Available online at www.sciencedirect.com

ScienceDirect

Biomedical Journal

journal homepage: www.elsevier.com/locate/bj

News and Perspectives

Development of a semi-structured, multifaceted, computer-aided questionnaire for outbreak investigation: e-Outbreak Platform

Wei J. Chen [a,b,c,*], Shih-Yi Yang [d], Jung-Chih Chang [e], Wei-Che Cheng [d], Tzu-Pin Lu [b,c], Yu-Neng Wang [f], Ming-Hao Juan [f], Ruey-Tzer Hsu [f], Song-Ren Huang [g], Jia-Jang Tu [h], Pang-Chieh Wang [h], Vincent W.-S. Feng [h], Pei-Zen Chang [i]

[a] Center for Neuropsychiatric Research, National Health Research Institutes, Zhunan, Miaoli, Taiwan
[b] Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan
[c] Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan
[d] Public Sector, Microsoft Taiwan, Taipei, Taiwan
[e] Solutions Technical Unit, Microsoft Taiwan, Taipei, Taiwan
[f] Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[g] Department of Computer Science & Information Engineering, National Taiwan University, Taipei, Taiwan
[h] Computational Intelligence Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan
[i] Industrial Technology Research Institute, Hsinchu, Taiwan

ARTICLE INFO

Article history:
Received 10 May 2020
Accepted 16 June 2020
Available online 20 June 2020

Keywords:
Outbreak
Epidemiological investigation
Platform
Knowledge graph
COVID-19
SARS-CoV-2

ABSTRACT

Aggressive tracing of contacts of confirmed cases is crucial to Taiwan's successful control of the early spread of COVID-19. As the pandemic lingers, an epidemiological investigation that can be conducted efficiently in a timely manner can help decrease the burden on the health personnel and increase the usefulness of such information in decision making. To develop a new tool that can improve the current practice of epidemiological investigation by incorporating new technologies in digital platform and knowledge graphs. To meet the various needs of the epidemiological investigation, we decided to develop an *e-Outbreak Platform* that provides a semi-structured, multifaceted, computer-aided questionnaire for outbreak investigation. There are three major parts of the platform: (1) a graphic portal that allows users to have an at-glance grasp of the functions provided by the platform and then choose the one they need; (2) disease-specific questionnaires that can accommodate different formats of the information, including text typing, button selection, and pull-down menu; and (3) functions to utilize the stored information, including report generation, statistical analyses, and knowledge graphs displaying contact tracing. When the number of outbreak investigation increases, the knowledge graphs can be extended to encompass

* *Corresponding author*. Center for Neuropsychiatric Research, National Health Research Institutes, 35 Keyan Rd., Zhunan, Miaoli 35053, Taiwan.
E-mail address: wjchen@nhri.edu.tw (W.J. Chen).
Peer review under responsibility of Chang Gung University.
https://doi.org/10.1016/j.bj.2020.06.007

other persons appearing in the same location at the same time, i.e., constituting a potential contact cluster. The information extracted can also be used to display the tracing on a map in animation. Overall, this system can provide a basis for further refinement that can be generalized to a variety of outbreak investigations.

Epidemiological investigation in the control for COVID-19

A thorough epidemiological investigation is essential for controlling the spread of infectious diseases, as empirically implemented in the U.K. during the pandemic of 2009 influenza [1] or after the importation of Ebola virus [2]. One key component of such epidemiological investigation is contact tracing. Especially in the early stages of an infectious disease outbreak, contact tracing is a critical control measure to contain its spread. For a highly transmissible infectious disease such as the coronavirus disease 2019 (COVID-19), which has an estimated basic reproduction number (R_0) of 2.6 [3] to 3.1 [4] in preliminary studies, a combination of contact tracing and cases isolation has been demonstrated in mathematical modeling to be crucial to the containment of COVID-19 [5]. Further simulations suggest that about 70% of contacts will have to be successfully traced to control early spread [6]. Since an investigation of the first 100 confirmed cases of COVID-19 in Taiwan found SARS-CoV-2 to have high transmissibility before and immediately after symptom onset [7], early identification and/or isolation of asymptomatic cases become crucial in containing the transmission, as shown in another mathematical modeling study [8]. Hence, an epidemiological investigation during an outbreak needs to be conducted efficiently in a timely manner.

