Methods and Applications

Epidemic Surveillance of Influenza Infections: A Network-Free Strategy — Hong Kong Special Administrative Region, China, 2008–2011

Zhanwei Du[1,&]; Qi Tan[2,&]; Yuan Bai[1,2]; Lin Wang[3]; Benjamin J. Cowling[1,2]; Petter Holme[4,#]

ABSTRACT

Introduction: The ease of coronavirus disease 2019 (COVID-19) non-pharmacological interventions and the increased susceptibility during the past COVID-19 pandemic could be a precursor for the resurgence of influenza, potentially leading to a severe outbreak in the winter of 2022 and future seasons. The recent increased availability of data on Electronic Health Records (EHR) in public health systems, offers new opportunities to monitor individuals to mitigate outbreaks.

Methods: We introduced a new methodology to rank individuals for surveillance in temporal networks, which was more practical than the static networks. By targeting previously infected nodes, this method used readily available EHR data instead of the contact-network structure.

Results: We validated this method qualitatively in a real-world cohort study and evaluated our approach quantitatively by comparing it to other surveillance methods on three temporal and empirical networks. We found that, despite not explicitly exploiting the contacts' network structure, it remained the best or close to the best strategy. We related the performance of the method to the public health goals, the reproduction number of the disease, and the underlying temporal-network structure (e.g., burstiness).

Discussion: The proposed strategy of using historical records for sentinel surveillance selection can be taken as a practical and robust alternative without the knowledge of individual contact behaviors for public health policymakers.

Influenza infections were reported to be low between the months of September 2021 and January 2022 (*1*). The relaxation of coronavirus disease 2019 (COVID-19) non-pharmacological interventions and the increased susceptibility during the past COVID-19 pandemic have provided an opportunity for an increase of more severe influenza epidemics to occur in upcoming winters in temperate locations.

Infectious disease surveillance systems would provide historical information on the occurrence of infections and allow early detection of influenza outbreaks before they are past the point of being contained. The surveillance strategies that map out individual contact behaviors fall into two general categories — those based on static contact networks and those on temporal contact networks. Due to contact networks being essentially dynamic with temporal-network structures [such as burstiness — individual activities often happen in periods of intense activity (*2*)], the problem remains somewhat more practical in the context of temporal contact networks.

Retrospective studies have demonstrated that temporal network structures can influence the spreading speed and the outbreak size but also surveillance strategies (*3*–*5*). Cowling et al. compared two temporal-network strategies to select sentinels [sampling the recent contact, as the recent strategy, and most frequent contact, as the frequent strategy, with random individuals (*2*)], as well as two static-network strategies (acquaintance and random), on temporal networks for sentinel surveillance of outbreak detection (*3*). The two temporal-network strategies both derive earlier signals than static-network strategies for early epidemic detection on networks with strong temporal structures. However, due to physical contact data being difficult to obtain, these strategies are difficult to be applied to practical public health systems.

To detect an early signal for the emerging outbreak using sentinel surveillance, the digital data on Electronic Health Records (EHR) provide a unique chance to test cutting-edge sentinel surveillance strategies. The EHR of influenza viruses can help detect other viruses, and have temporal characteristics with records of when individuals were infected. Our previous study found treatment records can be used to

monitor emerging epidemic outbreaks (e.g., influenza) and proposed a simple EHR-based strategy that identifies the most vulnerable individuals who acquired the earliest infections during historical influenza seasons and could be a theoretically optimal surveillance subset (*6*). However, it does not account for the real-world data validation, the temporal contact networks (in which the contact structure may not be persistent enough), and the cross-strain immunity (which could be gained during an influenza season to protect the previously infected individuals from the reinfection of a group of strains).

In the current study, we produced a practical data-driven surveillance strategy by targeting previously infected nodes with low cross-strain immunity to accelerate outbreak detection using sentinel surveillance of previous earliest infected individuals. We validated this strategy with a real-world cohort study and further validated it by simulations using mathematical epidemic models in temporal networks. We quantified the early warning and Peak lag gained by these selected individuals over different transmission scenarios of effective reproduction numbers, R_es.

METHODS

Cohort Data

In the published dataset (*7*), serum specimens were collected from a cohort of participants from 2008 to 2011, with each annual record for 956 individuals. We identified an individual to be infected or not using a 4-fold criterion of hemagglutination inhibition (HAI) titers for each study year. Informed by the infection history, we could evaluate the infection probability of a case infected in a year based on the historical records, together with its 95% confidence interval, using the method of one-sample t-test.

Surveillance Strategies

We investigated three temporal-network strategies for designing network-based surveillance systems. To test the generality of our methodology, we compared it with available temporal-network surveillance strategies shown in Cowling et al. using three classes of temporal complex networks with distinct temporal features (Supplementary Material, available in http://weekly.chinacdc.cn/) (*3*).

Epidemic Model

We simulated epidemic outbreaks using a stochastic chain-binomial model in contact networks with nodes as individuals and edges as epidemiologically relevant contacts between individuals. We used the following two criteria to evaluate the performance of each surveillance strategy in the test epidemic season: Early warning, and Peak lag (Supplementary Material).

Proposed Surveillance Based on Individual Historical Vulnerability

In the static-network study, nodes with higher eigenvector centrality were the theoretical optimal surveillance subset (*8*). Conversely, nodes, which were infected earlier, tended to have a higher eigenvector centrality (*6*). We proposed a novel surveillance strategy to target previously infected nodes (History) based on estimating the vulnerability of each individual using their infection history of seasonal respiratory disease (e.g., influenza) registered in the EHR (Figure 1). Let τ_j^i the time at which an individual j acquires infection in season i. We defined the vulnerability of an individual j to contract a respiratory disease by their infection times in influenza season one year ago. This vector ranked each individual in the population by its historical vulnerability to getting infected, which we used as a surveillance strategy.

