CCDR | CANADA COMMUNICABLE DISEASE REPORT

canada.ca/ccdr

Vol. 46 No. 6 (June 2020)

# ARTIFICIAL INTELLIGENCE IN PUBLIC HEALTH

## GUEST EDITOR: RITA FINLEY



NEXT ISSUE JULY 2, 2020

Public Health Agency of Canada | Agence de la santé publique du Canada

Canada

# CCDR
## CANADA COMMUNICABLE DISEASE REPORT


The *Canada Communicable Disease Report* (CCDR) is a bilingual, peer-reviewed, open-access, online scientific journal published by the Public Health Agency of Canada (PHAC). It provides timely, authoritative and practical information on infectious diseases to clinicians, public health professionals, and policy-makers to inform policy, program development and practice.

The CCDR Editorial Board is composed of members based in Canada, United States of America, European Union and Australia. Board members are internationally renowned and active experts in the fields of infectious disease, public health and clinical research. They meet four times a year, and provide advice and guidance to the Editor-in-Chief.

## Editorial Team



## CCDR Editorial Board members




**Indexed**
in PubMed, Directory of Open Access (DOAJ)/Medicus

**Available**
in PubMed Central (full text)


## Photo Credit

The illustration on the cover of this issue shows the interaction between human and machine in natural language processing as applied to the study of viruses and bacteria in public health. This was based on a Shutterstock photo that was adapted by Lyal Saikaly. (https://www.shutterstock.com/image-illustration/microbiology-future-futuristic-concept-study-viruses-1083123254).

## Contact the Editorial Office


phac.ccdr-rmtc.aspc@canada.ca

613.301.9930

CCDR
CANADA COMMUNICABLE DISEASE REPORT

# ARTIFICIAL INTELLIGENCE IN PUBLIC HEALTH

## TABLE OF CONTENTS

COMMENTARY

# Precision public health: Dream or reality?

Maureen Dobbins[1]*, David Buckeridge[2]

***Suggested citation:*** Dobbins M, Buckeridge D. Precision public health: Dream or reality? Can Commun Dis Rep 2020;46(6):160. https://doi.org/10.14745/ccdr.v46i06a01



**Affiliations**

[1] McMaster University, School of Nursing, National Collaborating Centre for Methods and Tools, Hamilton, ON

[2] McGill University, School of Population and Global Health, Montreal, QC

***Correspondence:**
dobbinsm@mcmaster.ca

Precision public health (PPH) is defined as an "emerging practice to more granularly predict and understand public health risks and customize treatments for more specific and homogenous sub-populations, often using new data, technologies and methods"(1). In Canada, public health practitioners are beginning to focus their attention on precision public health and researchers are developing methods to implement this practice. As this effort ramps up, it is important to ask if the anticipated benefits are likely to be realized, or if we are chasing an unachievable dream?

In a manner similar to evidence-informed public health (2), precision public health (3) has the potential to alter the decision-making process significantly and impact population health positively. Both approaches are based on the use of the best available evidence to classify population health status and identify the optimal interventions for a population, given the health status (2,4). Both approaches have been enabled by the explosive growth in data and computing power over the last decade. For example, the PopHR software platform integrates a wide range of population health data using epidemiologic knowledge to guide decisions about public health interventions (5). Together with a repository of synthesized, quality appraised research evidence evaluating the effectiveness of public health interventions (Health Evidence™) (6), such software platforms may enable the realization of benefits of precision public health.

A pilot project to code the research evidence of interventions related to influenza control illustrated that it is possible to create a coded knowledge base of influenza control interventions that could be "read" by a computer and then paired with local data to identify interventions suited to a specific circumstance. However, the human effort to achieve this automation was not inconsequential and this approach is likely not scalable without additional automation of the process to encode evidence. Furthermore, while a 2019 survey of public health professionals illustrated highly favourable perceptions towards an electronic evidence service for public health, decision-makers also identified issues that would need to be resolved, such as accounting for community values and resources, in order to be comfortable using such a resource. So, while there is much reason for optimism regarding the potential benefits of precision public health, the reality is that there is still much to be learned and considerable work to be done. Recent investments in public health and informatics, however, place Canada at the forefront of this emerging field.

## References

1. Dolley S. Big data's role in precision public health. Front Public Health 2018 Mar 7;6:68. DOI

2. Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. Annu Rev Public Health 2009;30(1):175–201. DOI

3. Desmond-Hellmann S. Progress lies in precision. Science 2016 Aug 19;353(6301):731–1. DOI

4. Horton R. Offline: in defence of precision public health. Lancet 2018 Oct 17;392(10157):1504. DOI

5. Shaban-Nejad A, Lavigne M, Okhmatovskaia A, Buckeridge DL. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. Ann N Y Acad Sci 2017 Jan;1387(1):44–53. DOI PubMed

6. Dobbins M, DeCorby K, Robeson P, Husson H, Tirilis D, Greco L. A knowledge management tool for public health: health-evidence.ca. BMC Public Health 2010 Aug 18;10:496. DOI

# Challenges and opportunities for public health made possible by advances in natural language processing

Oliver Baclic[1]*, Matthew Tunis[1], Kelsey Young[1], Coraline Doan[2], Howard Swerdfeger[2], Justin Schonfeld[3]*

## Abstract

Natural language processing (NLP) is a subfield of artificial intelligence devoted to understanding and generation of language. The recent advances in NLP technologies are enabling rapid analysis of vast amounts of text, thereby creating opportunities for health research and evidence-informed decision making. The analysis and data extraction from scientific literature, technical reports, health records, social media, surveys, registries and other documents can support core public health functions including the enhancement of existing surveillance systems (e.g. through faster identification of diseases and risk factors/at-risk populations), disease prevention strategies (e.g. through more efficient evaluation of the safety and effectiveness of interventions) and health promotion efforts (e.g. by providing the ability to obtain expert-level answers to any health related question). NLP is emerging as an important tool that can assist public health authorities in decreasing the burden of health inequality/inequity in the population. The purpose of this paper is to provide some notable examples of both the potential applications and challenges of NLP use in public health.



**Affiliations**

[1] Centre for Immunization and Respiratory Infectious Disease, Public Health Agency of Canada, Ottawa, ON

[2] Data, Partnerships and Innovation Hub, Public Health Agency of Canada, Ottawa, ON

[3] National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB

***Correspondence:**

oliver.baclic@canada.ca

justin.schonfeld@canada.ca

***Suggested citation:*** Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J. Challenges and opportunities for public health made possible by advances in natural language processing. Can Commun Dis Rep 2020;46(6):161–8. https://doi.org/10.14745/ccdr.v46i06a02

***Keywords:*** natural language processing, NLP, artificial intelligence, machine learning, public health

## Introduction

There is a growing interest in deploying artificial intelligence (AI) strategies to achieve public health outcomes, particularly in response to the global coronavirus disease 2019 (COVID-19) pandemic where novel datasets, surveillance tools and models are emerging very quickly.

The objective of this manuscript is to provide a framework for considering natural language processing (NLP) approaches to public health based on historical applications. This overview includes a brief introduction to AI and NLP, suggests opportunities where NLP can be applied to public health problems and describes the challenges of applying NLP in a public health context. Particular articles were chosen to emphasize the breadth of potential applications for NLP in public health as well as the not inconsiderable challenges and risks inherent in incorporating AI/NLP in public health analysis and decision support.

## Artificial intelligence and natural language processing

AI research has produced models that can interpret a radiograph (1,2), detect irregular heartbeats using a smartwatch (3), automatically identify reports of infectious disease in the media (4), ascertain cardiovascular risk factors from retinal images (5) and find new targets for existing medications (6,7). The success of these models is built from training on hundreds, thousands and sometimes millions of controlled, labelled and structured data points (8). The capacity of AI to provide constant, tireless and rapid analyses of data offers the potential to transform society's approach to promoting health and preventing and managing diseases. AI systems have the potential to "read" and triage all of the approximately 1.3 million research articles indexed by PubMed each year (9); "examine" comments from 1.5 billion Facebook users or "monitor" 500 million tweets of people struggling with mental illness on a daily basis, foodborne illness or the flu (10,11); and simultaneously interact with each and every person seeking answers to their health questions, concerns, problems and challenges (12).

NLP is a subfield of AI that is devoted to developing algorithms and building models capable of using language in the same way humans do (13). It is routinely used in virtual assistants like "Siri" and "Alexa" or in Google searches and translations. NLP provides the ability to analyze and extract information from unstructured sources, automate question answering and conduct sentiment analysis and text summarization (8). With natural language (communication) being the primary means of knowledge collection and exchange in public health and medicine, NLP is the key to unlocking the potential of AI in biomedical sciences.

Most modern NLP platforms are built on models refined through machine learning techniques (14,15). Machine learning techniques are based on four components: a model; data; a loss function, which is a measure of how well the model fits the data; and an algorithm for training (improving) the model (16). Recent breakthroughs in these areas have led to vastly improved NLP models that are powered by deep learning, a subfield of machine learning (17).

Innovation in the different types of models, such as recurrent neural network-based models (RNN), convolutional neural network-based models (CNN) and attention-based models, has allowed modern NLP systems to capture and model more complex linguistic relationships and concepts than simple word presence (i.e. keyword search) (18). This effort has been aided by vector-embedding approaches to preprocess the data that encode words before feeding them into a model. These approaches recognize that words exist in context (e.g. the meanings of "patient," "shot" and "virus" vary depending on context) and treat them as points in a conceptual space rather than isolated entities. The performance of the models has also been improved by the advent of transfer learning, that is, taking a model trained to perform one task and using it as the starting model for training on a related task. Hardware advancements and increases in freely available annotated datasets have also boosted the performance of NLP models. New evaluation tools and benchmarks, such as GLUE, superglue and BioASQ, are helping to broaden our understanding of the type and scope of information these new models can capture (19–21).

## Opportunities

Public health aims to achieve optimal health outcomes within and across different populations, primarily by developing and implementing interventions that target modifiable causes of poor health (22–26). Success depends on the ability to effectively quantify the burden of disease or disease risk factors in the population and subsequently identify groups that are disproportionately affected or at-risk; identify best practices (i.e. optimal prevention or therapeutic strategies); and measure outcomes (27). This evidence-informed model of decision making is best represented by the PICO concept (patient/problem, intervention/exposure, comparison, outcome). PICO provides an optimal knowledge identification strategy to frame and answer specific clinical or public health questions (28). Evidence-informed decision making is typically founded on the comprehensive and systematic review and synthesis of data in accordance with the PICO framework elements.

Today, information is being produced and published (e.g. scientific literature, technical reports, health records, social media, surveys, registries and other documents) at unprecedented rates. By providing the ability to rapidly analyze large amounts of unstructured or semistructured text, NLP has opened up immense opportunities for text-based research and evidence-informed decision making (29–34). NLP is emerging as a potentially powerful tool for supporting the rapid identification of populations, interventions and outcomes of interest that are required for disease surveillance, disease prevention and health promotion. For example, the use of NLP platforms that are able to detect particular features of individuals (population/problem, e.g. a medical condition or a predisposing biological, behavioural, environmental or socioeconomic risk factor) in unstructured medical records or social media text can be used to enhance existing surveillance systems with real-world evidence. One recent study demonstrated the ability of NLP methods to predict the presence of depression prior to its appearance in the medical record (35). The ability to conduct real-time text mining of scientific publications for a particular PICO concept provides opportunities for decision makers to rapidly provide recommendations on disease prevention or management that are informed by the most current body of evidence when timely guidance is essential, such as during an outbreak. NLP-powered question-answering platforms and chatbots also carry the potential to improve health promotion activities by engaging individuals and providing personalized support or advice. **Table 1** provides examples of potential applications of NLP in public health that have demonstrated at least some success.

## Challenges

Despite the recent advances, barriers to widespread use of NLP technologies remain.

Similar to other AI techniques, NLP is highly dependent on the availability, quality and nature of the training data (72). Access and availability of appropriately annotated datasets (to make effective use of supervised or semi-supervised learning) are fundamental for training and implementing robust NLP models. For example, the development and use of algorithms that are able to conduct a systematic synthesis of published research on a particular topic or an analysis and data extraction from electronic health records requires unrestricted access to publisher or primary care/hospital databases. While the number of freely accessible biomedical datasets and pre-trained models has been increasing in recent years, the availability of those dealing with public health concepts remains limited (73).

**Table 1: Examples of existing and potential applications of natural language processing in public health**

| Type of activity | Public health objective | Example of NLP use |
|---|---|---|
| Identification of at-risk populations or conditions of interest | To continuously measure the incidence and prevalence of diseases and disease risk factors (i.e. surveillance) | Analysis of unstructured or semistructured text from electronic health records or social media (36–42) |
| | To identify vulnerable and at-risk populations | Analysis of risk behaviours using social media (43–45) |
| Identification of health interventions | To develop optimal recommendations/ interventions | Automated systematic review and analysis of the information contained in scientific publications and unpublished data (46–50) |
| | To identify best practices | Identification of promising public health interventions through analysis of online grey and peer reviewed literature (51) |
| Identification of health outcomes using real-world evidence | To evaluate the benefits of health interventions | Analysis of unstructured or semistructured text from electronic health records, online media and publications to determine the impact of public health recommendations and interventions (52,53) |
| | To identify unintended adverse outcomes related to interventions | Analysis of unstructured or semistructured text from electronic health records, social media and publications to identify potential adverse events of interventions (54–58) |
| Knowledge generation and translation | To support public health research | Analysis and extraction of information from electronic health records and scientific publications for knowledge generation (59–62) |
| | To support evidence-informed decision making | Use of chatbots, question/answer systems and text summarizers to provide personalized information to individuals seeking advice to improve their health and prevent disease (63–65) |
| Environmental scanning and situational awareness | To conduct public health risk assessments and provide situational awareness | Analysis of online content for real-time critical event detection and mitigation (66–70) |
| | To monitor activities that may have an impact on public health decision making | Analysis of decisions of international and national stakeholders (71) |

Abbreviation: NLP, natural language processing

The ability to de-bias data (i.e. by providing the ability to inspect, explain and ethically adjust data) represents another major consideration for the training and use of NLP models in public health settings. Failing to account for biases in the development (e.g. data annotation), deployment (e.g. use of pre-trained platforms) and evaluation of NLP models could compromise the model outputs and reinforce existing health inequity (74). However, it is important to note that even when datasets and evaluations are adjusted for biases, this does not guarantee an equal impact across morally relevant strata. For example, use of health data available through social media platforms must take into account the specific age and socioeconomic groups that use them. A monitoring system trained on data from Facebook is likely to be biased towards health data and linguistic quirks specific to a population older than one trained on data from Snapchat (75). Recently many model agnostic tools have been developed to assess and correct unfairness in machine learning models in accordance with the efforts by the government and academic communities to define unacceptable AI development (76–81).