After the emergence of COVID-19 in January 2020 in Wu-Han, China, Taiwan has succeeded in containing its spread, beating the odds of an early prediction of high risk due to close proximity to China. As of June 7, Taiwan had 443 confirmed cases with only 55 of them being indigenous cases, which ranks below 144 countries and regions. Taiwan government adopts an aggressive epidemiological investigation to identify all the individuals at certain level of risk and put them under a variety of control measures, including self-health management, quarantine (quarantine at designated institutes, home quarantine, and home isolation), or isolation (admitted to isolation ward in a designated hospital) [7,9]. As the pandemic lingers, any means that help increase the efficiency of epidemiological investigation can decrease the burden on the personnel in charge of the control of COVID-19. In this study, we aim to develop a new tool that can improve the current practice of epidemiological investigation by introducing new technologies in digital platform and knowledge graphs.

Current practice in epidemiological investigation in Taiwan

Although the basic components of an epidemiological investigation for various infectious diseases are similar, there are some aspects specific to individual outbreaks. Using the COVID-19 as an example, the Taiwan Centers of Disease Control (TCDC) issued the COVID-19 Epidemiological Investigation Questionnaire (version 3) on January 30, 2020. There are four major components in the questionnaire: (A) basic demographic characteristics, including date of notification, date of symptom onset, and occupation; (B) clinical features, including symptoms, health-seeking experiences, and co-morbid chronic diseases; (C) source of exposure in the 14 days prior to the onset of illness, including traveling overseas, particularly to the Hu-Bei Province of China, contact with patients who had fever or respiratory symptoms, and contact with animals; and (D) contact history from the date of symptom onset (and could be extended to up to four days before symptom onset when epidemiologically indicated) till isolation, including taking bus or mass transportation, being to major public facilities, and having face-to-face meeting with some people.

For every laboratory-confirmed case to have SARS-CoV-2 infection using real-time reverse transcriptase-polymerase chain reaction (RT-PCR), a staff member of either TCDC or local health authorities will be assigned to conduct the epidemiological investigation. The investigator needs to build rapport with empathy first and then obtain relevant information in the conversation, not using the questionnaire as a structured interview. After the interview, the investigator will go back to office and type semi-structured reports as requested by the TCDC and local health authorities.

Input platform for various sources of information

Although an epidemiological investigation questionnaire highlights important items to ask during the investigation, it is not rare that important information cannot be obtained from the face-to-face interview. A variety of sources of information, such as records in smart phones, closed circuit television footage, or phone communication might also be sought-after in an appropriate manner. Hence, some investigators might prefer to type the semi-structured report directly without filling the structured questionnaire.

To meet the various needs of the epidemiological investigation, we decided to develop an *e-Outbreak Platform* that provides a semi-structured, multifaceted, computer-aided questionnaire for outbreak investigation. There are three major parts of the platform: portal, data entry, and functions to utilize the stored information.

Portal to the platform

To make the system more user-friendly, we create a graphic portal that allows users to have an at-glance grasp of the

Fig. 1 The *e-Outbreak Platform* that is created using the framework of Microsoft software: (A) the portal to the platform, which allows users to choose functions; (B) data entry form for information collected for the epidemiologic investigation; and (C) a semi-structured report, left panel, generated by the *e-Outbreak Platform* following the format chosen by the user and the activities-date summary sheet.

C

Fig. 1 (*continued*).

functions provided by the platform and then choose the one they need [Fig. 1A]. As the *e-Outbreak Platform* grows, there will be more options on the portal.

Input platform

To solve the issue of different questionnaire items for different infectious diseases, e.g., COVID-19, measles, and dengue fever, the platform allows health authorities to easily upload any newly designed epidemiological questionnaires to the system. Then users only need to choose a disease-specific form of the task assigned. Using COVID-19 as an example [Fig. 1B], the data entry form can accommodate different formats of the information, including text typing, button selection, and pull-down menu. When users key-in information to individual items, the platform can do some logical checking, reminding the user of missing information, and allows the user to modify the value or edit the text later.

Two types of information are critical for the control of COVID-19, i.e., source of exposure and contact tracing. Hence, this input platform provides further flexibility to deal with complicated scenarios. For example, the platform allows the investigator to enter the date, place and people of activities not by chronological order. Later the user can choose to sort the input by date.