To be convenient to use by policymakers (e.g., doctors in hospitals), Figure 2 provides a schematic overview of the proposed surveillance strategy with four guidelines. Individuals who were uninfected in the current season but infected in the last season had higher rankings than others, which could have been higher if they had been infected earlier in the last season and had had more infection records in previous seasons.

RESULTS

Real-world Evaluation Using Historical Influenza Infections from a Cohort Study

Informed by a cohort of around 1,000 participants from 2008 to 2011, when influenza A(H1N1)pdm09 circulated, with an annual electronic record of influenza HAI titers (*7*), we identified an individual to be infected or not using a 4-fold criterion of HAI titers for each study year and estimated the infection probability of a case infected in the third year with/without infection during the past two years (Figure 1A and Supplementary Table S1, available in http://weekly.chinacdc.cn). The participants, who were

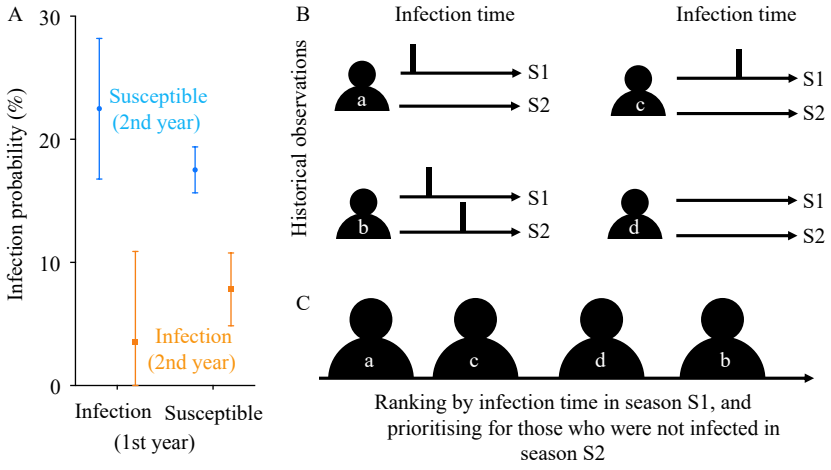

FIGURE 1. A schematic of the proposed surveillance strategy to target previously infected nodes (History). (A) Infection probability of influenza virus in Hong Kong Special Administrative Region. (B) Informed by available historical observations of individuals (a, b, c, and d) over two seasons from S1 to S2, each for one year. (C) In our proposed surveillance strategy, individuals are ranked by the infection time in season S1 for season S2.
Note: In panel A, we studied a cohort of 956 participants from 2008 to 2011 with annual electronic records for three years (2008 to 2010, and 2009 to 2011) (7). We estimated the probability of a case being infected in the third year, which was infected or not in the past two years (Supplementary Table S1, available in http://weekly.chinacdc.cn/). Vertical bars and error bars represent the estimated mean and 95% CIs. In panel B, the average historical vulnerability of an individual is estimated from the historical infection time. In panel C, the black bars denote the observed infection timing of individuals in the first and second/third historical seasons.
Abbreviation: S1=the first season; S2=the second season.

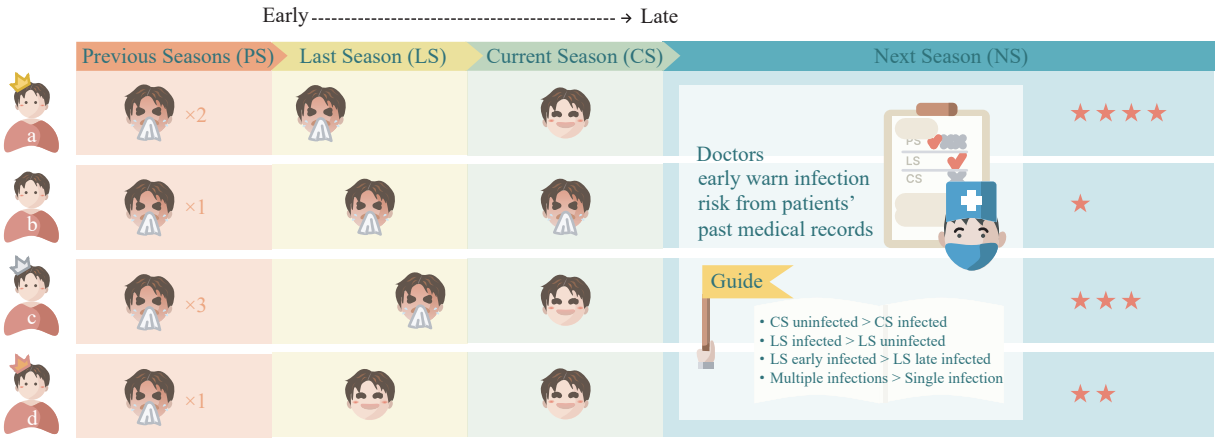

FIGURE 2. A simplified schematic illustration of the proposed surveillance strategy.
Note: The proposed strategy ranks the included individuals following four guidelines. Individuals who were uninfected in the current season but infected in the last season have higher rankings than others. And individuals can get a higher ranking if they were infected earlier in the last season and have more infection records in previous seasons. By taking four individuals (a, b, c, and d) as an example, users (e.g., doctors in hospitals) can assess their electronic records of historical seasons at the end of the current season. Individuals a and c have one infection record in the last season, none in the current season, and are ranked the highest. Given that a has been infected earlier than c, a has a higher ranking than c. If a and c have the same infection times in the last season, we could compare the number of infection records in previous seasons. Finally, the ranking of the four individuals is a, c, d, and b. We used golden/silver/copper crowns and red stars to mark their ranking from high to low.

infected in the first year but not in the second year, had the highest infection probability of 22.49% [95% confidence interval (CI): 16.78%, 28.20%] in the third year (Figure 1A). In the contact network structure, those participants may have more contact behaviors and low cross-strain immunity in the third year and thus have a higher probability of infection. And historical EHR data can help identify those with higher

infection risks of influenza outbreaks to some extent.