Currently, one of the biggest hurdles for further development of NLP systems in public health is limited data access (82,83). Within Canada, health data are generally controlled regionally and, due to security and confidentiality concerns, there is reluctance to provide unhindered access to these systems and their integration with other datasets (e.g. data linkage). There have also been challenges with public perception of privacy and data access. A recent survey of social media users found that the majority considered analysis of their social media data to identify mental health issues "intrusive and exposing" and they would not consent to this (84).

Before key NLP public health activities can be realized at scale, such as the real-time analysis of national disease trends, jurisdictions will need to jointly determine a reasonable scope and access to public health–relevant data sources (e.g. health record and administrative data). In order to prevent privacy violations and data misuse, future applications of NLP in the analysis of personal health data are contingent on the ability to embed differential privacy into models (85), both during training and postdeployment. Access to important data is also limited through the current methods for accessing full text publications. Realization of fully automated PICO-specific knowledge extraction and synthesis will require unrestricted access to journal databases or new models of data storage (86).

Finally, as with any new technology, consideration must be given to assessment and evaluation of NLP models to ensure that they are working as intended and keeping in pace with society's changing ethical views. These NLP technologies need to be assessed to ensure they are functioning as expected and account for bias (87). Although today many approaches are posting equivalent or better-than-human scores on textual analysis tasks, it is important not to equate high scores with true language understanding. It is, however, equally important not to view

a lack of true language understanding as a lack of usefulness. Models with a "relatively poor" depth of understanding can still be highly effective at information extraction, classification and prediction tasks, particularly with the increasing availability of labelled data.

## Natural language processing and the coronavirus disease 2019 (COVID-19)

With the emergence of the COVID-19, NLP has taken a prominent role in the outbreak response efforts (88,89). NLP has been rapidly employed to analyze the vast quantity of textual information that has been made available through unrestricted access to peer-review journals, preprints and digital media (90). NLP has been widely used to support the medical and scientific communities in finding answers to key research questions, summarization of evidence, question answering, tracking misinformation and monitoring of population sentiment (91–97).

## Conclusion

NLP is creating extraordinary opportunities to improve evidence-informed decision making in public health. We anticipate that broader applications of NLP will lead to the creation of more efficient surveillance systems that are able to identify diseases and at-risk conditions in real time. Similarly, with an ability to analyze and synthesize large volumes of information almost instantaneously, NLP is expected to facilitate targeted health promotion and disease prevention activities, potentially leading to population-wide disease reduction and greater health equity. However, these opportunities are not without risks: biased models, biased data, loss of data privacy and the need to maintain and update models to reflect the evolving language and context of public communication are all existing challenges that will need to be addressed. We encourage the public health and computer science communities to collaborate in order to mitigate these risks, ensure that public health practice does not fall behind in these technologies or miss opportunities for health promotion and disease surveillance and prevention in this rapidly evolving landscape.

## Authors' statement

OB — Writing – original draft, review & editing and conceptualization
MT — Writing – original draft, review & editing and conceptualization
KY — Writing – review & editing, and conceptualization
CD — Writing – review & editing
HS — Writing – review & editing
JS — Writing – original draft, review & editing and conceptualization

## Conflict of interest

None.

## Acknowledgements

We thank J Nash and J Robertson who were kind enough to offer feedback and suggestions.

## Funding

This work is supported by the Public Health Agency of Canada. The research undertaken by JS was funded by the Canadian federal government's Genomic Research and Development Initiative.

## References

1. Majkowska A, Mittal S, Steiner DF, Reicher JJ, McKinney SM, Duggan GE, Eswaran K, Cameron Chen PH, Liu Y, Kalidindi SR, Ding A, Corrado GS, Tse D, Shetty S. Chest radiograph interpretation with deep learning models: assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. Radiology 2020;294(2):421–31. DOI PubMed

2. Liu X, Faes L, Kale A, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C, Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A comparison of deep learning performance against health care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health 2019. DOI

3. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, Balasubramanian V, Russo AM, Rajmane A, Cheung L, Hung G, Lee J, Kowey P, Talati N, Nag D, Gummidipundi SE, Beatty A, Hills MT, Desai S, Granger CB, Desai M, Turakhia MP; Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019;381(20):1909–17. DOI PubMed

4. Feldman J, Thomas-Bachli A, Forsyth J, Patel ZH, Khan K. Development of a global infectious disease activity database using natural language processing, machine learning, and human expertise. J Am Med Inform Assoc 2019;26(11):1355–9. DOI PubMed

5. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L, Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2018;2(3):158–64. DOI PubMed

6. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 2019;18(6):463–77. DOI PubMed

7. Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, Humeidi R, Peck D, Wu X, Tang AA, Wang VM, Bender SA, Lemire E, Narayan R, Montgomery P, Ben-David U, Garvie CW, Chen Y, Rees MG, Lyons NJ, McFarland JM, Wong BT, Wang L, Dumont N, O'Hearn PJ, Stefan E, Doench JG, Harrington CN, Greulich H, Meyerson M, Vazquez F, Subramanian A, Roth JA, Bittker JA, Boehm JS, Mader CC, Tsherniak A, Golub TR. Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nat Can 2020;1:235–48. DOI

8. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44–56. DOI PubMed

9. MEDLINE PubMed Production Statistics. Bethesda (MD): U.S. National Library of Medicine (updated 2019-11-19; accessed 2020-01-27). https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html

10. Twitter usage statistics. Internet LiveStats.com (updated 2013-08-16; accessed 2020-01-27). https://www.internetlivestats.com/twitter-statistics/

11. Searching for health. Google News Lab, Schema; 2017 (accessed 2020-01-27). https://googlenewslab.gistapp.com/searching-for-health

12. Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Shortliffe E, Cimino J, editors. Biomed Informatics London: Springer; 2014. DOI

13. Ruder S. NLP-progress. London (UK): Sebastian Ruder (accessed 2020-01-18). https://nlpprogress.com/

14. Jurafsky D, Martin JH. Speech and language processing. Stanford (CA): Stanford University; 2019 (updated 2019-11-16; accessed 2020-01-18). https://web.stanford.edu/~jurafsky/slp3/

15. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544–51. DOI PubMed

16. Nilsson N. Introduction to machine learning. Stanford (CA): Robotic Library, Department of Computer Science, Stanford University; 1998. http://robotics.stanford.edu/people/nilsson/MLBOOK.pdf

17. Zhou M, Duan N, Liu S, Shum HY. Progress in neural NLP: modeling, learning, and reasoning. Engineering 2020;6(3):275–90. DOI

18. Tang B, Pan Z, Yin K, Khateeb A. Recent advances of deep learning in bioinformatics and computational biology. Front Genet 2019;10:214. DOI PubMed

19. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015;349(6245):261–6. DOI PubMed

20. Wang A, Singh A, Michael J, Hill F, Levy O, Bowman S. GLUE: a multi-task benchmark and analysis platform for natural language understanding. Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. Brussels (BE): 2018 Nov; p. 353–5. DOI

21. The Big Bad NLP Database. New York (NY): Quantum Stat; 2020 (updated 2020-01-21; accessed 2020-01-27). https://quantumstat.com/dataset/dataset.html

22. Jackson B, Huston P. Advancing health equity to improve health: the time is now. Health Promot Chronic Dis Prev Can 2016;36(2):17–20. DOI PubMed

23. Pan American Health Organization. Just societies: health equity and dignified lives. Report of the Commission of the Pan American Health Organization on Equity and Health Inequalities in the Americas. Washington (DC): Pan American Health Organization (updated 2019-11; accessed 2020-01-18). http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsebk&AN=2329553

24. Marmot M, Allen J, Goldblatt P, Boyce T, McNeish D, Grady M, Geddes I; The Marmot Review. Fair society, healthy lives: strategic review of health inequalities in England post-2010. UCL Institute of Health Equity. http://www.parliament.uk/documents/fair-society-healthy-lives-full-report.pdf

25. Arcaya MC, Arcaya AL, Subramanian SV. Inequalities in health: definitions, concepts, and theories. Glob Health Action 2015;8:27106. DOI PubMed

26. Public Health Agency of Canada. The Chief Public Health Officer's report on the state of public health in Canada: addressing health inequalities. Ottawa (ON): Public Health Agency of Canada; 2008. Report No.: HP2-10/2008E. http://www.phac-aspc.gc.ca/cphorsphc-respcacsp/2008/fr-rc/index-eng.ph

27. Ndumbe-Eyoh S, Dyck L, Clement C. Common agenda for public health action on health equity. Antigonish (NS): National Collaborating Centre for Determinants of Health, St Francis Xavier University; 2016. http://nccdh.ca/images/uploads/comments/Common_Agenda_EN.pdf

28. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A, Guyatt GH, Oxman AD; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ 2016;353:i2016. DOI PubMed

29. Kim ES, James P, Zevon ES, Trudel-Fitzgerald C, Kubzansky LD, Grodstein F. Social media as an emerging data resource for epidemiologic research: characteristics of social media users and non-users in the Nurses' Health Study II. Am J Epidemiol 2020;189(2):156–61. DOI PubMed

30. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. J Am Med Inform Assoc 2019;26(4):364–79. DOI PubMed

31. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Syst Rev 2019;8(1):163. DOI PubMed

32. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc 2019;26(6):561–76. DOI PubMed

33. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform 2017;73:14–29. DOI PubMed

34. The Office of the National Coordinator for Health Information Technology. Health IT dashboard: Quick stats. Washington (DC): U.S. Department of Health and Human Services. https://dashboard.healthit.gov/quickstats/quickstats.php

35. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J; Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014;63(32):681–5. PubMed

36. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, Tozzi AE. Influenza-like illness surveillance on Twitter through automated learning of naïve language. PLoS One 2013;8(12):e82489. DOI PubMed

37. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, Asch DA, Schwartz HA. Facebook language predicts depression in medical records. Proc Natl Acad Sci USA. 2018;115(44):11203–8. DOI

38. Şerban O, Thapen N, Maginnis B, Hankin C, Foot V. Real-time processing of social media with SENTINEL: a syndromic surveillance system incorporating deep learning for health classification. Inf Process Manage 2019;56(3):1166–84. DOI

39. Edo-Osagie O, Smith G, Lake I, Edeghere O, De La Iglesia B. Twitter mining using semi-supervised classification for relevance filtering in syndromic surveillance. PLoS One 2019;14(7):e0210689. DOI PubMed

40. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016;23(5):1007–15. DOI PubMed

41. Dorr D, Bejan CA, Pizzimenti C, Singh S, Storer M, Quinones A. Identifying patients with significant problems related to social determinants of health with natural language processing. Stud Health Technol Inform 2019;264:1456–7. DOI PubMed

42. Carrell DS, Cronkite D, Palmer RE, Saunders K, Gross DE, Masters ET, Hylan TR, Von Korff M. Using natural language processing to identify problem usage of prescription opioids. Int J Med Inform 2015;84(12):1057–64. DOI PubMed

43. Cacheda F, Fernandez D, Novoa FJ, Carneiro V. Early detection of depression: social network analysis and random forest techniques. J Med Internet Res 2019;21(6):e12554. DOI PubMed

44. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and consumer generated data. Yearb Med Inform 2019;28(1):208–17. DOI PubMed

45. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: from linguistic signal to clinical reality. Baltimore (MA): 27 June 2014;p. 51–60. DOI

46. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, Hartling L. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst Rev 2019;8(1):278. DOI PubMed

47. Przybyła P, Soto AJ, Ananiadou S. Identifying personalised treatments and clinical trials for precision medicine using semantic search with Thalia. Manchester (UK): TREC; 2017. https://www.researchgate.net/publication/323629465_Identifying_Personalised_Treatments_and_Clinical_Trials_for_Precision_Medicine_using_Semantic_Search_with_Thalia

48. Bannach-Brown A, Przybyła P, Thomas J, Rice AS, Ananiadou S, Liao J, Macleod MR. Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Syst Rev 2019;8(1):23. DOI PubMed

49. Norman C, Leeflang M, Spijker R, Kanoulas E, Névéol A. A distantly supervised dataset for automated data extraction from diagnostic studies. ACL Workshop on Biomedical Natural Language Processing, Florence (IT): 2019 Aug. DOI

50. Tsafnat G, Glasziou P, Karystianis G, Coiera E. Automated screening of research studies for systematic reviews using study characteristics. Syst Rev 2018;7(1):64. DOI PubMed

51. Lerner I, Créquit P, Ravaud P, Atal I. Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. J Clin Epidemiol 2019;108:86–94. DOI PubMed

52. Tucker TC, Durbin EB, McDowell JK, Huang B. Unlocking the potential of population-based cancer registries. Cancer 2019;125(21):3729–37. DOI PubMed

53. Mohammadhassanzadeh H, Sketris I, Traynor R, Alexander S, Winquist B, Stewart SA. Using natural language processing to examine the uptake, content, and readability of media coverage of a pan-Canadian drug safety research project: cross-sectional observational study. JMIR Form Res 2020;4(1):e13296. DOI PubMed

54. Banerji A, Lai KH, Li Y, Saff RR, Camargo CA Jr, Blumenthal KG, Zhou L. Natural language processing combined with ICD-9-CM codes as a novel method to study the epidemiology of allergic drug reactions. J Allergy Clin Immunol Pract 2020;8(3):1032–1038.e1. DOI PubMed

55. Young IJ, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019;132:103971. DOI PubMed

56. Henry S, Buchan K, Filannino M, Stubbs A, Uzuner O. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. J Am Med Inform Assoc 2020;27(1):3–12. DOI PubMed

57. Fan B, Fan W, Smith C, Garner H. Adverse drug event detection and extraction from open data: a deep learning approach. Inf Process Manage 2020;57(1):102131. DOI

58. Yu W, Zheng C, Xie F, Chen W, Mercado C, Sy LS, Qian L, Glenn S, Tseng HF, Lee G, Duffy J, McNeil MM, Daley MF, Crane B, McLean HQ, Jackson LA, Jacobsen SJ. The use of natural language processing to identify vaccine-related anaphylaxis at five health care systems in the Vaccine Safety Datalink. Pharmacoepidemiol Drug Saf 2020;29(2):182–8. DOI PubMed

59. Liu F, Weng C, Yu H. Advancing clinical research through natural language processing on electronic health records: traditional machine learning meets deep learning. In: Richesson RL, Andrews JE, editors: Clinical Research Informatics. Springer International Publishing; 2019. p. 357–78. DOI

