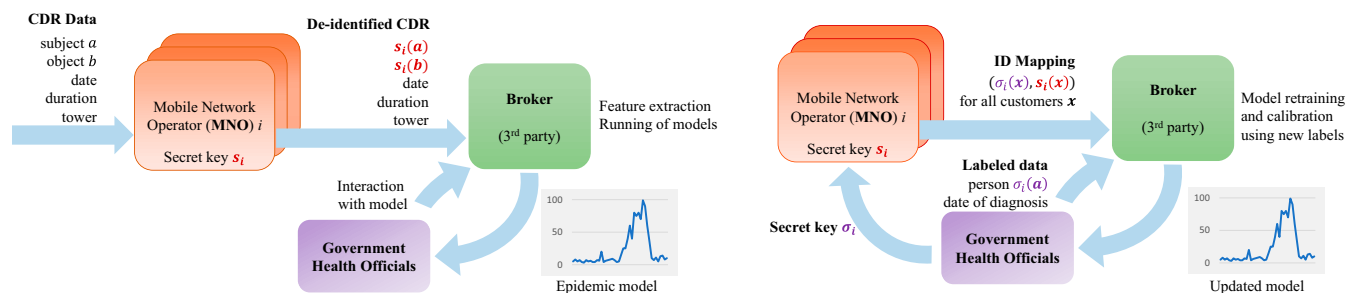
www.manaraa.com

Fig. 3. Privacy-preserving data-sharing protocol. Privacy-preserving architecture for syndromic surveillance using CDR data for future experimental design. An independent third-party broker is provided with real-time deidentified CDR data, extracts features, and runs the prediction models to generate an epidemic curve (*Left*; **O1**). The broker could also be provided labeled anonymous health information to join with the CDR data to calibrate or retrain the classifiers (*Right*; **O2**). The design accommodates mutual distrust, ensuring that health officials cannot monitor behavior or track mobility of individuals, that MNOs are not provided with any health information of customers, and that the broker only operates on deidentified data.

H1N1v in unprecedented detail. Observations of behavior in CDRs are consistent with our knowledge of influenza pathology: Individuals become infected and begin showing symptoms, which their behavior reflects; they then access healthcare, receive a diagnosis, and display activity patterns different from normal for a period, after which they return to normality. This picture depicts a group trend; however, in an effort to avoid ecological inference fallacy (39), we observe that individuals' changed behavior varied widely within a group. The variability of individuals' behavioral responses suggests that CDR data are best suited for aggregate analysis of symptomatic behavior.

Although we cannot know the exact cause in each individual case, collectively, the duration of anomalies is consistent with estimates of influenza symptom duration (40). The use case in Iceland demonstrates that disease-monitoring systems could be expanded with CDRs, already passively collected by local mobile operators, that can discern behavior consistent with ILI symptoms while following a protocol to preserve user privacy, and our approach provides a complementary way of estimating the duration of symptoms and, therefore, an important component for estimating the economic impact of an outbreak.

The results presented here have important implications for modeling disease dynamics. As individuals change behavior due to symptom onset, their potential to transmit is modified, yet modeling efforts that have been central to mitigation measures for novel pathogens tend to ignore behavioral effects, due largely to a dearth of quantitative information. Such limitation is evident in the case of modeling of SARS-CoV-2 transmission—for instance, where different groups vary in their ability to alter their behavior in response to exposure or illness (31, 41). Here, we quantify the direction and magnitude of the behavioral change effect for H1N1v on an atypical population that exhibits fewer sources of variability than most. Other pathogens and populations will have different properties that will require a context-specific investigation. Our work provides a methodology for capturing and quantifying behavior change that can be used to improve the predictive power of models in future outbreaks. We argue that such an approach would have an important part to play in outbreak response for novel pathogens.

A separation of access to private data is vital for ensuring public trust. While aggregation helps protect privacy (31), enabling health officials to interact with the data increases the risk to individual or group privacy. Concerns have been raised over government responses to COVID-19, where contacts of those infected are traced from historical CDR data (34). Our data-sharing protocol (Fig. 3 and *SI Appendix, Privacy-Preserving Data Sharing*) mitigates risk by ensuring that: 1) Mobile operators that hold cell-phone metadata do not have access to any new health information for their customers held by health officials; and 2) health officials do not access cell-phone metadata. To further strengthen the separation, differential privacy methods can be used to introduce controlled noise to the data in such a manner that aggregate statistics remain unchanged, while provably protecting the privacy of individuals and small groups (42, 43). At the same time, communicating the collective benefit of studies such as this one, and the effort taken to protect data, is necessary to help the public decide when the public health value of the information provided is worth the risk to their privacy.