Simulation Evaluation of Surveillance Strategies

We compared our proposed strategy (History) (Figure 1) with three conventional temporal-network-based surveillance strategies, including 1) the recent strategy; 2) the frequent strategy, which uses the top 10% of individuals with the highest numbers of most recent and frequent connections, respectively; and 3) the random acquaintance strategy, which first randomly selects 10% of a randomly selected individual as the surveillance node (Figure 3).

In temporal networks, different from static networks, the experience period for strategy learning may not correlate with the evaluation period for testing, and individual interactions may happen in a short period. This may fail surveillance strategies if the high centrality nodes in the experience period decrease their centrality significantly or interactions in the evaluation period. Thus, we investigate the impact of burstiness [the phenomenon that human activities often happen in intense periods separated by periods of inactivity (*2*)] and persistence [measured by the fraction of edges that is present both in the first and last 5% of the contacts by the Jaccard similarity coefficient (*2*)] on the performance of early warning and Peak lags (Supplementary Table S2, available in http://weekly.chinacdc.cn). We collected and summarized the properties of the study temporal networks in Supplementary Table S2, with burstiness ranging from 0.39 to 0.72 and persistence ranging from 0 to 0.18.

Burstiness may correlate with the early warning negatively (Supplementary Table S3, available in http://weekly.chinacdc.cn/). In the evaluation criteria for early warning, the proposed strategy outstrips the random strategy in all three networks. The performance of our strategy is also comparable to the frequent and recent strategies in the Prostitution and Email networks, but not in the Dating network (which has the highest burstiness) (Figure 4 and Supplementary Table S3). For example, in the Prostitution network, under a low transmission scenario (R_e=1.2) and a high transmission scenario (R_e=3), the history strategy has an early warning of 1.78 days (95% CI: 0.00, 3.56) and 0.71 days (95% CI: −0.54, 1.95), respectively, which is around a half day earlier than the recent and frequent strategies. In the Dating network, the history strategy is better than the random strategy but not the other two. As for Peak lag, there is no significant difference among study strategies in the three temporal networks, given the large 95% CI (Figure 4 and Supplementary Table S4, available in http://weekly.chinacdc.cn).

DISCUSSION

Although influenza outbreaks are extremely challenging to predict (*9*), alerting the onset of an influenza pandemic would be extremely important for public health agencies to respond before it pasts the point of being contained.

Building on the availability of electronic health record systems, we proposed a novel surveillance strategy of selecting previously infected individuals for sentinel surveillance in temporal networks. The beauty of this approach is to exploit contact structure without having to measure it — which is both difficult and may change over time, inspired by that history of infection will reflect temporal contact structure if it is persistent enough. In practical use, for a new emerging or reemerging infectious disease, it is not necessary for our proposed strategy to ask for patient infection history or prior knowledge of the same disease.

The proposed strategy does not work well in the Dating network, perhaps due to its high bursty coefficient (Supplementary Table S2). Individuals just have intensive contact for a short period and hinder the subsequent infections. For instance, one individual who has intensive contacts in the training seasons has a higher probability of getting infected than other individuals, and thereby is ranked higher in our proposed history strategy. However, this individual may make few contacts in the test season and has a lower probability to be infected and fail the history strategy.

EHRs are the most essential component of health information technology, with hospital adoption rates representing a country's level of medical digitalization. In China, it was 83.6% and 86.6% in 2018 in China's economically underdeveloped and developed hospitals, respectively (*10*). Although the growth rate of EHR adoption has been slow since 2013, as a result of many hospitals in underdeveloped areas lacking sufficient financial support and staffing funds, the Chinese government has begun to reform relevant policies, implementing both financial assistance and policy-guiding measures gradually (*10*). With the implementation of EHR systems in underdeveloped