60. Chan L, Beers K, Yau AA, Chauhan K, Duffy Á, Chaudhary K, Debnath N, Saha A, Pattharanitima P, Cho J, Kotanko P, Federman A, Coca SG, Van Vleck T, Nadkarni GN. Natural language processing of electronic health records is superior to billing codes to identify symptom burden in hemodialysis patients. Kidney Int 2020;97(2):383–92. DOI PubMed

61. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020;145(2):463–9. DOI PubMed

62. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S, Liu H. Clinical information extraction applications: A literature review. J Biomed Inform 2018;77:34–49. DOI PubMed

63. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, Surian D, Gallego B, Magrabi F, Lau AY, Coiera E. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018;25(9):1248–58. DOI PubMed

64. Head to health. COVID-19 support. Department of Health; Australian Government (accessed 2020-01-27). https://headtohealth.gov.au/sam-the-chatbot

65. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst 2019;43(5):135. DOI PubMed

66. Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). Can Commun Dis Rep 2015;41(9):209–14. DOI PubMed

67. Ghosh S, Chakraborty P, Lewis BL, Majumder M, Cohn E, Brownstein JS, Marathe M, Ramakrishnan N. GELL: Automatic extraction of epidemiological line lists from open sources. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017 Aug 13-17; Halifax (NS): Association for Computing Machinery; 2017. p. 1477–86. DOI

68. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EH, Olsen JM, Pavlin JA, Shigematsu M, Streichert LC, Suda KJ, Corley CD. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. PLoS One 2015;10(10):e0139701. DOI PubMed

69. Jordan S, Hovet S, Fung I, Liang H, Fu KW, Tsz Ho Tse Z. Using Twitter for public health surveillance from monitoring and prediction to public response. Data (Basel) 2018;4(1):6. DOI

70. Abbood A, Ullrich A, Busche R, Ghozzi S. EventEpi—A natural language processing framework for event-based surveillance medRxiv 2019;19006395. DOI

71. Anglin K. Gather-narrow-extract: A framework for studying local policy variation using web-scraping and natural language processing. J Res Educ Eff 2019;12(4):685–706. DOI

72. Tatman R, Conner K. Effects of talker dialect, gender & race on accuracy of Bing speech and YouTube automatic captions. Proc Interspeech 2017;934–8. DOI

73. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform 2020;8(3):e17984. DOI PubMed

74. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018;169(12):866–72. DOI PubMed

75. Gramlich J. 10 facts about Americans and Facebook. Washington (DC): Pew Research Center (accessed 2020-01-27). https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/

76. Xu C, Doshi T. Fairness indicators: scalable infrastructure for fair ML system. Mountain View (CA): Google (accessed 2020-01-27). DOI

77. Holstein K, Vaughan JW, Daumé H, Dudík M, Wallach H. Improving fairness in machine learning systems: What do industry practitioners need? CHI ‹19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Paper No.: 600. p. 1–16. DOI

78. Wiens J, Price WN 2nd, Sjoding MW. Diagnosing bias in data-driven algorithms for healthcare. Nat Med 2020;26(1):25–6. DOI PubMed

79. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. Nat Med 2020;26(1):16–7. DOI PubMed

80. Montreal Declaration for a Responsible Development of Artificial Intelligence. Forum on the Socially Responsible Development of AI: 2017 Nov 2-3: Montréal (QC) (accessed 2020-01-18). https://www.montrealdeclaration-responsibleai.com/the-declaration

81. Treasury Board Secretariat. Directive on automated decision-making. Ottawa (ON): Government of Canada (modified 2019-02-05; accessed 2020-01-27). https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592

82. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. J Biomed Inform 2013;46(5):765–73. DOI PubMed

83. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019;7(2):e12239. DOI PubMed

84. Ford E, Curlewis K, Wongkoblap A, Curcin V. Public opinions on using social media content to identify users with depression and target mental health care advertising: mixed methods survey. JMIR Ment Health 2019;6(11):e12942. DOI PubMed

85. Radebaugh C, Erlingsson U. Introducing tensorflow privacy: learning with differential privacy for training data. Medium.com (accessed 2020-01-27). https://medium.com/tensorflow/introducing-tensorflowprivacy-learning-with-differential-privacy-for-trainingdata-b143c5e801b6

86. Penning de Vries BB, van Smeden M, Rosendaal FR, Groenwold RH. Title, abstract, and keyword searching resulted in poor recovery of articles in systematic reviews of epidemiologic practice. J Clin Epidemiol 2020;121:55–61. DOI PubMed

87. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447–53. DOI PubMed

88. Coronavirus tech handbook: natural language processing. https://coronavirustechhandbook.com/nlp

89. COVID-19 Open Research Dataset Challenge (CORD-19): An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House. San Francisco (CA): kaggle.com (accessed 2020-01-27). https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

90. PubMed Central. Public Health Emergency COVID-19 Initiative. Bethesda (MD): US National Library of Medicine (accessed 2020-01-27). PubMed

91. Bullock J, Luccioni A, Pham KH, Lam CS, Luengo-Oroz M. Mapping the landscape of artificial intelligence applications against COVID-19. arXiv:2003.11336 [cs.CY]. https://vectorinstitute.ai/wp-content/uploads/2020/03/arxiv-mappingai.pdf

92. Allen Institute for Artificial Intelligence (AI2). CORD-19 Explorer: explore the dataset. https://cord-19.apps.allenai.org/

93. Chen E, Lerman K, Ferrara E. COVID-19: the first public coronavirus Twitter dataset. Ithaca (NY): Cornell University (accessed 2020-01-27). https://arix.org/abs/2003.07372

94. LitCovid. Bethesda (MD): U.S. National Library of Medicine (accessed 2020-01-27). PubMed

95. Coronafiles: Chatbots take strain off Denmark's emergency helplines (accessed 2020-01-27). https://sifted.eu/articles/coronafiles-chatbots-helplines/

96. Kritikos M. At a glance: scientific foresight: What if we could fight coronavirus with artificial intelligence? Scientific Foresight Unit, European Parliament. https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/641538/EPRS_ATA(2020)641538_EN.pdf

97. Against AI. COVID-19 Canada. About us. Toronto (ON): CIFAR (accessed 2020-01-27). https://ai-against-covid.ca/

# A call for an ethical framework when using social media data for artificial intelligence applications in public health research

Jean-Philippe Gilbert[1]*, Victoria Ng[2], Jingcheng Niu[3], Erin E Rees[2]

## Abstract

Advancements in artificial intelligence (AI), more precisely the subfield of machine learning, and their applications to open-source internet data, such as social media, are growing faster than the management of ethical issues for use in society. An ethical framework helps scientists and policy makers consider ethics in their fields of practice, legitimize their work and protect members of the data-generating public. A central question for advancing the ethical framework is whether or not Tweets, Facebook posts and other open-source social media data generated by the public represent a human or not. The objective of this paper is to highlight ethical issues that the public health sector will be or is already confronting when using social media data in practice. The issues include informed consent, privacy, anonymization and balancing these issues with the benefits of using social media data for the common good. Current ethical frameworks need to provide guidance for addressing issues arising from the use of social media data in the public health sector. Discussions in this area should occur while the application of open-source data is still relatively new, and they should also keep pace as other problems arise from ongoing technological change.



**Affiliations**

[1] Université Laval, Québec, QC

[2] Public Health Agency of Canada, Ottawa, ON

[3] University of Toronto, Toronto, ON

*Correspondence:
jean-philippe.gilbert.5@ulaval.ca

***Suggested citation:*** Gilbert J-P, Ng V, Niu J, Rees EE. A call for an ethical framework when using social media data for artificial intelligence applications in public health research. Can Commun Dis Rep 2020;46(6):169–73. https://doi.org/10.14745/ccdr.v46i06a03

***Keywords:*** ethics, ethical research, social media, artificial intelligence

## Introduction

Rapid technological advancements in artificial intelligence (AI), and more specifically, natural language processing (NLP) using machine learning techniques, are enabling easy access and use of open-source big data. NLP allows computers to analyze datasets of natural language discourse (i.e. text not structured for quantitative analysis).

In public health, digital epidemiology has emerged as a new field that focuses on using non–public health sector data such as open-source internet data (e.g. Google Trends, news media) and social media data (e.g. Twitter and Facebook posts), whereas traditional epidemiology uses data collected for the purposes of health care, such as reporting of notifiable diseases by healthcare professionals to contribute to data for the surveillance of disease cases.

Researchers and policy makers recognize the potential of digital epidemiology data for advancing early warning of public health threats (1–3). Odlum & Yoon (4) used NLP to assess Twitter data and reported that Tweets related to Ebola increased in the days leading up to the official alert of the 2014 Ebola outbreak in Africa. Yousefinaghani *et al.* (5) showed that 75% of real-time outbreak notifications of avian influenza were identifiable from Twitter; one-third of outbreak notifications were reported on Twitter earlier than official reports. These observations support using Twitter volumes to predict the occurrence of outbreaks, and even forecast expected case counts, has also been shown with Google Trends data (1,6). Furthermore, refinement of social media data into various disease-relevant categories, by using NLP to classify Tweets into symptom types (e.g. fever, vomit), or focusing analysis on specific search terms from Google Trends, helps increase the accuracy in predictions of outbreak occurrence and forecast estimates.

Research that uses data from human participants requires ethical approval. A review process by a government body or university committee independent of the researchers assesses if use of these data ensures the safety, dignity and rights of the participants. Researchers need to demonstrate to the research ethics board (REB) that their study minimizes harm to participants

and respects their autonomy, generates and maximizes benefit (e.g. to society, science, participants) and acts with integrity, fairness and transparency to all stakeholders (e.g. participants, beneficiaries of the research). However, in a systematic review of the utilization of Twitter for health research, only 32% of the studies acquired ethical approval (7).

This is an example of technology moving faster than policy, in that the availability of newer data sources, such as from social media, have outpaced the need to assess the ethics of their use. This has led to studies with questionable ethical actions, which casts a shadow on all fields that use big data. An example is the "Tastes, Ties, and Time" study in 2007, where the researchers published an anonymized dataset of a group of university students and a codebook with information about the dataset; the dataset was identifiable from the codebook (8). Similarly, in 2012, evidence of online emotional contagion was sought, without prior consent, by manipulating the Facebook news feed of thousands of people to see if doing so changes sentiments in individuals' posts (9).

In this article, we explore issues to do with traditional ethical frameworks in relation to research based on AI, particularly in the field of public health and digital epidemiology. We then present ethical frameworks that allow scientists and policy makers to use data from social media and their applications.

## Contemporary ethics

In contemporary science, researchers need ethical approval for the use of human data. This very criterion is the main problem in big data–based research. It raises a seemingly simple question: Does a post or a Tweet represent human data or text data? (10). Several issues and points of view arise from this question, leading to a necessary debate given that the popularity of using social media data is increasing in several scientific fields, including digital epidemiology.

Currently, studies that use social media data are usually perceived as outside the scope of ethics committees' evaluation because these data are commonly not considered to be human data (11,12). Many researchers, policy makers and practitioners assume that they can use open-source data, for example, Tweets, public posts on Facebook, public photos on Instagram and Google Trends queries, which do not require passwords to access (8,13). However, for many users of social media, posting publicly does not equate with giving their consent for the post to be used for research (8,11,12). This issue is not covered by existing ethical review mechanisms (14).

Furthermore, the ease of access to social media data (in the absence of ethical regulations and using rapid data capture via AI) means that the number of data points is often much larger than from traditional epidemiological datasets. Therefore, decisions about the use and implications of social media data can potentially affect more people (14). For example, the number of people accidentally or maliciously reidentified in a Twitter database is only limited by the resources used to compile and analyse the database, which is far less than traditional surveillance systems (14).

### Informed consent

Informed consent in the way it exists in contemporary ethics fits poorly with social media data. Firstly, it is almost impossible to obtain the informed consent of people whose data contribute to digital epidemiology because there are often insufficient resources to contact such high numbers of people who can be living anywhere (15).

Secondly to obtain informed consent, scientists need to confirm the identity of the social media users (16). There is no way to ensure that the person behind the social media profile is who they claim to be or to confirm whether the social media post was not generated by a bot (i.e. "robot" responsible for computer-generated social media posts). Because of this complication, some researchers consider consent to the terms and services of a social media platform, which users must give to use the platform, to be a surrogate for informed consent (16). However, users often do not read the terms and services or understood them well (17–19); nor do these stipulate the terms and conditions under which the data will be used for research, which calls into question the legitimacy and integrity of using terms and services as a surrogate for informed consent. Many "participants" in digital epidemiology are not aware that their data were collected or used (20).

### Privacy and anonymization issues

We are becoming increasingly reliant on technology to structure and analyze the data proliferating in our digital societies. Data mining helps researchers find complex and unintuitive data patterns. However, data mining methods can also reveal confidential information from seemingly harmless social media data, for example, political affiliations (12,21). In addition, Wang *et al.* (22) reported being able to identify people's sexual orientation by processing pictures of people from a dating website.

An anonymized dataset is the minimal requirement to protect the identity of subjects in social science (23) or in traditional epidemiology (20). According to the Common Rule, also known as 45 CFR 46 Subpart A, the principal regulation for human research from the Department of Health and Human Services of the United States (24), 17 identifiers need to be removed to consider a dataset anonymized. These include, among others, name, location of residence, all dates except the year and biometric identifiers (25). The Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Social Sciences and Humanities Research Council (SSHRC), identify similar identifiers (26). However, removing the 17 Common Rule identifiers is often not enough to ensure a dataset is anonymized.

This is because social media data are highly complex (i.e. have high dimensionality). Many non-traditional attributes can enable identification, such as reidentification from assessing the structure of the social networks (i.e. human connections) from multiple social media platforms (15,27). The advancements in AI algorithms and computational power to extract information and assess patterns means it is no longer possible to have anonymous databases (28,29). Many examples in the scientific literature demonstrate this issue by reidentifying an anonymized and subsequently published dataset (12,21).

## The common good

The common good takes roots in the utilitarian vision of ethics. In this vision, the common good that research can do is considered versus the potential harm to individuals. A certain level of harm can be tolerated if the result is "positive morality". In the context of social media, the harm is mostly an invasion of privacy (30). People are more willing to sacrifice their privacy if they perceive that usage of their data will benefit the common good (31,32). For the most enthusiastic social media users in the Mikal *et al.* study (31), "it's cool when it's stuff [...] like the flu, because then that's how [public health decision-makers] know to get the vaccines to a place." Similarly, for the social media users in the Golder *et al.* study (32), it "could give a voice to patients and others groups, uncover true prevailing issues, and improve patient care." Factors that influence people's compliance in sharing their data for the common good include the type of research and the researchers affiliations (i.e. university, company, government) (32–34).

Ultimately, while the majority of people agree with the concept of the common good, there is no agreed-upon threshold for which an invasion of privacy can, and should, be tolerated for public health research.