Our results suggest that CDR metadata may allow surveillance of symptomatic diseases whose symptom intervals are sufficiently long and behavioral changes sufficiently pronounced that they produce a signal that is visible at the resolution afforded by the data. The granularity of these data is rapidly refining, both spatially, with denser tower infrastructure being built in response to population growth and newer generations of devices (e.g., 5G), and temporally, as mobile phones become increasingly used for Internet applications. Greater data resolution may help offset the relatively small effect sizes in our results, which are confounded by other brief interruptions to people's routines, and allow the approach to extend beyond a large-scale epidemic of a transmissible pathogen. Environments lacking health-monitoring infrastructure, but where mobile-phone use is prevalent and consistent (33), have the greatest potential gains from CDR-based epidemic surveillance. In particular, establishing the nature of symptomatic behavior provides an opportunity to use artificial intelligence to identify patterns suggesting that an individual or a group is symptomatic, and thus estimate the numbers of cases. We are optimistic that further study could establish the full generality and versatility of infectious disease surveillance using call-data records on their own.

Vigfusson et al.
Cell-phone traces reveal infection-associated behavioral change

PNAS | 5 of 6
https://doi.org/10.1073/pnas.2005241118

www.manaraa.com

1. World Health Organization, Coronavirus disease 2019 (COVID-19): Situation report (World Health Organization, Geneva, Switzerland, 2020), vol. 72.
2. World Health Organization, Ten threats to global health in 2019 (2019). https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019. Accessed 16 March 2020.
3. J. R. Ortiz et al., Strategy to enhance influenza surveillance worldwide. Emerg. Infect. Dis. 15, 1271 (2009).
4. Triple S Project, Assessment of syndromic surveillance in Europe. Lancet 378, 1833–1834 (2011).
5. M. Lipsitch, F. G. Hayden, B. J. Cowling, G. M. Leung, How to maintain surveillance for novel influenza A H1N1 when there are too many cases to count. Lancet 374, 1209–1211 (2009).
6. A. Sadilek, H. A. Kautz, V. Silenzio, "Predicting disease transmission from geo-tagged micro-blog data" in AAAI'12: Proceedings of the 26th AAAI Conference on Artificial Intelligence, J. Hoffmann, B. Selman, Eds. (AAAI, Palo Alto, CA, 2012), pp. 136–142.
7. A. Signorini, A. M. Segre, P. M. Polgreen, The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PloS One 6, e19467 (2011).
8. P. M. Polgreen, Y. Chen, D. M. Pennock, F. D. Nelson, R. A. Weinstein, Using internet searches for influenza surveillance. Clin. Infect. Dis. 47, 1443–1448 (2008).
9. J. Ginsberg et al., Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014 (2009).
10. C. Li et al., Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020. Euro Surveill. 25, 2000199 (2020).
11. S. Todd, P. J. Diggle, P. J. White, A. Fearne, J. M. Read, The spatiotemporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain. BMJ open 4, e004869 (2014).
12. D. Balcan et al., Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Natl. Acad. Sci. U.S.A. 106, 21484–21489 (2009).
13. M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns. Nature 453, 779–782 (2008).
14. M. Lajous et al., Mobile messaging as surveillance tool during pandemic (H1N1) 2009, Mexico. Emerg. Infect. Dis. 16, 1488–1489 (2010).
15. F. S. Lu et al., Accurate influenza monitoring and forecasting using novel internet data streams: A case study in the Boston metropolis. JMIR public health surveillance 4, e4 (2018).
16. C. E. Mills, J. M. Robins, M. Lipsitch, Transmissibility of 1918 pandemic influenza. Nature 432, 904–906 (2004).
17. G. J. Rubin, R. Amlôt, L. Page, S. Wessely, Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: Cross sectional telephone survey. BMJ 339, b2651 (2009).
18. M. Lipsitch et al., Managing and reducing uncertainty in an emerging influenza pandemic. N. Engl. J. Med. 361, 112–115 (2009).