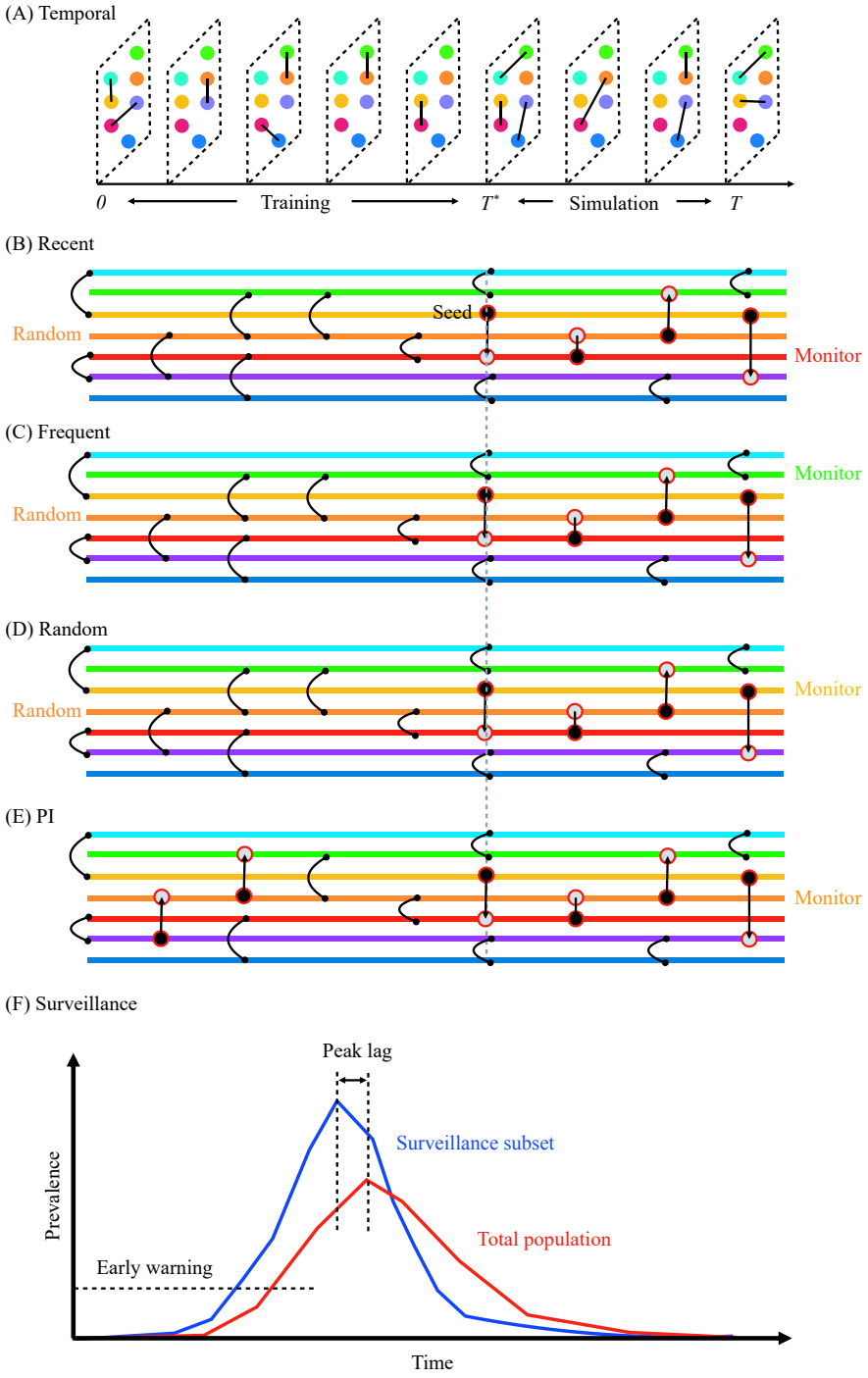

FIGURE 3. Schematic illustration of surveillance strategies. (A) The example of a temporal network with two time phases. (B) The schematic illustration of the Recent surveillance strategy. (C) The schematic illustration of the Frequent surveillance strategy. (D) The schematic illustration of the Random surveillance strategy. (E) The schematic illustration of the history surveillance strategy for previously infected (PI). (F) Surveillance objectives.
Note: In panel A, the first time phase is for training and the second one is for epidemic simulation. In panel B, C, D, and E, the horizontal line denotes an individual and the circles and vertical lines indicate the interaction. We marked the first infected node as "Seed" in the second time window, the node selected randomly to trigger surveillance strategy as "Random", the node for sentinel surveillance as "Monitor" following different strategies. In epidemic simulation, gray and black circles with red borders denote infectors and infectees, respectively. In panel F, we compared the prevalences between nodes in the surveillance subset and those in the whole population. We calculated the time lag between the two groups reaching 1% prevalence (early warning) and their epidemic peaks (peak lag).
* means the end time of the training phase.