## New ethical frameworks

New frameworks that respond to new ethical challenges regarding the use of AI for research have been proposed by the Association of Internet Researchers (AoIR) (35) and Zook *et al.* (36) (**Table 1**).

Following a framework can help to legitimize research for the population (37). Since the AoIR framework (35) is accepted in the scientific literature, with the Association being one of the most cited organizations in terms of ethics and big data, scientists may want to use this framework rather than the lesser-known Zook *et al.* framework. However, the Zook *et al.* (36) framework is less restrictive and easier to follow.

Many points in these guidelines are already considerations that public health scientists have to address (e.g. protection of the vulnerable population, the potential harms of the study, the anonymization process). Public health scientists already frequently use highly confidential data. The main difference between social media data and traditional data is the way the data are accessed; the original intent for which the data are produced; and the limited ability for social media users to provide informed consent. The data still represent humans, and can result in unintentional consequences such as identifying the individual behind their social media content. Public health scientists have an obligation to protect the individuals behind their data while balancing this with the common good; this subjective decision is extremely difficult to agree upon.

**Table 1: Proposed ethical frameworks**

| Authors | Guidelines |
|---|---|
| AoIR (35) | 1. Protect vulnerable populations<br>2. Assess potential harm from research studies on a case-by-case basis<br>3. Consider data from humans to be human<br>4. Balance the rights of all involved parties (i.e. the right of privacy for the subject and the right to do research for the scientist)<br>5. The temporal variability of ethical considerations must be resolved when it occurs<br>6. Discuss ethical problems with qualified professionals when these arise |
| Zook *et al.* (36) | 1. Acknowledge that data are people and can do harm<br>2. Recognize that privacy is more than a binary value<br>3. Guard against the reidentification of your data<br>4. Practice ethical data sharing<br>5. Consider the strengths and limitations of your data; big does not automatically mean better<br>6. Debate the tough, ethical choices<br>7. Develop a code of conduct for your organization, research community or industry<br>8. Design your data and systems for auditability<br>9. Engage with the broader consequences of data and analysis practices<br>10. Know when to break these rules |

Abbreviation: AoIR, Association of Internet Researchers

# Discussion

As technology advances rapidly and more research is done with AI and social media data, an established ethical framework is essential to prevent improper use of social media data in public health applications. Researchers in public health, computer science and ethics need to come together to develop a framework that will help scientists conduct responsible research. In general, existing frameworks have been developed for use in every scientific field. Public health-related decisions can have an important impact on the population, however, going as far as to restrict the freedom of movement of persons in the case of a highly infectious disease, as an example (20).

The REB is an important part of the process to ensure the research is within the ethical framework. Inherent in using open-source social media data is that people do not know, or do not have the opportunity to consent, with their data being used. Thus, the REB provides the means to defend the safety, dignity and rights of the participants as stipulated through the ethical framework.

The REB and ethical framework are also needed to address the limitations of social media data. Many social media platforms are available, and the predominance in their use can differ by location. For example, Twitter and Facebook are used extensively in Western countries but banned in the People's Republic of China; the Chinese government authorizes the use of Sina Weibo and WeChat as the respective Twitter and Facebook equivalents. Furthermore, the demographics of use can vary among applications. Older generations tend to use Twitter and Facebook, while younger generations tend to use Snapchat, Instagram and TikTok. This is known as the digital divide (38). Some profiles may be underrepresented (e.g. children and elderly), depending of the social media platforms.

### Conclusion

The ethical issues to do with using social media data for AI applications in public health research centre around whether these data are considered human. Current ethical frameworks are inadequate for public health research. To prevent further misuse of social media data, we argue that considering social media to be human would facilitate an REB process that ensures the safety, dignity and rights of social media data providers. We further propose that there needs to be more consideration towards the balance between the common good and the intrusion of privacy. Collaboration between ethics researchers and digital epidemiologists is needed to develop ethics committees, guidelines and to oversee research in the field.

## Authors' statement

JPG — Writing–original draft, project administration, conceptualization
VN — Writing–reviewing & editing, conceptualization, supervision
NJ — Writing–reviewing & editing, conceptualization, supervision
EER — Writing–reviewing & editing, conceptualization, supervision

### Conflict of interest

None.

## Acknowledgements

The authors would also like to acknowledge S de Montigny, N Barrette and P Gachon for their comments.

## Funding

This work is supported by the Public Health Agency of Canada.

## References

1. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457(7232):1012–4. DOI PubMed

2. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P; Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. PLoS One 2013;8(3):e57252. DOI PubMed

3. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018;14(1):1–5. DOI PubMed

4. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015;43(6):563–71. DOI PubMed

5. Yousefinaghani S, Dara R, Poljak Z, Bernardo TM, Sharif S. The assessment of Twitter's potential for outbreak detection: avian influenza case study. Sci Rep 2019;9(1):18147. DOI PubMed

6. Rangarajan P, Mody SK, Marathe M. Forecasting dengue and influenza incidences using a sparse representation of Google trends, electronic health records, and time series data. PLOS Comput Biol 2019;15(11):e1007518. DOI PubMed

7. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017;107(1):e1–8. DOI PubMed

8. Zimmer M. "But the data is already public": on the ethics of research in Facebook. Ethics Inf Technol 2010;12(4):313–25. DOI

9. Jouhki J, Lauk E, Penttinen M, Sormanen N, Uskali T. Facebook's emotional contagion experiment as a challenge to research ethics. Media Commun 2016;4(4):75–85. DOI

10. Buchanan E. Considering the ethics of big data research: A case of Twitter and ISIS/ISIL. PLoS One 2017;12(12):e0187155. DOI PubMed

11. Fiesler C, Proferes N. "Participant" perceptions of Twitter research ethics. Social Media Soc 2018;4(1):14. DOI

12. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One 2018;13(10):e0204937. DOI PubMed

13. Gehner M, Oughton D. Ethical challenges in social media engagement and research: considerations for code of engagement practices. J Radiol Prot 2016;36(2):S187–92. DOI PubMed

14. Lee EC, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. J Infect Dis 2016;214 suppl_4:S409–13. DOI PubMed

15. Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and epistemology in big data research. J Bioeth Inq 2017;14(4):489–500. DOI PubMed

16. Webb H, Jirotka M, Stahl BC, Housley W, Edwards A, Williams ML, Procter R, Rana OF, Burnap P. The Ethical Challenges of Publishing Twitter Data for Research Dissemination. Proceedings of the 2017 ACM Web Science Conference. 2017, New York: Assoc Computing Machinery; 339–48.

17. Fiske ST, Hauser RM. Protecting human research participants in the age of big data. Proc Natl Acad Sci USA 2014;111(38):13675–6. DOI PubMed

18. Saqr M. Big data and the emerging ethical challenges. Int J Health Sci (Qassim) 2017;11(4):1–2. PubMed

19. Hesse A, Glenna L, Hinrichs C, Chiles R, Sachs C. Qualitative research ethics in the big data era. Am Behav Sci 2019;63(5):560–83. DOI

20. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. PLOS Comput Biol 2015;11(2):e1003904. DOI PubMed

21. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annu Rev Public Health 2018;39:95–112. DOI PubMed

22. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. J Pers Soc Psychol 2018;114(2):246–57. DOI PubMed

23. Phillips A, Borry P, Shabani M. Research ethics review for the use of anonymized samples and data: A systematic review of normative documents. Account Res 2017;24(8):483–96. DOI PubMed

24. Shade J, Coon H, Docherty AR. Ethical implications of using biobanks and population databases for genetic suicide research. Am J Med Genet B Neuropsychiatr Genet 2019;180(8):601–8. DOI PubMed

25. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010 Sep;10(9):3–11. DOI PubMed

26. Canadian Institutes of Health Research. Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council, Tri-Council policy statement: ethical conduct for research involving humans – TCPS 2 (2018). Ottawa (ON): Government of Canada; 2018 Dec. https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf

27. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy chal-lenges. Swiss Med Wkly 2018;148:w14571. DOI PubMed

28. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071. DOI PubMed

29. de Montjoye YA, Radaelli L, Singh VK, Pentland AS. Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata. Science 2015;347(6221):536–9. DOI PubMed

30. Herron M, Sinclair M, Kernohan WG, Stockdale J. Ethical issues in undertaking internet research of user-generated content: a review of the literature. Evid Based Midwifery 2011;9(1):9–15.

31. Mikal J, Hurst S, Conway M. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med Ethics 2016;17(1):22. DOI PubMed

32. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017;19(6):e195. DOI PubMed

33. Beninger K, Fry A, Jago N, Lepps H, Nass L, Silvester H. Research using social media; users' views. London (UK): NatCen Social Research; 2014. https://www.researchgate.net/publication/261551701_Research_using_Social_Media_Users%27_Views

34. Golder S, Scantlebury A, Christmas H. Understanding public attitudes toward researchers using social media for detecting and monitoring adverse events data: multi methods study. J Med Internet Res 2019;21(8):e7081. DOI PubMed

35. Markham A, Buchanan E. Ethical decision-making and internet research: recommendations from the AoIR Ethics Working Committee (Version 2.0). Association of Internet Researchers; 2012.

36. Zook M, Barocas S, Boyd D, Crawford K, Keller E, Gangadharan SP, Goodman A, Hollander R, Koenig BA, Metcalf J, Narayanan A, Nelson A, Pasquale F. Ten simple rules for responsible big data research. PLOS Comput Biol 2017;13(3):e1005399. DOI PubMed

37. Shilton K, Sayles S. "We aren't all going to be on the same page about ethics:" Ethical practices and challenges in research on digital and social media. In: Bui TX, Sprague RH, editors. Proceedings of the 49th Annual Hawaii International Conference on System Sciences. 2016, IEEE Computer Soc: Los Alamitos. 1909–18.

38. Hargittai E. Digital Na(t)ives? Variation in internet skills and uses among members of the "Net Generation". Sociol Inq 2010;80(1):92–113. DOI

OVERVIEW

# Towards automating systematic reviews on immunization using an advanced natural language processing–based extraction system

David Begert[1], Justin Granek[1], Brian Irwin[1], Chris Brogly[2]*

## Abstract

Evidence-informed decision making is based on the premise that the entirety of information on a topic is collected and analyzed. Systematic reviews allow for data from different studies to be rigorously assessed according to PICO principles (population, intervention, control, outcomes). However, conducting a systematic review is generally a slow process that is a significant drain on resources. The fundamental problem is that the current approach to creating a systematic review cannot scale to meet the challenges resulting from the massive body of unstructured evidence. For this reason, the Public Health Agency of Canada has been examining the automation of different stages of evidence synthesis to increase efficiencies.

In this article, we present an overview of an initial version of a novel machine learning–based system that is powered by recent advances in natural language processing (NLP), such as BioBERT, with further optimizations completed using a new immunization-specific document database. The resulting optimized NLP model at the core of this system is able to identify and extract PICO-related fields from publications on immunization with an average accuracy of 88% across five classes of text. Functionality is provided through a straightforward web interface.



**Affiliations**

[1] Xtract AI, Vancouver, BC

[2] Centre for Immunization and Respiratory Infectious Diseases, Public Health Agency of Canada, Ottawa, ON

***Correspondence:**
christopher.brogly@canada.ca

***Suggested citation:*** Begert D, Granek J, Irwin B, Brogly C. Towards automating systematic reviews on immunization using an advanced natural language processing–based extraction system. Can Commun Dis Rep 2020;46(6):174–9. https://doi.org/10.14745/ccdr.v46i06a04

***Keywords:*** automation, natural language processing, NLP, data extraction, systematic reviews, machine learning

## Introduction

Evidence-based medicine relies on systematic reviews as key sources of information on a variety of topics (1) because these provide rigorous assessments and analyses of data from different studies. Although rapid publication of relevant, high-quality systematic reviews is ideal, in practice the publication process has generally been slow (1,2). This is largely due to the massive amount of unstructured information that must be filtered. Synthesizing key information from multiple articles to create a systematic review requires considerable amounts of experts' time. Publication times often exceed one year (3), and costs run into hundreds of thousands of dollars (4).

Machine learning methods have previously been identified for the automation of systematic reviews (1,5). These allow for the development of software systems capable of automatically identifying distinct types of textual information, providing there are enough examples to learn to do this from. Natural language processing (NLP) methods have also been identified for the automation of systematic reviews (1,5) as these methods analyze written text through statistical and/or knowledge-based approaches, allowing for the identification of key items and patterns.

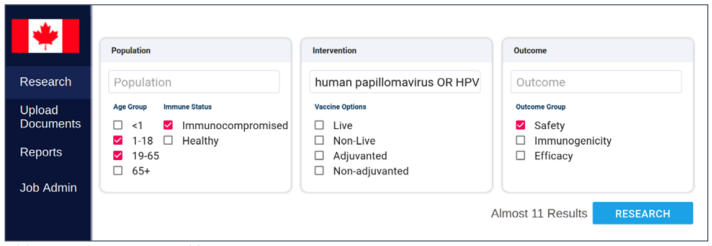
## Background

The Public Health Agency of Canada has been examining the automation of aspects of evidence synthesis based on PICO principles (population, intervention, control, outcomes) to eliminate some of the barriers to obtaining systematic reviews results, namely, direct involvement of experts, time and cost. To do this, the Agency collaborated with Xtract AI (Vancouver, British Columbia) to develop a system that uses state-of-the-art machine learning and NLP to focus on extracting the PICO principles from immunization-specific articles. The functionality of our system is a result of its learning to review articles in a database composed of 249 immunization-specific articles with manually labelled PICO elements. Once the system's accuracy at extracting relevant PICO-related text from previously unseen articles is shown to be high, we can rely on it to carry out work

that would normally be completed manually. The aim was to develop the system so that far fewer articles need to be manually reviewed.

In this article, we use the terms "NLP model" and "system." Strictly speaking, the "NLP model" refers to the collection of machine learning and NLP methods used to process immunization-related documents. The "system" refers to the NLP model with a web-based user interface that allows easy use of the model (see **Figure 1**). Although the NLP model is the core of the system, performing text extraction and prediction tasks, we use the term "system" more often in this article due to the interdependencies of the NLP model and system components.

**Figure 1: Screen capture of the homepage of the web interface of the NLP-based extraction system showing an example of a search query**




Abbreviation: NLP, natural language processing

The system was designed to automatically extract 27 PICO-relevant classes of text from previously unseen immunization-specific articles. To test this initial version of the system we measured performance based on the system's ability to identify text about the main vaccine, the study type, the population health status and the outcome as well as the outcome's descriptive text from 40 previously unseen immunization-specific articles. We considered these five classes to be appropriate means of measuring the initial performance of the extraction of PICO-related text because they cover a wide range of texts and we suspected they were more varied than many of the other text classes the system can extract. **Table 1** shows the five main classes of text used to measure the performance of the system.

