MAYO CLINIC

EDITORIAL

September 2018
Volume 93
Number 9

MAYO CLINIC
PROCEEDINGS

Social Media Posts and Search Engine Queries as the Canary in the Coal Mine for Public Health Surveillance

See also page 1185

A bird doesn't sing because it has an answer, it sings because it has a song
Maya Angelou

There are few images as evocative as "canary in the coal mine." The canary is a beautiful yellow tropical songbird native to the small North African island group that bears its name; its natural habitat affords freedom of flight, lush vegetation, and tropical sunlight. A coal mine is a fetid, damp hole in the ground, bereft of natural light and devoid of all photosynthetic life.

British coal miners at the turn of the 20th century risked death by asphyxiation from coal damp—carbon monoxide and methane gas (then undetectable toxins). The solution—both simple and brutal—was proposed by John Scott Haldane, a noted Scottish physiologist at the University of Glasgow. Haldane, remembered in British medical history as "the father of oxygen therapy," correctly reasoned that birds would be more sensitive than humans to hypoxia and airborne toxins because of the high metabolic requirements of flight. The canary, a domestic caged bird commonly available in Britain, would be the ideal coal mine sentinel. From 1911 until their final retirement in late 1986, canaries caged in tiny, barred, wooded cages accompanied British miners deep underground. If the canary showed signs of fainting or death, the miners would leave the mine; this strategy saved countless lives before the advent of modern carbon monoxide and methane detectors. It is the contrast between the beautiful world of the free canary and the *Inferno*esque world of the coal mine that strikes us as shocking today; at the time, it was an expedient solution to the real threat of death from toxic coal mine gas. The term *canary in the coal mine* has now come to mean a method of early detection of a threat by a sensitive indicator that can forewarn and trigger countermeasures.

In June 1999, crows started dying at record numbers in New York City. Initially, there was no major public health concern. Two months later (in August 1999), the crow deaths were followed by an outbreak of human West Nile meningoencephalitis resulting from a single-stranded RNA flavivirus; this virus infects both humans and birds alike and had never previously been observed in the Western Hemisphere.[1] Fifty-nine patients were hospitalized in the initial outbreak; 7 of those 59 died. In a fascinating TEDxUCLA talk,[2] Tracy McNamara (the lead veterinary pathologist on the case) recounts the details of the 2-month delay between the crow deaths and the human deaths and the missed opportunity for detection in New York in 1999.

Public health surveillance is a complex set of processes that allows governmental agencies to track disease outbreaks and public health threats at a near real-time rate. It serves a critical national role and is a patchwork of largely physician-dependent processes designed to funnel near real-time data on the diagnoses of emerging public health threats such as acute influenza to the US Centers for Disease Control and Prevention (CDC).[3] Surveillance efforts have been strengthened recently[4] and include a broadened focus on collecting data from emergency department visits, public health visits,

more timely reporting of laboratory results, and better integration of existing information technology platforms.[5-9] Publicly reported disease surveillance tactics do not currently include the analysis of social media postings by the public or Internet search engine queries, despite adaptation of these specific tools for political, economic, and consumer surveillance.[1,2]

It is especially challenging for the CDC to monitor disease-specific mortality trends, such as cardiovascular mortality, in the midst of seasonal influenza outbreaks. The CDC revised action plan[5,8] now strives to get near real-time updates from state and federal mortality registries. In this issue of *Mayo Clinic Proceedings*, Kumar et al[10] demonstrate for the first time a strong association between the use of specific search engine terms within the Google search platform and cardiovascular disease (CVD) mortality in both the Northern and Southern hemispheres. Notably, this association followed a "winter peak and summer trough" profile, this being the well-recognized cycle for seasonal influenza outbreaks. These data thus uncover the temporal peaking of CVD mortality occurring at the same time when the monitoring by CDC for such mortality may be diverted because of CDC's attention to influenza and other seasonal outbreaks.

Kumar et al studied publicly available information by using a total of 28 search query terms across 5 CVD categories. These authors demonstrated a significant correlation between Google search volumes (GSVs) and CVD mortality when using state-specific GSV data. The correlation between GSV and mortality from ischemic heart disease, stroke, and overall CVD ranged from 0.52 to 0.62, consistent with a reasonably strong correlation. The correlations were strongly associated with Northern and Southern hemisphere trends and paralleled previously published real-world data.