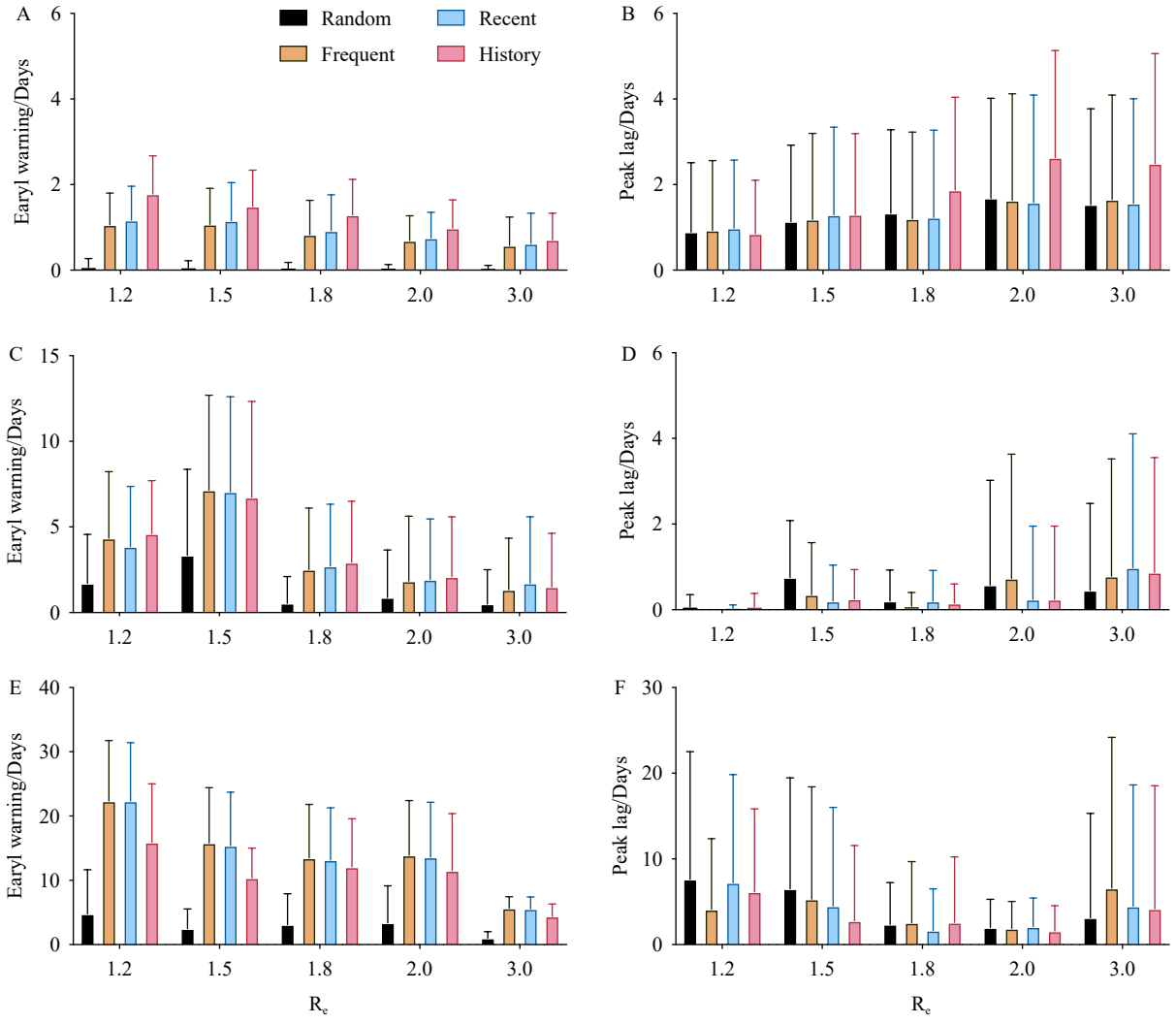

FIGURE 4. Early warning and Peak lag of the random, recent, frequent, and history strategies. (A) Early warning in the Prostitution network. (B) Peak lag in the Prostitution network. (C) Early warning in the Email network. (D) Peak lag in the Email network. (E) Early warning in the Dating network. (F) Peak lag in the Dating network.
Note: The history strategy here uses the EHR records obtained from historical seasons two years ago. The horizontal and vertical axes present the early warning (days) in the left panels and Peak lags (days) in the right panels for each strategy over effective reproduction numbers (R_e). Bars and error bars indicate the mean and standard deviations across 100 simulations of each temporal network. The burstiness of Prostitution, Email, and Dating are 0.39, 0.62, 0.72 (Supplementary Table S2, available in http://weekly.chinacdc.cn).
Abbreviation: EHR=Electronic Health Record.

and developed areas in near future, we expect our EHR-based strategy to become increasingly widely used in China.

We concluded that the proposed strategy of using historical records for sentinel surveillance selection is competitive with other existing surveillance strategies in temporal networks and can be taken as a practical and robust alternative without the knowledge of individual contact behaviors for public health policymakers. This study can deepen the understanding of sentinel surveillance and guide future strategies with diverse data sources, especially digital health data.

Conflicts of interest: BJC consults for AstraZeneca, Fosun Pharma, GlaxoSmithKline, Moderna, Pfizer, Roche and Sanofi Pasteur. BJC is supported by the AIR@innoHK program of the Innovation and Technology Commission of the Hong Kong SAR Government. No other conflicts of interest.

Funding: Supported by Key Projects of Intergovernmental International Scientific and Technological Innovation Cooperation of National Key R&D Programs (No. 2022YFE0112300) and AIR@InnoHK administered by Innovation and

Technology Commission of the Research Grants Council of the Hong Kong SAR Government.

doi: 10.46234/ccdcw2022.207

[#] Corresponding author: Petter Holme, petter.holme@aalto.fi.

[1] WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; [2] Laboratory of Data Discovery for Health Limited (D24H), Hong Kong Science and Technology Park, Hong Kong SAR, China; [3] Department of Genetics, University of Cambridge, Cambridge, UK; [4] Department of Computer Science, Aalto University, Espoo, Finland.
[&] Joint first authors.

Submitted: November 11, 2022; Accepted: November 14, 2022

REFERENCES

1. World Health Organization. Recommended composition of influenza virus vaccines for use in the 2022–2023 northern hemisphere influenza season. 2022. https://www.who.int/publications/m/item/recommended-composition-of-influenza-virus-vaccines-for-use-in-the-2022-2023-northern-hemisphere-influenza-season. [2022-10-15].
2. Lee S, Rocha LEC, Liljeros F, Holme P. Exploiting temporal network structures of human interaction to effectively immunize populations. PLoS One 2012;7(5):e36439. http://dx.doi.org/10.1371/journal.pone.0036439.
3. Bai Y, Yang B, Lin LJ, Herrera JL, Du ZW, Holme P. Optimizing sentinel surveillance in temporal network epidemiology. Sci Rep 2017;7(1):4804. http://dx.doi.org/10.1038/s41598-017-03868-6.
4. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature 2005;438(7066):355 – 9. http://dx.doi.org/10.1038/nature04153.
5. Gao C, Zhu JY, Zhang F, Wang Z, Li XL. A novel representation learning for dynamic graphs based on graph convolutional networks. IEEE Trans Cybern 2022. http://dx.doi.org/10.1109/TCYB.2022.3159661.
6. Du ZW, Bai Y, Wang L, Herrera-Diestra JL, Yuan ZL, Guo RZ, et al. Optimizing COVID-19 surveillance using historical electronic health records of influenza infections. PNAS Nexus 2022;1(2):pgac038. http://dx.doi.org/10.1093/pnasnexus/pgac038.
7. Tsang TK, Perera RAPM, Fang VJ, Wong JY, Shiu EY, So HC, et al. Reconstructing antibody dynamics to estimate the risk of influenza virus infection. Nat Commun 2022;13(1):1557. http://dx.doi.org/10.1038/s41467-022-29310-8.
8. Herrera JL, Srinivasan R, Brownstein JS, Galvani AP, Meyers LA. Disease surveillance on complex social networks. PLoS Comput Biol 2016;12(7):e1004928. http://dx.doi.org/10.1371/journal.pcbi.1004928.
9. World Health Organization. Influenza: are we ready? World Health Organization. https://www.who.int/news-room/spotlight/influenza-are-we-ready. [2022-10-15].
10. Liang J, Li Y, Zhang ZA, Shen DX, Xu J, Zheng X, et al. Adoption of Electronic Health Records (EHRs) in China during the past 10 years: consecutive survey data analysis and comparison of Sino-American challenges and experiences. J Med Internet Res 2021;23(2):e24813. http://dx.doi.org/10.2196/24813.

SUPPLEMENTARY MATERIAL

Methods

Surveillance strategies: We investigated three temporal-network strategies for designing network-based surveillance systems. In a temporal network, we divide the time duration of individual contacts into two sequential periods: the experience period and the evaluation period. The reference strategies used for surveillance are:

- **Random**: target an individual randomly;
- **Recent**: target the most recent contact individual of random individuals in the experience period;
- **Frequent**: target the most frequent contact individual of random individuals in the experience period.