At the time of writing, the average accuracy of the system across these five key text classes was 88%. A summary of the accuracy results is shown in Table 1. To achieve this degree of accuracy, the system learned from 209 of the 249 examples in our document database. The automation task was tested using the remaining 40 examples in the document database. Two versions of these 40 test documents existed; the first version was labelled by an expert who scanned the documents for instances of the 27 text classes, and the other was unlabelled. To test the PICO-related text extraction capability of the system, the system processed the unlabelled versions of the test documents. We then compared the system's automated extractions to the text in the expert's labels. If the system extracted text that was comparable in quality to one of the expert's labelled texts, this was counted as a success that contributed to the 88% average accuracy score.

**Table 1: Key text extraction classes with examples and accuracy scores**

| Text class | Description | Extracted example | Accuracy score (%) |
|---|---|---|---|
| Main vaccine | What vaccine is this article about? | Quadrivalent human papillomavirus vaccine, or heptavalent pneumococcal vaccine | 90 |
| Study type | What kind of study is this? | Randomized, placebo-controlled trial | 92.5 |
| Population health status | Population health status | HIV-positive | 85 |
| Outcome – adverse event | Any adverse outcome described in the article | "The most common adverse event was pain; other common events were neurological, gastrointestinal and skin related." | 85 |
| Outcome – description sentence | Any important sentence related to outcomes | Safety/ immunogenicity outcome, e.g. "VE was 93.0% (85.1–97.3) in the TVC-E (Table S1)." | 87.5 |

As development work continues and more expert-labelled immunization-specific articles are added to the document database, we expect to be able to demonstrate similarly high accuracy scores when testing the PICO-related text extraction task on many more documents.

In this article we describe the technical approach to the development of the NLP model and the process by which the NLP model learned to perform this task. We then provide a more detailed analysis of accuracy using several performance measures.

## Technical approach

The NLP model was designed as a multi-class sequence extraction model. A multi-class sequence extraction model works by processing the full text of a previously unseen document and then extracting sequences of text that correspond to each text class it learned to extract. In this case, based on the expert-labelled domains in 209 of our 249 immunization-specific documents, the system extracts up to 27 classes of text. Duplicates for each class may be included.

BioBERT, the biomedical language variant of Bidirectional Encoder Representations from Transformers (BERT), was used as a basis for the NLP model in order to increase performance (6,7).

BERT, a recent development in NLP (7), is essentially a model that has processed and learned from a massive corpus of text. BERT and variants like BioBERT are being increasingly used as the basis for new machine learning and NLP software systems. Their use has resulted in a considerable increase in the accuracy of these systems. BioBERT, as the biomedical language variant of the original BERT model, was considered more appropriate for use in this work.

## Dataset creation

The initial learning data used for the system's NLP model was the evidence-based medicine NLP (EBM-NLP) corpus for PICO extraction. The EBM-NLP corpus contains 5,000 annotated abstracts of medical articles describing clinical randomized controlled trials (8). EBM-NLP corpus annotations labelled key parts of these abstracts, such as description of the participants (e.g. age range, condition), interventions (e.g. pharmacological) and outcomes (e.g. pain, adverse effects or mortality). These fields/classes of text were determined by the EBM-NLP corpus developers. A complete description of the annotation methodology used for the EBM-NLP corpus can be found at https://ebm-nlp.herokuapp.com/annotations.

We used the EBM-NLP corpus to "teach" our system to work with these fields because we needed to be able to extract the same fields/classes of text. However, because the EBM-NLP corpus annotates a small number of text fields that are not immunization-specific, we also generated the 249-document immunization-specific database by annotating data in-house. While the specific amount of time for including and labelling each article in the document database was not noted, adding and labelling a document would be equivalent to the time typically taken to manually review an article that might potentially be included in a systematic review. However, the effort at this stage will ultimately mean less human involvement in the overall systematic review process, as the system's accuracy should show that it is capable of reliably performing the task on its own.

Currently the model processes all types of documents, without pause, but returns nil results for those that do not contain any recognizable text similar to learned examples. We expect future versions of the system to be able to identify documents not related to immunization based on a lack of results. Using the BRAT annotation tool (9), we annotated the title, abstract, methods and results sections of the 249 articles in the new database. The articles were sourced from PubMed Central using a keyword search. We annotated the fields not included in the EBM-NLP corpus, that is, "study type," "main vaccine," "outcomes – adverse event," "outcomes – description sentence" and "population health status." The NLP model understands these text classes because they were manually labelled in the new immunization-specific document database. Examples of these text classes are listed in Table 1.

## System learning process/accuracy testing

The BioBERT-based NLP model was initialized from the original BioBERT model, with subsequent learning using the EBM-NLP corpus. The model then completed more learning using our immunization-specific document database. While 209 samples from the document database were used during this final learning stage, 40 labelled articles were excluded in order to test the system's performance.

Although we focused on testing five key text classes in this article, the system only needs to extract and output a small amount of simple text to test several of the 27 extractable text classes (see **Table 2** for examples). For the remaining classes, the system does not output the extracted text on finding it but rather flags the article as "true" or "false" depending on the content of the research article. The EBM-NLP corpus used in the initial training of the NLP model contained examples for some of these classes. These examples were expected to allow the NLP model to identify anything included in the EBM-NLP corpus accurately.

**Table 2: Additional text extraction classes with examples[a]**

| Text Class[b] | Extracted example |
|---|---|
| Safety | True |
| Efficacy | True |
| Pharmacological[b] | Quadrivalent human papillomavirus (types 6, 11, 16 and 18) recombinant vaccine 0.5 mL intramuscularly |
| Condition | HIV-infected |
| Country | Mali |
| Age | Adults aged 27 years or older |
| Sample size | 535 |
| Sex | Women |

[a] The system also extracts "vaccine_pathogen_target_main," "adjuvanted," "immunogenicity," "immunocompromised," "healthy," "non-live," "non-adjuvanted," "live," "sex," "pregnancy," "doi," "score," "abstract," "methods," "results," "ai_version" and "keywords," but these are not listed here for reasons of brevity

[b] For some text classes, e.g. pharmacological, the system was shown many thousand learning examples via the evidence-based medicine natural language processing corpus

# Evaluation

To evaluate the prediction performance for the five key text classes, we computed common measures of performance for machine learning and NLP-based systems, namely, precision, recall and F1 score (another measure of accuracy based on precision and recall). We further computed the number of successes, errors and general accuracy percentage. All performance measures are defined in **Table 3**. The measures listed apply to entire documents. The number of true positives, true negatives, false positives and false negatives are used to compute the successes and errors. Those numbers are the result of the degree of correctness of text extractions made by the

system. The general accuracy percentages shown in Table 1 are calculated based on the successes and errors.

### Table 3: Definitions of model performance measures

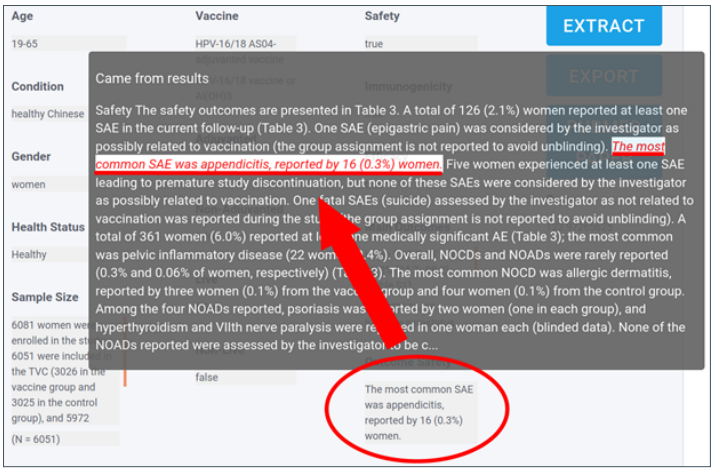
| Measure of accuracy | Means or formula of calculation | Meaning of results |
|---|---|---|
| TP | Number of documents | Correctly identified documents<br>Higher scores are better |
| TN | Number of documents | Correctly identified documents<br>Higher scores are better |
| FP | Number of documents | Documents incorrectly identified<br>Lower scores are better |
| FN | Number of documents | Documents incorrectly identified<br>Lower scores are better |
| Successes | $TP + TN$ | The sum of correctly identified documents<br>Higher scores are better |
| Errors | $FP + FN$ | The sum of incorrectly identified documents<br>Lower scores are better |
| General accuracy percentage | $\frac{(TP + TN)}{(TP + TN + FP + FN)} * 100$ | Overall accuracy from 0% to 100% |
| Precision | $\frac{TP}{(TP + FP)}$ | Measure of document retrieval<br>Scores from 0.0 to 1.0, the higher the better |
| Recall | $\frac{TP}{(FN + TP)}$ | Measure of document retrieval<br>Scores from 0.0 to 1.0, the higher the better |
| F1 score | 2*(P*R) / (P+R) | Scores from 0.0 to 1.0, the higher the better |

Abbreviations: FN, false negative; FP, false positive; P, precision; R, recall; TN, true negative; TP, true positive

It is important to note that the system extracts free-form text, where the length and content of an extracted prediction can vary greatly from the correct, labelled sequence of text on a test document (for examples, see **Figure 2**). This being the case, it is very important to clearly define what constitutes a success (true positive or true negative). For example, if the extracted prediction is "pneumococcal vaccine" while the correct answer is "heptavalent pneumococcal vaccine," the result may be classified as an error if "heptavalent" is deemed to be too important to be left out of the "main vaccine" text for this document.

If the problems associated with definitions are not addressed, the performance measures have no context. In this initial stage, the extracted predictions were manually inspected for accuracy based on defined criteria. Although not yet completed at the time of writing, we expect the application of accuracy criteria to extracted predictions to be automated in future versions of the system. These accuracy criteria were imposed upon the five key text classes for an extracted prediction to count as a predicted true positive (PTP). Since a document can have many extracted predictions, a high number of correct PTPs are needed for the document classification task to be counted as a complete success, or true positive. The easiest way to think about a PTP compared to a true positive is that a PTP is at the text level while a true or false positive is assigned to an entire document based on the number of PTPs. We define the general criteria for this in **Table 4**.

### Figure 2: Screen capture of the NLP model web interface showing an extracted prediction[a]




Abbreviation: NLP, natural language processing
[a] The extracted prediction is underlined in the web interface and circled on the main results panel of the web interface

### Table 4: Custom criteria for predicted true positives and the formulas used to calculate true positives and true negatives

| PTP criteria 1[a] | PTP criteria 2[a] | Requirement for a success (true positive)[b,c] | Requirement for a success (true negative)[b,c] |
|---|---|---|---|
| The EP contains one or more of the labelled answers OR unlabelled but correct answer(s), including all important information | The EP cannot have too much unnecessary information | $\frac{PTP}{(PTP+ FPF + PFN)} > 0.8$<br>>0.5 for Outcome text classes due to subjectivity | PTP+PFP+PFN = 0 |

Abbreviations: EP, extracted prediction; PFN, predicted false negative; PFP, predicted false positive; PTP, predicted true positive
[a] PTPs are determined based on the extracted predictions made by the system from the document text after evaluation by an expert reviewer. There could be many extracted predictions per text class that might count as PTPs or otherwise
[b] PFPs and PFNs are equally important but the numbers of each may vary depending on the article
[c] If the requirements for a true positive or true negative are not met, a false positive may be assigned where PFP > PFN or a false negative may be assigned where PFP < PFN

The full results of the 40-document accuracy testing are shown in **Table 5**.

**Table 5: All performance results for five key classes**

| Performance measure | Key text classes | | | | |
|---|---|---|---|---|---|
| | Main vaccine | Study type | Outcome - adverse event[a] | Outcome - description sentence[a] | Population - health status |
| F1 score[b] | 0.8824 | 0.947 | 0.727 | 0.9315 | 0.75 |
| Precision[c] | 1 | 0.964 | 1 | 0.9444 | 0.9 |
| Recall[d] | 0.7895 | 0.931 | 0.571 | 0.9189 | 0.643 |
| TP[e] | 15 | 27 | 4 | 34 | 9 |
| TN[e] | 21 | 10 | 33 | 1 | 25 |
| FP[f] | 0 | 1 | 0 | 2 | 1 |
| FN[f] | 4 | 2 | 3 | 3 | 5 |
| Successes (TP or TN)[g] | 36 | 37 | 37 | 35 | 34 |
| Errors (FP or FN)[h] | 4 | 3 | 3 | 5 | 6 |
| Accuracy percentage for class, %[i] | 90 | 92.5 | 92.5 | 87.5 | 85 |

Abbreviations: FN, false negative; FP, false positive; P, precision; R, recall; TN, true negative; TP, true positive
[a] Text classes had noticeably imbalanced positive and negative examples. Overall accuracy may be skewed in favour of the group with the greater number of examples. However, an imbalance between examples may also occur in real-world data
[b] Another measure of accuracy based on precision and recall. Scores range from 0.0 to 1.0, the higher the better
[c] Scores range from 0.0 to 1.0, the higher the better
[d] Scores range from 0.0 to 1.0, the higher the better
[e] A measure of correctly identified documents. Higher scores are better
[f] A measure of incorrectly identified documents. Lower scores are better
[g] A measure of the sum of correctly identified documents. Higher scores are better
[h] The sum of incorrectly identified documents. Lower scores are better
[i] Overall accuracy, with scores ranging from 0% to 100%

The system consistently performed well. The success rate was high and the error rate low, which demonstrates overall effectiveness at the PICO extraction task. A balance of both positive and negative test examples was not possible for every text class due to limited data, although a balance may not necessarily reflect real-world performance. For instance, there were many more true negatives for "population – health status" because the articles did not contain any text that could be extracted for this class. Regardless, one issue resulting from this imbalance is that the accuracy scores for these text classes may be skewed in favour of the group (either positive or negative) with more test examples. However, we expect that scores will remain high as this issue is addressed through the expansion of the immunization-specific documents database.

As shown in Figure 1, PICO-related extraction results are accessible through a user-friendly web interface. **Figure 3** shows an example of a completed search displaying results for many of the text classes.

## System limitations

As previously stated, a balance of positive and negative test example groups for all text classes was not possible due to limited data. This may skew the accuracy scores in favour of the group with the higher number of test examples. However, it is important to note that there may be an imbalance between these positive and negative examples on unseen documents in real-world situations.

**Figure 3: Example extraction results for HPV after submitting search terms**


Developing the new immunization-specific document database required some involvement by experts, that is, it was not automated. There was also some manual effort in reviewing extracted predictions from the document text for correctness. This early manual effort is ultimately required to enable automation later.