Source of exposure

Activities within 14 days prior to the symptom onset are inquired. This part is important for the identification of the source of infection and possible date of contracting the infection. Thus, these date and persons contacted are important to record.

Contact history

Once an infected person became infectious, all the persons who had close contact with the case need to be put under certain control measures, e.g., self-heath management, collection of swab for virus detection, home quarantine, home isolation, or being admitted to an isolation ward. Sometimes the infected case might forget or not know exactly the name of a close contact (e.g., a taxi driver). For health authorities to formally notify these people for respective control measures, the investigator may need to seek outside information based on the date-location information obtained from the epidemiological investigation. Originally, it was recommended to trace from the symptom onset till isolation. Later, the period was extended to four days before the symptom onset as

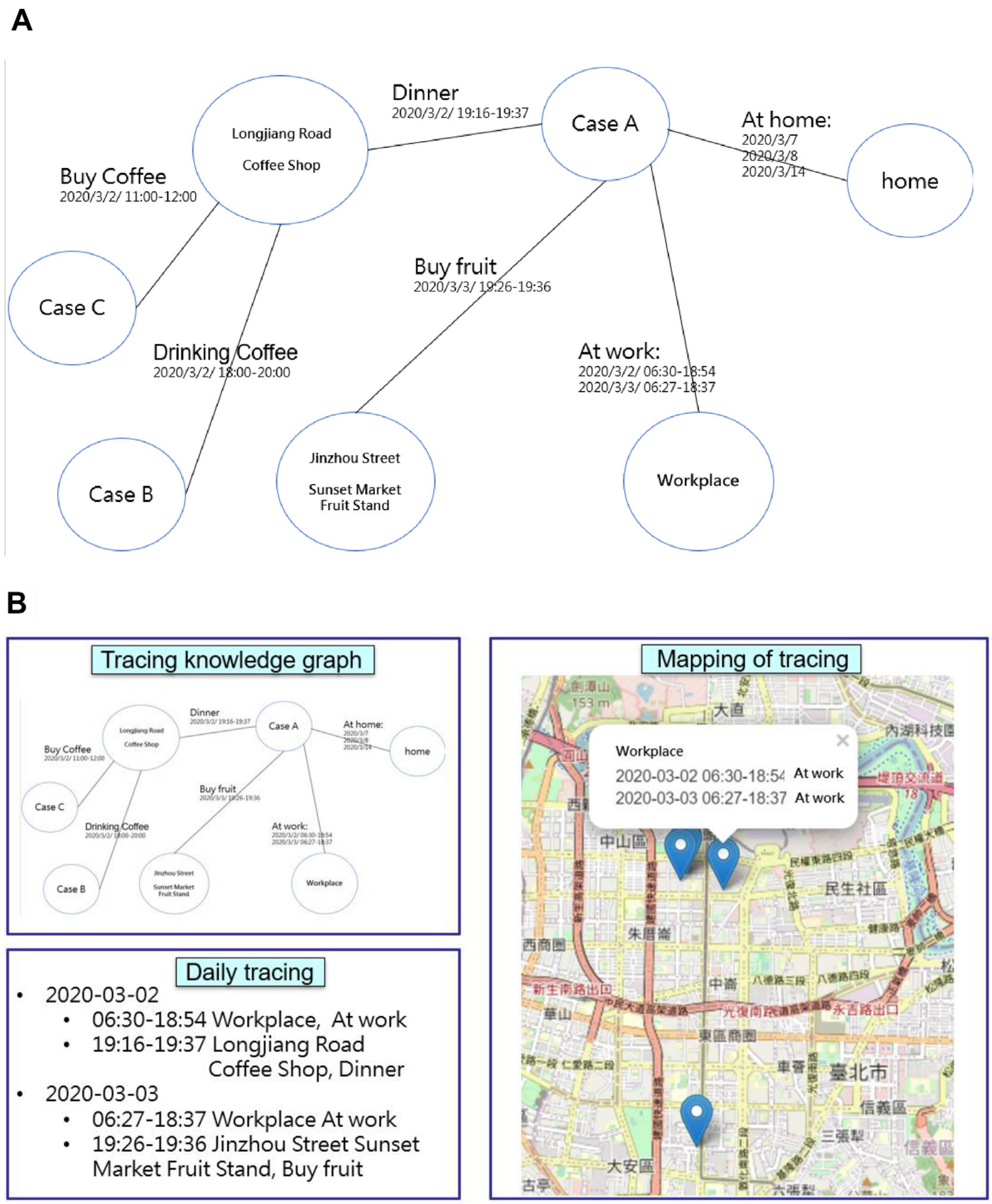

Fig. 2 Knowledge Graphs generated using the information extracted from a semi-structured epidemiological report submitted by investigators: (A) the knowledge graph displaying a confirmed case's tracing records; (B) a summary graph consisting of tracing knowledge graph, daily tracing in text, and the mapping of tracing.

epidemiological data indicated that a COVID-19 patient might have pre-symptom transmission [7].

Report generation

Once all the raw information is input, then a user can choose from the list a particular layout form of outbreak report that can fulfill the requirements of the assigned task. Typically, the report has a semi-structured format. The platform can help generate a draft of such document for further editing [Fig. 1C]. In addition to the text, the platform can supplement a summary of the activities-date. When there are different layouts required by different reporting channels, it can be done simply by choosing the pre-constructed format. This can save investigators much time.

Data security

e-Outbreak Platform is compliant with many data safety standards such as the United States' Health Insurance Portability and Accountability Act (HIPAA) [10] and the European Union's General Data Protection Regulation (GDPR) [11]. Since this platform is built following Microsoft's framework, it is also compliant with many security standards adopted in Microsoft 365.

Display of statistical analyses and applications

This platform incorporates many statistical analysis and application tools available in Microsoft's framework to help reveal the summary trend and statistics [Fig. 1A].

Knowledge graphs of contact tracing

A typical feature of a semi-structured epidemiological investigation report is about the social network and the time-place of contacts. More often than not, there is time strain on the epidemiological investigation because many control measures depend on such information. Displaying important information of the contact tracing extracted from a semi-structured text as a graph allows both highly efficient querying and assisted reasoning. Hence, we adopted the technique of "Knowledge Graphs" [12] that can apply natural language processing (NLP) to extract information and display the relationships graphically.