Contact network data sets: To test the generality of our methodology, we also compared it with available temporal-network surveillance strategies shown in Bai et al. (*1*), using three classes of temporal complex networks with distinct temporal features:

- **Prostitution**: A sexual network, collected from a Brazilian online forum, includes N=16,730 individuals and E=50,632 sexual contacts over T=2,232 days. The contact at a given time represents the reported sexual encounter between a male sex buyer and a female escort (*2*–*3*).
- **Email**: A communication network, collected from one of the main mail servers of a university (*4*), includes N=2,997 email accounts and E=202,694 contacts over T=82 days. The contact at time t represents the interaction of two email accounts by sending or receiving a message (*2*,*4*).
- **Dating**: A sexual network, collected from an Internet dating community (*5*), includes N=28,972 individuals as well as E=430,827 sexual contacts over a span of T=512 days. Each contact demonstrates the initial dating interaction between two individuals (*2*,*5*).

The degree distributions of all three temporal networks follow a power law (*1*). Additionally, they show different degrees of burstiness (the phenomenon that human activities often happen intensively in short periods separated by periods of quiescence (*2*)] and persistence [the fraction of edges that is present both in the first and last 5% of the contacts by the Jaccard similarity coefficient (*2*)].

The three temporal networks show high values of burstiness with Prostitution at 0.39, Dating at 0.62, and Email at 0.72. At the same time, they possess different values of persistence, with Prostitution and Dating as 0, and Email as 0.18 (*1*).

The time scales of the temporal contacts in the above three datasets are greatly different. The mean number of contacts per individual in 7 days in prostitution, email, and dating datasets are 0.8, 18.4, and 16.1, respectively. We rescaled the time scales of the temporal contacts in these datasets to reduce the difference in physical contacts that

SUPPLEMENTARY TABLE S1. Infection probability of influenza virus in Hong Kong Special Administrative Region.

Infected in first year	Infected in second year	Mean infection probability in third year	Number of cases
Yes	No	22.49% (95% CI: 16.78%, 28.20%)	209
No		17.53% (95% CI: 15.66%, 19.39%)	1,592
Yes	Yes	3.57% (95% CI: 0.00%, 10.90%)	28
No		7.81% (95% CI: 4.86%, 10.77%)	320

Note: We estimated from a cohort of 956 participants from 2008 to 2011 with annual electronic records (*1*). We separated the four-year dataset into two three-year sub-datasets of 2008 to 2010 and 2009 to 2011. Cases were selected following different criteria from the two sub-datasets and aggregated together to estimate the infection probability in the third year. We estimated the probability of a case infected in the third year, which was infected or not in the past two years.
Abbreviation: CI=confidence interval.

SUPPLEMENTARY TABLE S2. Properties of temporal networks.

Dataset	Number of nodes	Number of edges	Time steps	Time unites	Average degree	Burstiness	Persistence
Prostitution	16,730	50,632	2,232 days	1 day	15	0.39	0
Email	2,997	202,694	82 days	1 second	15	0.62	0.18
Dating	28,972	430,827	512 days	1 second	8	0.72	0

SUPPLEMENTARY TABLE S3. Early warning (days) of optimal testing strategies for various influenza transmission scenarios.

R	Network	Random Mean (95% CI)	Frequent Mean (95% CI)	Recent Mean (95% CI)	History Mean (95% CI)
1.2	Prostitution	0.06 (−0.35, 0.47)	1.06 (−0.43, 2.54)	1.17 (−0.41, 2.75)	1.78 (0.00, 3.56)
1.5		0.05 (−0.28, 0.39)	1.07 (−0.61, 2.76)	1.16 (−0.61, 2.94)	1.49 (−0.21, 3.18)
1.8		0.04 (−0.24, 0.33)	0.83 (−0.77, 2.42)	0.92 (−0.77, 2.61)	1.29 (−0.37, 2.94)
2		0.03 (−0.17, 0.24)	0.69 (−0.47, 1.85)	0.75 (−0.45, 1.95)	0.98 (−0.35, 2.31)
3		0.02 (−0.16, 0.21)	0.58 (−0.73, 1.89)	0.62 (−0.79, 2.04)	0.71 (−0.54, 1.95)
1.2	Email	1.71 (−4.01, 7.43)	4.35 (−3.40, 12.11)	3.86 (−3.15, 10.87)	4.60 (−1.58, 10.77)
1.5		3.34 (−6.71, 13.40)	7.15 (−3.96, 18.26)	7.05 (−4.08, 18.18)	6.72 (−4.47, 17.91)
1.8		0.54 (−2.60, 3.68)	2.52 (−4.66, 9.69)	2.72 (−4.50, 9.95)	2.92 (−4.22, 10.05)
2		0.88 (−4.65, 6.42)	1.84 (−5.74, 9.41)	1.93 (−5.14, 9.00)	2.08 (−4.91, 9.07)
3		0.51 (−3.47, 4.50)	1.34 (−4.68, 7.35)	1.72 (−6.04, 9.48)	1.49 (−4.77, 7.74)
1.2	Dating	4.79 (−8.91, 18.49)	22.35 (3.52, 41.17)	22.35 (4.21, 40.49)	15.91 (−2.26, 34.08)
1.5		2.48 (−3.68, 8.65)	15.81 (−1.39, 33.01)	15.42 (−1.15, 31.99)	10.39 (1.14, 19.64)
1.8		3.10 (−6.52, 12.72)	13.49 (−3.14, 30.12)	13.17 (−3.05, 29.40)	12.07 (−2.96, 27.11)
2		3.38 (−8.14, 14.89)	13.90 (−3.17, 30.97)	13.60 (−3.51, 30.70)	11.52 (−6.18, 29.22)
3		1.01 (−1.01, 3.03)	5.69 (2.18, 9.21)	5.59 (1.98, 9.21)	4.43 (0.64, 8.22)

Abbreviation: CI=confidence interval.

SUPPLEMENTARY TABLE S4. Peak time (days) of optimal testing strategies for a range of influenza transmission scenarios.