### Next steps

At the time of writing, the system was still being developed. Future work will include increasing the number of labelled documents in the new immunization-specific document database to improve system learning. The web interface will also continue to be refined. Ideally, the system will identify documents that are not related to immunization and stop processing them immediately to prevent even the brief delay that is currently needed to scan a text. A related system, designed to encompass all biomedical literature (based on the same technology in this article), is also being developed.

Finally, the effectiveness of a more complete system will need to be tested in consultation with public health decision makers.

## Conclusion

We described a system based on machine learning and NLP methods for automating the repetitive manual work of analyzing documents that is part of the systematic review process. This system focuses on immunization-specific documents only. The promising performance results in this initial work demonstrate that there is potential to move away from the manual and laborious approaches of systematic reviews and move towards

automated systems, in an effort to eventually eliminate (or significantly reduce) expert involvement in the repetitious tasks of the process.

The system's overall design presents a promising way for public health decision makers to utilize unstructured data more quickly and economically when making policy decisions and applying the principles of evidence-based medicine. Our unique contribution to this area is the system's ease of use via the straightforward web interface combined with the performance resulting from the application of state-of-the-art machine learning and NLP methods on our new immunization-specific document database.

## Authors' statement

DB — System design/implementation, testing, writing, review, editing
JG — System design/implementation, testing, writing, review, editing
BI — System design/implementation, testing, writing, review, editing
CB — Testing requirements, writing, review, editing

### Conflicts of interest

None.

## Acknowledgements

The authors would like to acknowledge the contribution of Xtract AI and the following people from the Public Health Agency of Canada: O Baclic, M Tunis, M Laplante, K Young, H Swerdfeger, C Doan.

## Funding

This work was supported by the Public Health Agency of Canada.

## References

1. Tsafnat G, Glasziou P, Choong MK, Dunn A, Galgani F, Coiera E. Systematic review automation technologies. Syst Rev 2014;3(74):1-15. DOI PubMed

2. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. Syst Rev 2015;4(78):1-16. DOI PubMed

3. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open 2017;7(2):e012545. DOI PubMed

4. Lau J. Editorial: systematic review automation thematic series. Syst Rev 2019;8(70):1-2. DOI PubMed

5. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. BMC Syst Rev 2019;8(163):1-10 DOI PubMed

6. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb;36(4):1234–40. DOI PubMed

7. Devlin J, Chang MW, Lee K, Toutanova K. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); BERT: pre-training of deep bidirectional transformers for language understanding; 2019-06. NAACL-HLT. Minneapolis, Minnesota (US): Association for Computational linguistics: 2019;4171–86. DOI

8. Nye B, Li JJ, Patel R, Yang Y, Marshall I, Nenkova A, Wallace B. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics; A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. Melbourne (AUS): Association for Computational linguistics; July 15-20, 2018. DOI

9. Stenetorp P, Pyysalo S, Topíc G, Ohta T, Ananiadou S, Tsujii J. BRAT: A web-based tool for NLP-assisted text annotation. Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. 2012 April. Avignon (FR): EACL.

OVERVIEW

# Application of artificial intelligence to the *in silico* assessment of antimicrobial resistance and risks to human and animal health presented by priority enteric bacterial pathogens

Rylan Steinkey[1], Janice Moat[2,3], Victor Gannon[1], Athanasios Zovoilis[3,4,5], Chad Laing[2]*

## Abstract

Each year, approximately one in eight Canadians are affected by foodborne illness, either through outbreaks or sporadic illness, with animals being the major reservoir for the pathogens. Whole genome sequence analyses are now routinely implemented by public and animal health laboratories to define epidemiological disease clusters and to identify potential sources of infection. Similarly, a number of bioinformatics tools can be used to identify virulence and antimicrobial resistance (AMR) determinants in the genomes of pathogenic strains.

Many important clinical and phenotypic characteristics of these pathogens can now be predicted using machine learning algorithms applied to whole genome sequence data. In this overview, we compare the ability of support vector machines, gradient-boosted decision trees and artificial neural networks to predict the levels of AMR within *Salmonella enterica* and extended-spectrum ß-lactamase (ESBL) producing *Escherichia coli*. We show that minimum inhibitory concentrations (MIC) for each of 13 antimicrobials for *S. enterica* strains can be accurately determined, and that ESBL-producing *E. coli* strains can be accurately classified as susceptible, intermediate or resistant for each of seven antimicrobials.

In addition to AMR and bacterial populations of greatest risk to human health, artificial intelligence algorithms hold promise as tools to predict other clinically and epidemiologically important phenotypes of enteric pathogens.



**Affiliations**

[1] National Microbiology Laboratory at Lethbridge, Public Health Agency of Canada, Lethbridge, AB

[2] National Centre for Animal Diseases, Canadian Food Inspection Agency, Lethbridge, AB

[3] Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB

[4] Southern Alberta Genome Sciences Centre, Lethbridge, AB

[5] Canadian Centre for Behavioural Neuroscience, Lethbridge, AB

***Correspondence:**
chad.laing@canada.ca

***Suggested citation:*** Steinkey R, Moat J, Gannon V, Zovoilis A, Laing C. Application of artificial intelligence to the *in silico* assessment of antimicrobial resistance and risks to human and animal health presented by priority enteric bacterial pathogens. Can Commun Dis Rep 2020;46(6):180–5. https://doi.org/10.14745/ccdr.v46i06a05

***Keywords:*** machine learning, bacterial pathogens, whole genome sequence, predictive genomics, antimicrobial resistance

## Introduction

Every year, about one in eight Canadians will be affected by a foodborne illness, resulting in an average of 11,600 hospitalizations and 238 deaths nationwide (1). Animals are often the reservoir for major bacterial pathogens such as *Salmonella enterica* and *Escherichia coli*. These pathogens are associated with both sporadic cases and outbreaks of foodborne disease. Antimicrobial resistance (AMR) among these organisms is a growing concern, with treatment being more difficult and expensive. For example, extended-spectrum ß-lactamase (ESBL) producing *E. coli* are multidrug resistant, with treatment costs up to three times that of non-ESBL-producing *E. coli* (2).

National and provincial public health agencies are very effective at identifying sources and halting exposure to pathogens. Historically, AMR determination has been performed in a wet lab setting (3,4). Two of the most commonly used diagnostic methods are diffusion and dilution tests. Diffusion methods, such as the Kirby–Bauer method, require growing a bacterial lawn in either a disk of known concentration of antimicrobials or a strip with a gradient of concentrations of antimicrobials; the zone of growth inhibition around the antimicrobial is compared with a standard to determine the resistance of the bacteria (3). Dilution methods involve liquid cultures in serial dilution of

each antimicrobial, where growth of the organism is used to determine the minimum inhibitory concentration (MIC) (3,4).

These methods are time consuming because they rely on the growth of bacteria, and expensive because they require trained personnel and specialized equipment to carry out.

Whole genome sequence (WGS) analyses have become integral to public health work flows. *In silico* tests have largely replaced many costly and time-consuming wet lab tests in outbreak response and routine surveillance (5–7). Artificial intelligence is being increasingly used to analyse these datasets.

Artificial intelligence involves training machines to make predictions based on large amounts of data. It has been used in fields as disparate as handwriting recognition (8) and autonomous weapons systems (9).

Supervised machine learning (ML) better describes the application of artificial intelligence to the prediction of bacterial phenotypes based on WGS data. ML algorithms are trained on known data ("features") and subsequently predict or classify unknown data using the trained models. In general, data used for ML training are application specific and can include images or information about weather or outbreaks of infectious disease. Biological data, and in particular WGS data from populations of organisms, provide an extremely large number of features for training ML models and predicting phenotypes of interest. Use of these algorithms in infectious disease research has not yet been fully exploited but holds significant promise.

ML algorithms have been used to predict important phenotypes such as AMR (10,11) and to determine if different groups of pathogens from the same species pose different risks to human health (12–14). The ability to predict important bacterial phenotypes based solely on WGS data would be of enormous benefit to both Canadian public health and the animal agriculture industry.

In this study, we trained three ML models on WGS data to predict the levels of resistance to 13 antimicrobials in *S. enterica* isolates and to classify ESBL-producing *E. coli* strains as susceptible, intermediate or resistant (SIR) to seven antimicrobials.

## Methods

*S. enterica* WGS were collected from the National Center for Biotechnology Information GenBank. These 5,853 sequences were primarily isolated within North America between 2002 and 2017; the data included 63 serotypes with at least five members, along with phenotypic MICs for 13 antimicrobials (15). WGSs were decomposed into sequence substrings, called *k*-mers, of length 11, and their occurrences were counted using Jellyfish (16).To limit the selection of features to those most associated with the phenotype being examined, we used an ANOVA F-value, keeping the top 1,000 *k*-mers most associated with each antimicrobial agent prior to model training. This feature selection allows the model to focus on statistically important *k*-mers, which can improve accuracy and saves substantial amounts of time and computing resources.

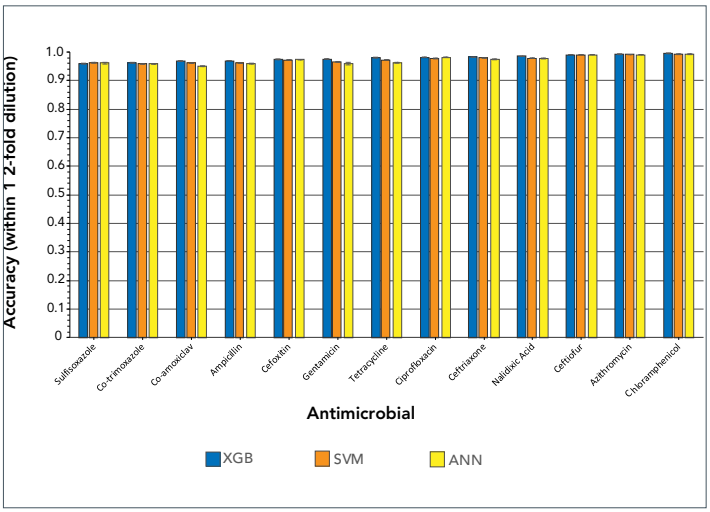
We implemented gradient-boosted decision trees using XGBoost (17) and support vector machines using SciKit-learn (18). Data analyses were conducted using five-fold cross-validation where 80% of the data was used to train a model and the remaining 20% was withheld to evaluate model performance. This was repeated five times, with each 20% being used once for evaluating performance. An average of the accuracy for the five evaluations was calculated for each experimental replicate. Ten separate experimental replicates with random assignment of genomes to each fold were performed, with the total model accuracy and standard deviation calculated from these.

Artificial neural networks were implemented using Keras (19) with a TensorFlow (20) backend and hyperparameter optimizations conducted with Hyperas (21). The five-fold cross-validation for the neural network consisted of a 60-20-20 split for training, hyperparameter optimization and testing, respectively, for each fold. Early stopping mechanisms were used to prevent over-fitting by monitoring diminishing or negative returns with successive training epochs. In addition, a random selection of nodes in the network and their connections were removed via dropout to prevent over-fitting or co-adaptation (22).

As shown in **Figure 1**, MICs were predicted within one dilution with an accuracy of 97.88% (± 1.13) using XGBoost, 97.48% (± 1.20) using support vector machines and 97.16% (± 1.48) using artificial neural networks. XGBoost classifiers averaged a major error and major error rate of 0.19% (± 0.19) and 0.71% (± 0.60), respectively. To prevent inflating model accuracies, co-trimoxazole, ciprofloxacin and ceftriaxone, which had low MIC class diversity, were removed from these averages. XGBoost classifiers trained to predict MICs for a single antimicrobial used eight cores (Intel Xeon Gold 6154 CPU), had a mean training time of 15 minutes and 12 seconds, and peaked at 84.74 GB of random access memory (RAM).

We also examined a set of 2,413 *E. coli* sequences containing ESBL producers, but no MIC data were available for these strains. Instead, they were classified as SIR for seven antimicrobials. The set included bovine, clinical and environmental samples isolated between 1970 and 2017 in Canada, Thailand and the United Kingdom (11,23,24). We analyzed the sequences with the *k*-mer approach described above and used them to train models to classify isolates as SIR for each antimicrobial. The

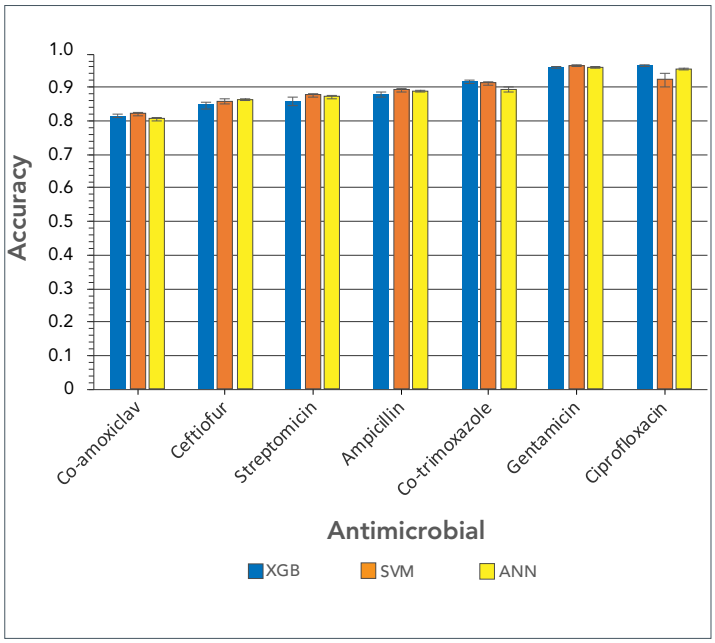
**Figure 1: Accuracies within one two-fold dilution for three machine learning models trained on the top 1,000 11-mers and used to predict minimum inhibitory concentrations for 13 *Salmonella enterica* antimicrobials**




Abbreviations: ANN, artificial neural network; SVM, support vector machine; XGB, XGBoost

average accuracies of the models across the seven antimicrobials were 89.18% (± 5.44) for XGBoost, 89.25% (± 4.43) for support vector machines and 89.18% (± 5.20) for artificial neural networks (**Figure 2**).

**Figure 2: Accuracies of three machine learning models trained on the top 1,000 11-mers, and used to predict susceptible, intermediate and resistant classifications for seven *Escherichia coli* antimicrobials**




Abbreviations: ANN, artificial neural network; SVM, support vector machine; XGB, XGBoost

## Discussion

As we have shown, the ML methods we employed did not rely on specific reference genomes, or *a priori* knowledge of the mechanisms of resistance, but on the classification of organisms into broad phenotypic groups. It is the ML models that identify the underlying genomic differences that are most associated with the phenotype. This has the double benefit of not requiring mechanistic knowledge and has the potential for identifying novel genomic determinants of the phenotype under study. These novel features extracted from the models have enormous potential benefit: as in the case of AMR, they can be used to grow established public databases of resistance mechanisms, and they can be used as potential targets for rapid diagnostics in subsequent *in silico* or wet lab assays.