Knowledge graphs have been applied in research on predictions of unknown adverse drug reactions [13], effective search for healthcare and biomedicine [14,15] or cancer data analytics [16]. Several features of knowledge graphs make it suitable for displaying the information obtained from contact tracing. It can ingest data from various sources, in the form of relational databases, semi-structured data, and unstructured data such as free text, images, and documents. Knowledge graphs can extract knowledge using natural language processing, text mining, and machine learning. After the extraction, knowledge graphs can perform knowledge fusion. Once created, a knowledge can be stored, retrieved, and visually represented.

Knowledge graphs of contact tracing

Knowledge graphs represent facts as edges between nodes that represent entities (e.g. people, food, and vehicle) or concepts (e.g. teacher, tour, and festival). As a preliminary trial, we constructed a knowledge graph containing information extracted from a semi-structured epidemiological report imitating those submitted by investigators [Fig. 2B]. When the number of outbreak investigation increases, the knowledge graphs can be extended to encompass other persons appearing in the same location at the same time, e.g., constituting a potential contact cluster. This may alert the investigator to clarify the relationship by going back to ask the index case.

Furthermore, the information extracted can also be used to display the tracing on a map in animation. All the information useful for decision making can be summarized in a panel [Fig. 2B].

Outlook

There remain several challenges for *e-Outbreak Platform* to be utilized in practice. First, after the preliminary trial by some medical officers from the Epidemic Intelligence Center at TCDC and field investigators from Regional Control Centers during the initial development stage, trial use of the platform in a larger scale by more diverse medical officers as well as field investigators of TCDC are currently under way.

Second, many new digital innovations applied in different countries in responding to COVID-19 have raised certain ethical concerns, including the issues of respecting privacy, respecting autonomy, equity concerns, minimizing the risk of error, and accountability [17,18]. There has been no consensus on the solutions to these issues yet but two principles are recommended when considering the ethics of digital surveillance: the principles of referring to the counterfactual and being judged the least burdensome alternative that would accomplish the public health objectives [18]. Although *e-Outbreak Platform* allows investigators to input information from a variety of sources, users should follow the updated ethical guidelines adopted in the country.