R	Network	Random Mean (95% CI)	Frequent Mean (95% CI)	Recent Mean (95% CI)	History Mean (95% CI)
1.2	Prostitution	0.89 (−2.35, 4.12)	0.93 (−2.32, 4.18)	0.98 (−2.19, 4.16)	0.85 (−1.66, 3.35)
1.5		1.14 (−2.42, 4.70)	1.19 (−2.84, 5.22)	1.29 (−2.81, 5.38)	1.30 (−2.48, 5.08)
1.8		1.33 (−2.57, 5.22)	1.20 (−2.86, 5.26)	1.23 (−2.84, 5.31)	1.87 (−2.47, 6.21)
2		1.68 (−3.00, 6.36)	1.63 (−3.35, 6.61)	1.58 (−3.45, 6.61)	2.63 (−2.37, 7.62)
3		1.53 (−2.96, 6.02)	1.65 (−3.22, 6.52)	1.56 (−3.34, 6.46)	2.49 (−2.65, 7.62)
1.2	Email	0.06 (−0.53, 0.65)	0.00 (0.00, 0.00)	0.01 (−0.18, 0.20)	0.06 (−0.58, 0.69)
1.5		0.75 (−1.92, 3.42)	0.35 (−2.10, 2.79)	0.20 (−1.48, 1.89)	0.25 (−1.14, 1.64)
1.8		0.20 (−1.26, 1.65)	0.08 (−0.55, 0.72)	0.20 (−1.25, 1.64)	0.15 (−0.74, 1.04)
2		0.58 (−4.31, 5.46)	0.73 (−5.07, 6.53)	0.24 (−3.18, 3.66)	0.24 (−3.18, 3.66)
3		0.45 (−3.61, 4.51)	0.78 (−4.70, 6.26)	0.98 (−5.27, 7.24)	0.87 (−4.49, 6.22)
1.2	Dating	7.61 (−22.17, 37.38)	4.12 (−12.37, 20.60)	7.23 (−17.98, 32.43)	6.17 (−13.17, 25.50)
1.5		6.50 (−19.41, 32.41)	5.31 (−20.90, 31.53)	4.51 (−18.49, 27.52)	2.78 (−14.79, 20.34)
1.8		2.37 (−7.37, 12.11)	2.56 (−11.69, 16.80)	1.67 (−8.00, 11.34)	2.58 (−12.72, 17.87)
2		1.99 (−4.61, 8.59)	1.90 (−4.39, 8.18)	2.09 (−4.59, 8.76)	1.58 (−4.33, 7.49)
3		3.15 (−21.15, 27.45)	6.60 (−28.58, 41.79)	4.48 (−23.82, 32.78)	4.18 (−24.55, 32.90)

Abbreviation: CI=confidence interval.

the infectious virus transmits on (*6*). Specifically, we sped up the time scale of the prostitution dataset 40-fold and slowed down the time scales of the email and dating by mapping the one-day time scale in the email dataset and the two-day time scale in the dating dataset to seven-day, respectively. Finally, the mean estimates are closer among prostitution, email, and dating datasets, which are 2.97, 3.97, and 9.06 days on average in 7 days.

SUPPLEMENTARY TABLE S5. Values of parameters in the SIRS epidemic model in networks across effective reproduction numbers.

Network	Effective reproduction number R_e	Transmission probability β	Recover rate γ (days)
Prostitution	1.2	0.3	7
	1.5	0.4	7
	1.8	0.5	7
	2	0.7	7
	3	0.9	7
Email	1.2	0.15	7
	1.5	0.3	7
	1.8	0.5	7
	2	0.6	7
	3	0.8	7
Dating	1.2	0.2	7
	1.5	0.3	7
	1.8	0.35	7
	2	0.4	7
	3	0.8	7

Abbreviation: SIRS=susceptible-infectious-recovered-susceptible.

Epidemic Model

We simulated epidemic outbreaks using a stochastic chain-binomial model in contact networks with nodes as individuals and edges as epidemiologically relevant contacts between individuals.

We simulated the susceptible-infectious-recovered-susceptible (SIRS) model: each individual has three states: susceptible (S), infectious (I), or recovered (R). The transmission probability of the disease in each contact is β. The node i will remain infectious for $1/\gamma$ days, after which it will recover. For each dataset, we simulate multiple instances of β and γ. We set γ to 7 days as the base recovery period (*7*). The γ is calibrated to match the expected R_e. The values of β and γ used in the experiments are listed in Supplementary Table S5. We estimated R_e by the secondary infections of the earliest infected cases, which count for 1% of the population. We started simulations in temporal networks by randomly sampling one seed to be infectious.

The disease prevalence is counted as the number of infectious people over time. We divided the temporal contacts into η=4 phases, with each phase as one season. In each phase, we randomly selected one individual as the source of infection. To simulate the cross-immunity, we asserted that the individual infected in the previous season will not be infected in the current season, as that the immunity period of influenza had been over one year (*8*–*9*). We used the last season as the test epidemic season and used the contact information data in the previous η-1 seasons for training surveillance strategies.

Following (*10*–*12*), we used the following two criteria to evaluate the performance of each surveillance strategy in the test epidemic season:

● Early warning. Let t_μ^{EP} be the time at which the disease prevalence reaches a predefined threshold μ in the entire population (EP), and t_μ^{SG} the time at which the disease prevalence reaches the same threshold μ=1% in the surveillance group (SG). The early warning criterion measures the time lag $t_\mu^{EP} - t_\mu^{SG}$.

● Peak lag. Let t_{peak}^{EP}, t_{peak}^{SG} be the time at which the disease prevalence reaches the peak prevalence in the entire population and in the surveillance group, respectively. Peak lag criterion measures the time lag $t_{\text{peak}}^{EP} - t_{\text{peak}}^{SG}$.