ML models can rapidly and accurately predict AMR using WGS data, from SIR classification to quantitative MIC values. For AMR predictions, XGBoost models were shown to train faster, use less memory and be more accurate than deep-learning methods. In addition, XGBoost and support vector machine models can be used to determine the specific regions of the genome that are most predictive of a phenotype. This is very difficult with the "black box" implementation of a neural network; however, artificial neural networks still excel in complicated network modelling and therefore should not be excluded from future studies in genomics.

AMR data typically suffer from substantial class imbalance, which can result in high accuracy models that are of no value, such as the case of co-trimoxazole in our *Salmonella* data, where more than 95% of the samples were within one dilution of each other, resulting in a model capable of 95% accuracy without learning anything from the underlying data.

Nguyen *et al.* (10) trained XGBoost regressors on a dataset containing 4,500 non-typhoidal *S. enterica* whole genome sequences (from a larger dataset of 5,278 samples, of which 4,595 were also in our dataset). These models had a cross-validation accuracy of 95% for the same 10 antimicrobials included in our current study. Nguyen *et al.* (10) used a single regressor trained on all 15 antimicrobials at once, which took 51 hours to train and peaked at 1,184 GB of memory on 170 cores (Intel Xeon E5-4669v4 CPU) (10). The XGBoost classifiers trained in our current study improved upon these training times as well as memory usage and accuracy. The XGBoost classifiers did this by creating per-antimicrobial models and initially selecting only the 1,000 most statistically important features. To better compare the accuracies of these models, an independent dataset should be used instead of relying on the reported cross-validation accuracies.

The *E. coli* dataset included 1,935 isolates from a previous study by Moradigaravand *et al.* (11). Their methods required the isolation year for each sequence and data preprocessing in the

form of pan-genome determination and population structure calculation (11). In contrast, our methods required only the genome sequence paired with laboratory-determined resistance phenotype, which allows classification as well as identification of novel regions not currently known to be associated with AMR. The regions could be used for subsequent *in silico* or wet lab diagnostic tests.

While broader classifications, such as SIR, are common for laboratory diagnostics, and useful for establishing treatment guidelines for a bacterial infection, the breakpoint criteria for these categories are established by committees, with some disparity between regions. The prediction of quantified values in the form of MICs will be of most use in future, even if they are subsequently used for classifying bacteria into broader categories such as SIR.

Though the results of these studies are encouraging, over-interpretation of results is a problem with genomic data due to the high number of features used to make predictions relative to the smaller sample size of the number of genomes. This can lead to over-fitting of data and poor performance of models, both of which we have tried to address in the methods of this study (25).

Use of ML has proved successful for AMR prediction in other pathogens, including *Mycobacterium tuberculosis*, where new resistant genetic signatures were identified (26). ML has also proved useful in the identification of novel antimicrobial compounds, which has historically been fraught with high failure rates in pharmaceutical companies (27).

ML research on *S. typhimurium* found that more than 80% of host source could be attributed using protein variants. This result was obtained using support vector machine (SVM), artificial neural networks and Random Forest models (28). What is particularly interesting from this study is the overlap between the animal reservoir and human cases. This indicates that not all isolates of a particular pathogen represent the same disease risk and suggests that more specific points of control could limit human infection. In addition, as more than 60% of human pathogens are of zoonotic origin, ML holds promise for identifying emerging pathogens by analyses of host adaptation of current animal pathogens (29).

Despite the proven usefulness of ML, bacteria are constantly evolving, and so our models, as they are only as good as the data they are trained on. The power of these techniques must be tempered by their judicious use. In addition, class and species-specific models are still required to generate meaningful results, for example, one model per drug per species for predicting AMR (30).

It should be noted that ML does not always accurately capture complex interactions and that improved modelling alone cannot compensate for sampling bias or an incomplete or error-prone dataset.

## Conclusion

As demonstrated in this overview, artificial intelligence has already improved infectious disease identification and characterization, the benefits of which will affect public health and animal health laboratories around the world. For example, genomic regions identified as predictive for specific AMR classes could be used for rapid downstream identification and classification, including *in silico* pipelines and wet lab applications such as polymerase chain reaction.

The near-future promises exciting developments, such as using ML to identify bacteriophages that lyse specific groups of pathogenic bacteria, enabling phage therapy in place of traditional antimicrobials (31). Lastly, "whole phenotype" characterization, with the ability to predict integral membrane protein expression, is becoming more likely (32); and biofilm formation (33).

Despite this, the size of the datasets required to effectively train ML models mean that desktop computers are often incapable of analyzing the data. Those without access to the necessary resources must instead use analytical approaches that reduce the computational burden (34). Fittingly, the use of ML itself has led to an increase in speed of mechanistic models, in some cases over four orders of magnitude (35).

We are just at the beginning of the coupling of vast amounts of genomic data and artificial intelligence, with the promise of new discoveries that will improve most aspects of animal and human health from the burden of enteric bacterial pathogens.

# Authors' statement

RJS — Data curation, formal analysis, methodology, software, validation, visualization, original draft, editing
JM — Data curation, formal analysis, methodology, software, validation, visualization, original draft, editing
VPJG — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, original draft, editing
AZ — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, original draft, editing
CRL — Conceptualization, funding acquisition, methodology, project administration, resources, supervision, validation, original draft, editing

## Conflict of interest

None.

## Funding

JM, AZ, CRL: This work has been supported by the Antimicrobial Resistance – One Health Consortium grant to AZ and CRL from the Alberta Ministry of Economic Development, Trade, and Tourism.

RJS, VPJG, CRL: This work has been supported by the Genomics Research and Development Initiative project on antimicrobial resistance. This work was additionally funded by the Public Health Agency of Canada, the Canadian Food Inspection Agency and the University of Lethbridge.

## References

1. Public Health Agency of Canada. Yearly food-borne illness estimates for Canada. Ottawa (ON): Government of Canada; 2015 (updated 2016-07-05). https://www.canada.ca/en/public-health/services/food-borne-illness-canada/yearly-food-borne-illness-estimates-canada.html

2. Maslikowska JA, Walker SA, Elligsen M, Mittmann N, Palmay L, Daneman N, Simor A. Impact of infection with extended-spectrum ß-lactamase-producing Escherichia coli or Klebsiella species on outcome and hospitalization costs. J Hosp Infect 2016;92(1):33–41. DOI PubMed

3. Schumacher A, Vranken T, Malhotra A, Arts JJ, Habibovic P. In vitro antimicrobial susceptibility testing methods: agar dilution to 3D tissue-engineered models. Eur J Clin Microbiol Infect Dis 2018;37(2):187–208. DOI PubMed

4. Andrews JM. Determination of minimum inhibitory concentrations. J Antimicrob Chemother 2001;48 (1 Suppl 1):5–16. DOI PubMed

5. Collineau L, Boerlin P, Carson CA, Chapman B, Fazil A, Hetman B, McEwen SA, Parmley EJ, Reid-Smith RJ, Taboada EN, Smith BA. Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: a review of opportunities and challenges. Front Microbiol 2019;10:1107. DOI PubMed

6. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K, Wise M, Gerner-Smidt P. Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. Foodborne Pathog Dis 2019;16(7):504–12. DOI PubMed

7. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. Front Public Health 2019;7:242. DOI PubMed

8. Muehlberger G, Seaward L, Terras M, Ares Oliveira S, Bosch V, Bryan M, Colutto S, Déjean H, Diem M, Fiel S, Gatos B, Greinoecker A, Grüning T, Hackl G, Haukkovaara V, Heyer G, Hirvonen L, Hodel T, Jokinen M, Kahle P, Kallio M, Kaplan F, Kleber F, Labahn R, Lang EM, Laube S, Leifert G, Louloudis G, McNicholl R, Meunier JL, Michael J, Mühlbauer E, Philipp N, Pratikakis I, Puigcerver Pérez J, Putz H, Retsinas G, Romero V, Sablatnig R, Sánchez JA, Schofield P, Sfikas G, Sieber C, Stamatopoulos N, Strauß T, Terbul T, Toselli AH, Ulreich B, Villegas M, Vidal E, Walcher J, Weidemann M, Wurster H, Zagoris K. Transforming scholarship in the archives through handwritten text recognition: transkribus as a case study. J Doc 2019;75:954–76. DOI

9. Sharkey A. Autonomous weapons systems, killer robots and human dignity. Ethics Inf Technol 2019;21:75–87. DOI

10. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. J Clin Microbiol 2019;57(2):e01260-18. DOI PubMed

11. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLOS Comput Biol 2018;14(12):e1006258. DOI PubMed

12. Fisch D, Yakimovich A, Clough B, Wright J, Bunyan M, Howell M, Mercer J, Frickel E. Defining host-pathogen interactions employing an artificial intelligence workflow. eLife 2019;8:e40560. DOI PubMed

13. Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL. Support vector machine applied to predict the zoonotic potential of E. coli O157 cattle isolates. Proc Natl Acad Sci USA 2016;113(40):11312–7. DOI PubMed

14. Lupolova N, Dallman TJ, Holden NJ, Gally DL. Patchy promiscuity: machine learning applied to predict the host specificity of Salmonella enterica and Escherichia coli. Microb Genom 2017;3(10):e000135. DOI PubMed

15. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2013;41(Database issue):D36–42. DOI PubMed

16. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 2011;27(6):764–70. DOI PubMed

17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York (NY): ACM; 2016. pp. 785–94. DOI

18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30. http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

19. Chollet F. Keras. GitHub repository; 2015. https://github.com/fchollet/keras

20. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y, Zheng X. TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation. 2016 Nov 2–4. Savannah (GA): OSDI 16. pp. 265–83. https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

21. Pumperla M. Keras + Hyperopt: A very simple wrapper for convenient hyperparameter optimization: Maxpumperla/Hyperas. 2019 (accessed 2020-03-25). http://maxpumperla.com/hyperas/

22. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929–58. http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf

23. Runcharoen C, Raven KE, Reuter S, Kallonen T, Paksanont S, Thammachote J, Anun S, Blane B, Parkhill J, Peacock SJ, Chantratita N. Whole genome sequencing of ESBL-producing Escherichia coli isolated from patients, farm waste and canals in Thailand. Genome Med 2017;9(1):81. DOI PubMed

24. Kallonen T, Brodrick HJ, Harris SR, Corander J, Brown NM, Martin V, Peacock SJ, Parkhill J. Systematic longitudinal survey of invasive Escherichia coli in England demonstrates a stable population structure only transiently disturbed by the emergence of ST131. Genome Res 2017;27(8):1437–49. DOI PubMed

25. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer Science & Business Media; 2009. DOI

26. Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, Heckmann D, Anand A, Yang L, Nizet V, Monk JM, Palsson BO. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. Nat Commun 2018;9(4306):1-9. DOI PubMed

27. Ivanenkov YA, Zhavoronkov A, Yamidanov RS, Osterman IA, Sergiev PV, Aladinskiy VA, Aladinskaya AV, Terentiev VA, Veselov MS, Ayginin AA, Kartsev VG, Skvortsov DA, Chemeris AV, Baimiev AK, Sofronova AA, Malyshev AS, Filkov GI, Bezrukov DS, Zagribelnyy BA, Putin EO, Puchinina MM, Dontsova OA. Identification of novel antibacterials using machine learning techniques. Front Pharmacol 2019;10:913. DOI PubMed

28. Lupolova N, Lycett SJ, Gally DL. A guide to machine learning for bacterial host attribution using genome sequence data. Microb Genom 2019 Dec;5(12):5. DOI PubMed

29. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. Nat Rev Genet 2018 Sep;19(9):549–65. DOI PubMed

30. Hicks AL, Wheeler N, Sánchez-Busó L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLOS Comput Biol 2019 Sep;15(9):e1007349. DOI PubMed

31. Leite DMC, Brochet X, Resch G, Que Y-A, Neves A, Peña-Reyes C. Computational prediction of inter-species relationships through omics data analysis and machine learning. BMC Bioinformatics 2018;19(S14 Suppl 14):420. DOI

32. Saladi SM, Javed N, Müller A, Clemons WM Jr. A statistical model for improved membrane protein expression using sequence-derived features. J Biol Chem 2018 Mar;293(13):4913–27. DOI PubMed

33. Yan J, Deforet M, Boyle KE, Rahman R, Liang R, Okegbe C, Dietrich LE, Qiu W, Xavier JB. Bow-tie signaling in c-di-GMP: machine learning in a simple biochemical network. PLOS Comput Biol 2017 Aug;13(8):e1005677. DOI PubMed

34. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci Rep 2019;9(1):4071. DOI PubMed

35. Wang S, Fan K, Luo N, Cao Y, Wu F, Zhang C, Heller KA, You L. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. Nat Commun 2019 Sep;10(1):4354. DOI PubMed

OVERVIEW

# Application of natural language processing algorithms for extracting information from news articles in event-based surveillance

Victoria Ng[1]*, Erin E Rees[1], Jingcheng Niu[2], Abdelhamid Zaghlool[3], Homeira Ghiasbeglou[3], Adrian Verster[4]

## Abstract

The focus of this article is the application of natural language processing (NLP) for information extraction in event-based surveillance (EBS) systems. We describe common information extraction applications from open-source news articles and media sources in EBS systems, methods, value in public health, challenges and emerging developments.

***Suggested citation:*** Ng V, Rees EE, Niu J, Zaghlool A, Ghiasbeglou H, Verster A. Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. Can Commun Dis Rep 2020;46(6):186–91. https://doi.org/10.14745/ccdr.v46i06a06

***Keywords:*** natural language processing, NLP, event-based surveillance, algorithms, information extraction, open-source data



**Affiliations**

[1] National Microbiology Laboratory, Public Health Agency of Canada

[2] Department of Computer Science, University of Toronto, Toronto, ON

[3] Centre for Emergency Preparedness and Response, Public Health Agency of Canada

[4] Food Directorate, Health Canada, Ottawa, ON

***Correspondence:**
victoria.ng@canada.ca

## Background

Natural language processing (NLP) methods enable computers to analyse, process and derive meaning from human discourse. Although the field of NLP has been around since the 1950s (1), progress in technology and methods in recent years have made NLP applications easier to implement, with some tasks outperforming human performance (2). There are many day-to-day applications of NLP including machine translation, spam recognition and speech recognition. NLP is a powerful tool in health care because of the large volumes of text data, for example, electronic health records, being produced. Indeed electronic health records have already been the focus of NLP applications, including detecting melanocytic proliferations (3,4), the risk of dementia (5) and neurological phenotypes (6). But NLP applications in health care extend beyond electronic health records, for example, it is possible to identify people with Alzheimer's disease based on their speech patterns (7).