Third, although *e-Outbreak Platform* can help investigators save time, the conduct of epidemiological investigation may need to be adjusted when the scale of confirmed cases increases to an extent that renders the manpower insufficient. In other words, the value of contact tracing should be considered in the context of control measures available or feasible, depending on whether there is community spreading and availability of medication or vaccine.

Fourth, the application of NLP in knowledge graph on the stored information in *e-Outbreak Platform* is of exploratory nature given the limited number of confirmed cases in the country.

Conclusions

We create an online input platform called *e-Outbreak Platform* that allows the health authorities to implement a disease-specific questionnaire. The platform incorporates many correcting algorithms that help decrease errors in texting input. Furthermore, the platform can generate a semi-structured report based on the information retrieved from the questionnaire or allow the user to text-in the report directly. Third, the platform can display the time-place contacts in animation using the techniques of knowledge graphs. Overall, this system can provide a basis for further refinement that can be generalized to a variety of outbreak investigations.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgements

All the authors responded to the call to improve the nation's control measures against COVID-19 on a voluntary basis. Although this work was not funded by any formal grant, we were grateful to our affiliated institutions that supported us to utilize the existing resources available in each institution in carrying out the research. The authors thank Mr. Po-Chang Hsiao of Population Health Research Center, College of Public Health, National Taiwan University, for creating an early version of web-based epidemiological investigation questionnaire for pilot testing.

REFERENCES

[1] McLean E, Pebody RG, Campbell C, Chamberland M, Hawkins C, Nguyen-Van-Tam JS, et al. Pandemic (H1N1) 2009 influenza in the UK: clinical and epidemiological findings from the first few hundred (FF100) cases. Epidemiol Infect 2010;138:1531–41.
[2] Crook P, Smith-Palmer A, Maguire H, McCarthy N, Kirkbride H, Court B, et al. Lack of secondary transmission of Ebola virus from healthcare worker to 238 contacts, United Kingdom, December 2014. Emerg Infect Dis 2017;23:2081–4.
[3] Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, et al. Report 3: Transmissibility of 2019-nCoV. London: Imperial College London; 2020.
[4] Read JM, Bridgen JRE, Cummings DAT, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv 2020. https://doi.org/10.110601/2020.2001.2023.20018549.
[5] Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health 2020;8:e488–96.
[6] Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health 2020;74:861–6.
[7] Cheng HY, Jian SW, Liu DP, Ng TC, Huang WT, Lin HH, et al. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Intern Med 2020;180:1156–63.
[8] Tsou HH, Cheng YC, Yuan HY, Hsu YT, Wu HY, Lee FJ, et al. Prevention of subclinical transmission on the containment of COVID-19: mathematical modeling and experience in Taiwan. Contemp Clin Trials 2020;96:106101.
[9] Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. JAMA 2020;323:1341–2.
[10] Moore W, Frye S. Review of HIPAA, Part 1: history, protected health information, and privacy and security rules. J Nucl Med Technol 2019;47:269–72.
[11] van Veen EB. Observational health research in Europe: understanding the general data protection regulation and underlying debate. Eur J Cancer 2018;104:70–80.
[12] Ehrlinger L, Wöß W. Towards a definition of knowledge graphs. Semantics 2016: Posters and Demos Track. 2016. September 13-14; Leipzig, Germany.
[13] Bean DM, Wu H, Iqbal E, Dzahini O, Ibrahim ZM, Broadbent M, et al. Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. Sci Rep 2017;7:16416.
[14] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 2018;34:i457–66.
[15] Kamdar MR, Stanley CE, Carroll M, Wogulis L, Dowling W, Deus HF, et al. Text snippets to corroborate medical relations: an unsupervised approach using a knowledge graph and embeddings. AMIA Jt Summits Transl Sci Proc 2020;2020:288–97.
[16] Hasan SMS, Rivera D, Wu XC, Durbin EB, Christian B, Tourassi GD. Knowledge graph-enabled cancer data analytics. IEEE J Biomed Health Inform 2020;24:1952–67.
[17] Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020;46:427–31.
[18] Mello MM, Wang CJ. Ethics and governance for digital disease surveillance. Science 2020;368:951.