Limitations

We would highlight several limitations in our study. First, we assumed the consequent outbreaks have similar circulating strains over seasons in our simulations, and individuals infected last season have immunity protection against infection in the consequent season. However, if the circulating strains are different during two consecutive

influenza seasons, there is weak cross-immunity protection. In this situation, the previously infected individuals in the last outbreak may be chosen directly for the sentinel surveillance of the studied outbreak. Second, the lack of influenza circulation in the past seasons during 2020 and 2021 would reduce the immunity level in the population level (*13*). For the sentinel surveillance of outbreaks in the winter of 2022, we suggest the selection of individuals based on EHR data before the COVID-19 pandemic, for example, in the winter of 2019. Third, we did not model the influenza vaccination in our simulations significantly. However, the effective reproduction number in our simulations has included the impact of vaccinations on transmissions, with higher vaccine coverage resulting in lower effective reproduction numbers. To reduce this potential bias in practice, we suggest excluding those individuals vaccinated in the near period in the sentinel surveillance set. Fourth, it is plausible that not all infected individuals will seek treatment and leave their records. According to our prior study (*14*), the EHR-based strategy did not work well when the probability of seeking treatment and having an influenza record does not exceed 25%, which we believe also holds in the temporal network qualitatively. Fifth, in the study of temporal networks, we mainly considered the individual contacts. Environmental variables may have an inevitable effect in determining whether a pathogen can become epidemic, for example, absolute humidity and temperature on influenza (*15*–*16*). We may include these additional effects on the transmission rate in the future for disease modeling of specific infectious diseases. Sixth, we divided the time duration of temporal networks into the experience and evaluation periods. However, human movements have seasonal patterns (*17*), which may enhance the performance of our proposed strategy. For the example of winter influenza, we may only use the history of winter influenza for the proposed strategy to learn, but not the overall annual history of influenza infection. Seventh, to validate our assumption of the impact of cross-immunity on early surveillance, we conducted a real-world evaluation using historical influenza infections from a cohort study. Although the group following the EHR-based strategy had a higher relative infection probability than other groups, influenza transmission is complex and has age-specific heterogeneity, which is not included in our modeling. However, we believe our result still holds qualitatively.

REFERENCES

1. Bai Y, Yang B, Lin LJ, Herrera JL, Du ZW, Holme P. Optimizing sentinel surveillance in temporal network epidemiology. Sci Rep 2017;7(1):4804. http://dx.doi.org/10.1038/s41598-017-03868-6.
2. Lee S, Rocha LEC, Liljeros F, Holme P. Exploiting temporal network structures of human interaction to effectively immunize populations. PLoS One 2012;7(5):e36439. http://dx.doi.org/10.1371/journal.pone.0036439.
3. Rocha LEC, Liljeros F, Holme P. Information dynamics shape the sexual networks of Internet-mediated prostitution. Proc Natl Acad Sci USA 2010;107(13):5706 – 11. http://dx.doi.org/10.1073/pnas.0914080107.
4. Eckmann JP, Moses E, Sergi D. Entropy of dialogues creates coherent structures in e-mail traffic. Proc Natl Acad Sci USA 2004;101(40):14333 – 7. http://dx.doi.org/10.1073/pnas.0405728101.
5. Holme P, Edling CR, Liljeros F. Structure and time evolution of an Internet dating community. Soc Networks 2004;26(2):155 – 74. http://dx.doi.org/10.1016/j.socnet.2004.01.007.
6. Feehan DM, Mahmud AS. Quantifying population contact patterns in the United States during the COVID-19 pandemic. Nat Commun 2021;12(1):893. http://dx.doi.org/10.1038/s41467-021-20990-2.
7. CDC. Key Facts About Influenza (Flu). Centers for Disease Control and Prevention. 2022. https://www.cdc.gov/flu/about/keyfacts.htm. [2022-11-3].
8. Patel MM, York IA, Monto AS, Thompson MG, Fry AM. Immune-mediated attenuation of influenza illness after infection: opportunities and challenges. Lancet Microbe 2021;2(12):e715 – 25. http://dx.doi.org/10.1016/S2666-5247(21)00180-4.
9. Pei S, Teng X, Lewis P, Shaman J. Optimizing respiratory virus surveillance networks using uncertainty propagation. Nat Commun 2021;12(1):222. http://dx.doi.org/10.1038/s41467-020-20399-3.
10. Herrera JL, Srinivasan R, Brownstein JS, Galvani AP, Meyers LA. Disease surveillance on complex social networks. PLoS Comput Biol 2016;12(7):e1004928. http://dx.doi.org/10.1371/journal.pcbi.1004928.
11. Thompson WW, Comanor L, Shay DK. Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease. J Infect Dis 2006;194(S2):S82 – 91. http://dx.doi.org/10.1086/507558.
12. Rath TM, Carreras M, Sebastiani P. Automated detection of influenza epidemics with hidden Markov models. In: Proceedings of the 5th international symposium on intelligent data analysis on advances in intelligent data analysis V. Berlin: Springer. 2003521 – 32. http://dx.doi.org/10.1007/978-3-540-45231-7_48.
13. Baker RE, Park SW, Yang WC, Vecchi GA, Metcalf CJE, Grenfell BT. The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. Proc Natl Acad Sci USA 2020;117(48):30547 – 53. http://dx.doi.org/10.1073/pnas.2013182117.
14. Du ZW, Bai Y, Wang L, Herrera-Diestra JL, Yuan ZL, Guo RZ, et al. Optimizing COVID-19 surveillance using historical electronic health records of influenza infections. PNAS Nexus 2022;1(2):pgac038. http://dx.doi.org/10.1093/pnasnexus/pgac038.
15. Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci USA 2009;106(9):3243 – 8. http://dx.doi.org/10.1073/pnas.0806852106.
16. Deyle ER, Maher MC, Hernandez RD, Basu S, Sugihara G. Global environmental drivers of influenza. Proc Natl Acad Sci USA 2016;113(46):13081 – 6. http://dx.doi.org/10.1073/pnas.1607747113.
17. Wesolowski A, Zu Erbach-Schoenberg E, Tatem AJ, Lourenço C, Viboud C, Charu V, et al. Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. Nat Commun 2017;8(1):2069. http://dx.doi.org/10.1038/s41467-017-02064-4.