Governments and health planning organizations constantly struggle to predict demands for health care services and understand emerging trends in disease burden. The need to detect drug adverse effects not identified by initial regulatory, mandated clinical trials or rare incidents of biological warfare, such as the use of the Novichok nerve agent in Britain, is rarer but still real. Nearly all surveillance for disease activity lags behind real-time trends because data collection and diagnosis validation are made on a delayed basis. The CDC has gone to great lengths to modernize its disease surveillance activity methods to make them as "real-time" as possible; despite those efforts, its methods remain retrospective because of their design methodologies.

It would greatly benefit society and health care organizations to have more proximate real-time disease prediction models. The analysis of GSVs and social media postings may offer such an opportunity. The data provided by Kumar et al represent the first peer-reviewed published article to suggest that Internet searches can predict incident rates of CVD mortality. As social media evolves into newer platforms and modalities, it may be possible to design better diagnostic and surveillance mechanisms into individually based choices of search engine query data. It is conceivable that search engine queries coupled with artificial intelligence tools may facilitate nearly instantaneous detection of acute CVD conditions.

The 21st century "canary in the coal mine" will not be an unfortunate caged bird used to alert humans to environmental toxins but instead will be a virtual reality program triggered by individual Web searches and social media postings collected in aggregate and analyzed by public health authorities to facilitate the early detection of symptoms and diseases gathered on a nearly instantaneous basis. The use of such a program would allow better targeting of resources and tools designed to reduce morbidity and mortality and empower individuals to live healthier and more productive lives. We predict that Internet search engines and social media postings will not only act like the proverbial "canary in the coal mine" in the early detection of disease but also "sing like a canary" in the dissemination of important public health warnings.

Joseph G. Murphy, MD
R. Scott Wright, MD
Department of Cardiology
Mayo Clinic
Rochester, MN

Potential Competing Interests: Dr Wright reports receiving consulting fees from The Medicines Company, Sanofi-Regeneron, and Boehringer Ingelheim. Dr Murphy reports no competing interests.

Correspondence: Address to R. Scott Wright, MD, Department of Cardiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (wright.scott@mayo.edu).

REFERENCES

1. Nash D, Mostashari F, Fine A, et al; 1999 West Nile Outbreak Response Working Group. The outbreak of West Nile virus infection in the New York City area in 1999. *N Engl J Med*. 2001;344(24):1807-1814.
2. McNamara T. Canaries in the coal mine. Presented at: TEDxUCLA; June 21, 2018, https://www.youtube.com/watch?v=qm8NnL582uc2018. Accessed July 20, 2018.
3. Richards CL, Iademarco MF, Anderson TC. A new strategy for public health surveillance at CDC: improving national surveillance activities and outcomes. *Public Health Rep*. 2014;129(6):472-476.
4. Buehler JW; Centers for Disease Control and Prevention. CDC's vision for public health surveillance in the 21st century: introduction. *MMWR Suppl*. 2012;61(3):1-2.
5. Richards CL, Iademarco MF, Atkinson D, et al. Advances in public health surveillance and information dissemination at the Centers for Disease Control and Prevention. *Public Health Rep*. 2017;132(4):403-410.
6. Richards C, Lee B. CDC Surveillance Strategy—a strategy for improving CDC activities in public health surveillance. *Online J Public Health Inform*. 2015;7(1):e49.
7. Centers for Disease Control and Prevention. Surveillance strategy: a strategy for improving the Centers for Disease Control and Prevention's activities in public health surveillance. https://www.cdc.gov/ophss/docs/CDC-Surveillance-Strategy-Final.pdf. Accessed July 20, 2018.
8. Centers for Disease Control and Prevention. CDC Health Information Innovation Consortium (CHIIC). https://www.cdc.gov/ophss/chiic/. Accessed July 20, 2018.
9. Mac Kenzie WR, Davidson AJ, Wiesenthal A, et al. The promise of electronic case reporting. *Public Health Rep*. 2016;131(6):742-746.
10. Kumar N, Pandey A, Garg N, Sampene E, Lavie CJ. Seasonal and geographic patterns in seeking cardiovascular health information: an analysis of the online search trends. *Mayo Clin Proc*. 2018;93(9):1185-1190.