19. G. Sigmundsdottir et al., Surveillance of influenza in Iceland during the 2009 pandemic. Euro Surveill. 15, 19742 (2010).
20. B. M. Althouse et al., Enhancing disease surveillance with novel data streams: Challenges and opportunities. EPJ Data Sci. 4, 17 (2015).
21. Office of National Statistics, "Statistical uses for mobile phone data: Literature review" (ONS Methodology Working Paper Series 8, Office of National Statistics, Newport, UK, 2019). https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseries-no8statisticalusesformobilephonedataliteraturereview. Accessed 7 March 2019.
22. International Telecommunication Union. Global and regional ICT data (2005-2019) (2019). https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx. Accessed 16 March 2020.
23. S. Asongu, The impact of mobile phone penetration on African inequality. Int. J. Soc. Econ. 42, 706–716 (2015).
24. N. Eagle, A. S. Pentland, D. Lazer, Inferring friendship network structure by using mobile phone data. Proc. Natl. Acad. Sci. U.S.A. 106, 15274–15278 (2009).
25. J. Blumenstock, G. Cadamuro, R. On, Predicting poverty and wealth from mobile phone metadata. Science 350, 1073–1076 (2015).
26. A. Wesolowski et al., Quantifying the impact of human mobility on malaria. Science 338, 267–270 (2012).
27. C. O Buckee, A. Wesolowski, N. N. Eagle, E. Hansen, R. W. Snow, Mobile phones and malaria: Modeling human and parasite travel. Trav. Med. Infect. Dis. 11, 15–22 (2013).
28. L. Bengtsson et al., Using mobile phone data to predict the spatial spread of cholera. Sci. Rep. 5, 8923 (2015).
29. M. Tizzoni et al., On the use of human mobility proxies for modeling epidemics. PLoS Comput. Biol., 10, e1003716 (2014).
30. C. C. Freifeld et al., Participatory epidemiology: Use of mobile phones for community-based health reporting. PLoS Med. 7, e1000376 (2010).
31. S. Y. Chang et al., Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. Nature 589, 82–87 (2020).
32. N. Ahmed et al., A survey of COVID-19 contact tracing apps. IEEE Access 8, 134577–134601 (2020).
33. S. L. Erikson. Cell phones ≠ self and other problems with big data detection and containment during epidemics. Med. Anthropol. Q. 32, 315–339 (2018).
34. D. M. Halbfinger, I. Kershner, R. Bergman. To track Coronavirus, Israel moves to tap secret trove of cellphone data. NY Times, 16 March 2020. https://www.nytimes.com/2020/03/16/world/middleeast/israel-coronavirus-cellphone-tracking.html. Accessed 11 January 2021.
35. A. Amato-Gauci et al., Surveillance trends of the 2009 influenza A (H1N1) pandemic in Europe. Euro Surveill. 16, 19903 (2011).
36. N. Kishore et al., Flying, phones and flu: Anonymized call records suggest that Keflavik International Airport introduced pandemic H1N1 into Iceland in 2009. Influ. other respiratory viruses 14, 37–45 (2020).
37. D. K. M. Ip et al., The dynamic relationship between clinical symptomatology and viral shedding in naturally acquired seasonal and pandemic influenza virus infections. Clin. Infect. Dis. 62, 431–437 (2016).
38. NewZoo, Global Mobile market report (2018). https://newzoo.com/insights/trend-reports/newzoo-global-mobile-market-report-2018-light-version/. Accessed 16 March 2019.
39. G. King et al., Ecological Inference: New Methodological Strategies (Cambridge University Press, New York, NY, 2004).
40. L. L. H. Lau et al., Viral shedding and clinical illness in naturally acquired influenza virus infections. J. Infect. Dis. 201, 1509–1516 (2010).
41. J. A. Patel et al., Poverty, inequality and COVID-19: The forgotten vulnerable. Publ. Health 183, 110–111 (2020).
42. Y. Cao, M. Yoshikawa, Y. Xiao, L. Xiong, "Quantifying differential privacy under temporal correlations" in Proceedings: 2017 IEEE 33rd International Conference on Data Engineering: ICDE 2017 (IEEE, Piscataway, NJ, 2017), pp. 821–832.
43. D. J. Mir, S. Isaacman, R. Cáceres, M. Martonosi, R. N. Wright, "DP-WHERE: Differentially private modeling of human mobility" in Proceedings: 2013 IEEE International Conference on Big Data, X. Hu, et al., Eds. (IEEE, Piscataway, NJ, 2013), pp. 580–588.

6 of 6 | PNAS
https://doi.org/10.1073/pnas.2005241118

Vigfusson et al.
Cell-phone traces reveal infection-associated behavioral change