The focus of this article is the application of NLP for information extraction in event-based surveillance (EBS) systems. We describe common information extraction applications from open-source news articles and media sources in EBS systems, methods, value in public health, challenges and emerging developments.

EBS systems mine the Internet for open-source data, relying on informal sources (e.g. social media activity) and formal sources (e.g. media or epidemiological reports from individuals, media outlets and/or health organizations) to help detect emerging threats (8). Operational systems include the Public Health Agency of Canada's Global Public Health Intelligence Network (9), HealthMap (10) and the World Health Organization's Epidemic Intelligence from Open Sources (11). Due to the growing volume, variety and velocity of digital information, a wealth of unstructured open-source data is generated daily, mainly as spoken or written communication (9). Unstructured open-source data contains pertinent information about emerging threats that can be processed to extract structured data from the background noise to aid in early threat detection (12). For EBS systems, this includes information about what happened (threat classification; number of cases), where it happened (geolocation) and when it happened (temporal information). The ability to identify this information allows governments and researchers to monitor and respond to emerging infectious disease threats.

One of the challenges in infectious disease surveillance, such as COVID-19, is that there is an immense amount of text data continuously being generated, and in an ongoing pandemic, this amount can be far more than humans are capable of processing. NLP algorithms can help in these efforts by automating the

filtering of large volumes of text data to triage articles into levels of importance and to identify and extract key pieces of information.

In this article, we discuss some important NLP algorithms and how they can be applied to public health. For a glossary of common technical terminology in NLP, see **Table 1**.

**Table 1: Glossary of common technical terminology in natural language processing**

| Term | Definition |
|---|---|
| (Linguistic) annotation | The association of descriptive or analytic notations with language data, generally performed to generate a corpus for algorithm training |
| Artificial intelligence (AI) | A branch of computer science dealing with the simulation of human intelligence by machines |
| Computational linguistics (CL) | The branch of computer science trying to model human language (including various linguistic phenomena and language related applications) using computational algorithms |
| Corpora (singular – corpus) | A set of articles where the unstructured text has been annotated (labelled) to identify different types of named entity. Corpus are developed for different domains to train ML algorithms to identify named entities (e.g. WikToR corpus of Wikipedia articles for geographic locations, TimeBank corpus of new report documents for temporal information) |
| F1 score | A performance measure used to evaluate the ability of NLP to correctly identify NEs by calculating the harmonic mean of precision and recall: F1 = 2 * Precision * Recall / (Recall + Precision). The F1 score privileges balanced algorithms because the score tends toward the least number, minimizing the impact of large outliers and maximizing the impact of small ones |
| Geocoding | Also known as georesolution, assigns geographic coordinates to toponyms |
| Geoparsing | The combined process of geotagging and geocoding |
| Geotagging | A subset of named NER that identifies geographic entities in unstructured text |
| Machine learning (ML) | The study of computer algorithms that learn patterns from experience. ML approaches may be supervised (the algorithm learn from labelled training samples), unsupervised (the algorithm retrieve patterns from unlabeled data), or semi-supervised (the algorithm perform learning with a small set of labelled data and a large set of unlabeled data) |
| Named entity (NE) | A word or phrase that identifies an item with particular attributes that make it stand apart from other items with similar attributes (e.g. person, organization, location) |
| Natural language processing (NLP) | A subfield of AI to process human (natural) language inputs for various applications, including automatic speech recognition, natural language understanding, natural language generation and machine translation |
| Named entity recognition (NER) | The process of identifying a word or phrase that represents a NE within the text. NER formerly appeared in the Sixth Message Understanding Conference (MUC-6), from which NEs were categorized into three labels: ENAMEX (person, organization, location), TIMEX (date, time) and NUMEX (money, percentage, quantity) |
| Polysemy | The association of a word or phrase with two or more distinct meanings (e.g. a mouse is a small rodent or a pointing device for a computer) |

**Table 1: Glossary of common technical terminology in natural language processing** (*continued*)

| Term | Definition |
|---|---|
| Precision (also known as positive predictive value) | Percentage of named entities found by the algorithm that are correct: (true positives) / (true positives + false positives) |
| Recall | Fraction of the total amount of relevant instances that were actually retrieved (true positives) / (true positives + false negatives) |
| Semi-supervised | Due to the high cost of to creating annotated data, semi-supervised learning algorithms combine the learning from a small set of labelled data (supervised) and a large set of unlabeled data (unsupervised) to achieve the tradeoff between cost and performance |
| RSS feed | RRS stands for Really Simple Syndication or Rich Site Summary, it is a type of web feed that allows users and applications to receive regular and automated updates from a website of their choice without having to visit websites manually for updates |
| Supervised learning | Supervised learning algorithms is the type of ML algorithms that learn from labelled input-output pairs. Features of the input data are extracted automatically through learning, and patterns are generalized from those features to make predictions of the output. Common algorithms include hidden Markov models (HMM), decision trees, maximum entropy estimation models, support vector machines (SVM) and conditional random fields (CRF) |
| Synonyms | Words of the same language that have the same or nearly the same meaning as another |
| Toponym | A NE of the place name for a geographic location such as a country, province and city |
| Unsupervised learning | A type of ML method that does not use labelled data, but instead, typically uses clustering and principal component analytical approaches so that the algorithm can find shared attributes to group the data into different outcomes |

## NLP algorithms and their application to public health

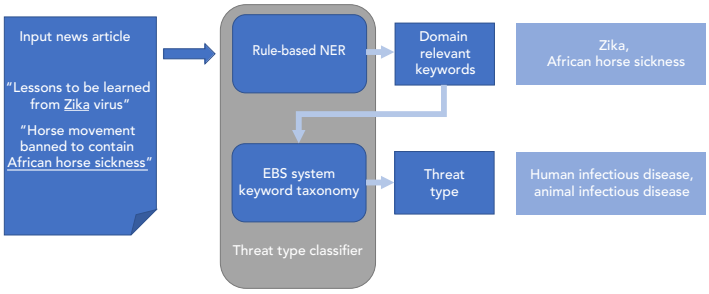
The simplest way to extract information from unstructured text data is by keyword search. Though effective, this ignores the issue of synonyms and related concepts (e.g. nausea and vomiting are related to stomach sickness); it also ignores the context of the sentence (e.g. Apple can be either a fruit or a company). The problem with identifying and classifying important words (entities) based on the structure of the sentence is known as named entity recognition (NER) (13). The most common entities are persons, organizations and locations. Many early NER methods were rule-based, identifying and classifying words with dictionaries (e.g. dictionary of pathogen names) and rules (e.g. using "H#N#" to classify a new influenza strain not found in the dictionary) (14). Synonyms and related concepts can be resolved using databases that organize the structure of words in the language (e.g. WordNet (15)). Newer NER methods use classifications and relationships predefined in corpora to

develop machine learning (ML) algorithms to identify and classify entities (13). For NER, terms are annotated to categories and the algorithm learns how to recognize other examples of the category from the term and surrounding sentence structure. Because language data are converted to word tokens as part of the analysis, NLP algorithms are not limited to languages using the Latin alphabet; they can also be used with character-based languages such as Chinese.

## 1. Article classification (threat type)

Classifying articles by taxonomy keywords into threat types allows EBS system users to prioritize emerging threats. For example, analysts monitoring an event can filter out articles to focus on a specific threat category. Rule-based NER identifies keywords to assign each article to different categories of health threats (e.g. disease type). Keywords are then organized into a predetermined, multilingual taxonomy (e.g. "Zika virus" is a human infectious disease, "African horse sickness" is an animal infectious disease, etc.) that can be updated as new threats are discovered. The taxonomy takes advantage of the structure of the language similar to WordNet (16). This mitigates part of the problem with keyword matching because it allows synonyms and related concepts to stand in for one another (**Figure 1**).

**Figure 1: Article classification**



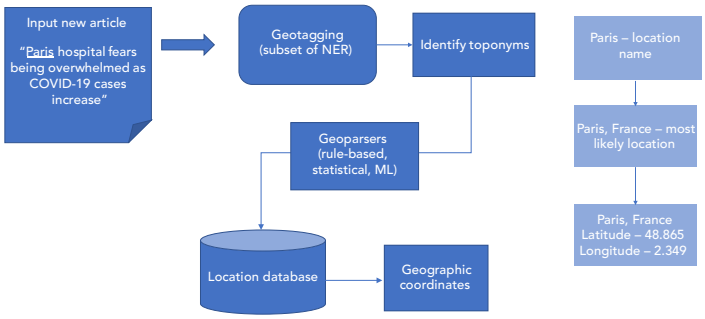

Abbreviations: EBS, event-based surveillance; NER, named entity recognition

## 2. Geoparsing

Identifying places where health-related events are reported from articles can help locate susceptible populations. Geoparsing is the task of assigning geographic coordinates to location entities (i.e. toponyms such as city, country) identified in unstructured text. The process starts with geotagging, a subset of NER for identifying the toponyms, and then geocoding to assign geographic coordinates from a dictionary such as from GeoNames (17). Geoparsers use computational methods that are rule-based, statistical and based on ML. The general approach of geoparsing is to characterize toponyms by a set of features (e.g. toponym name, first and last character position in text, character length). Feature information is then processed by computational methods to link each toponym to a geographic name in a location database (e.g. GeoNames (17)) and then assign the corresponding coordinates (18).

Advancements in geoparsing, like other NLP applications, focus on increasing leverage from unstructured text to resolve ambiguities. One advancement is using semi-supervised learning techniques that utilize programmatically generated corpora to train ML algorithms from larger datasets of annotated examples. Using code to annotate articles is faster and results in larger and more consistent corpora than from human annotation (19). Leveraging more context is also resulting from extending feature information to be topological (spatial relationships among toponyms, e.g. distance to closest neighbouring toponym) (20). A toponym from a phrase like "There are new cases of influenza in London" can be difficult to resolve because there are multiple potential locations. Toponym coordinates can be resolved by assigning a bias towards more populated areas because they are typically mentioned more often in discourse; however, emerging diseases do not always favour highly populated areas (**Figure 2**).

**Figure 2: Geoparsing**




Abbreviations: ML, machine learning; NER, named entity recognition

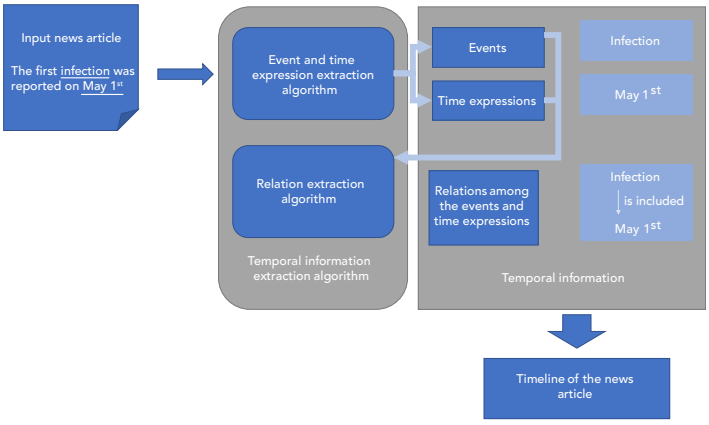
## 3. Temporal information extraction and temporal reasoning

Identifying the timing of events described in articles is necessary for coherent temporal ordering of those events. It is important to be able to differentiate an article reporting on a new event from an article reporting on a previous known event. The most common temporal identifiers in EBS systems are the article publication date and the received/import date (the timestamp for receiving the article into the EBS system). Neither of these dates extract the reported timing of event described in the articles. A subset of NLP—temporal information extraction—has been developed to extract this information. Temporal information extraction is used to identify tokens in text that contain temporal information of relevant events.

Two subtasks of temporal information extraction help resolve ambiguities arising from complicated narratives reporting on multiple events. First, temporal relation extraction focuses on classifying temporal relationships between the extracted events and temporal expressions. Using those relationships, EBS systems can anchor events to time (e.g. in the sentence "the first infection was reported on May 1st," the relation between the event "infection" and the date "May 1st" is used to timestamp the first infection). Second, temporal reasoning (21) focuses on chronological ordering of events through inference.

Multiple temporal information extraction systems have been developed including TimeML (developed for temporal extraction of news articles in finance) (22); ISO-TimeML (a revised version of TimeML) (23); and THYME (developed for temporal extraction in patient records) (24). Results have reached near-human performance (25–28). Based on these annotation standards, an annotation standard for news articles in the public health domain, Temporal Histories of Epidemic Events (THEE), was recently developed for EBS systems by the authors of this article (29) (**Figure 3**).

**Figure 3: Temporal information extraction and temporal reasoning**

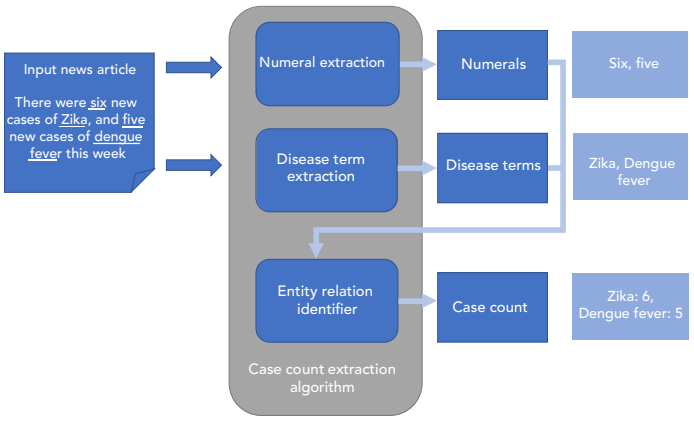



## 4. Case count extraction

Extracting the number of disease cases reported in articles would help EBS system users to monitor and forecast disease progression. Currently, there is no NLP algorithm incorporated into EBS systems capable of this task, however, there are algorithms capable of tackling related tasks that can be leveraged to develop a case count algorithm. News articles in epidemiology frequently mention the occurrence of disease cases (e.g. "There were six new cases of Zika this week") so that identifying cases requires identifying the relationships between a quantitative reference in the text (six new cases) and a disease term (of Zika). Many algorithms already identify relationships between entities in diverse fields, for example, the RelEx algorithm identifies relations between genes that are recorded in MEDLINE abstracts and performs with an F1 of 0.80 (30). Based on the RelEx algorithm, an algorithm has been developed to identify sentences in news articles that report on case counts of foodborne illnesses (31).

The authors of this article are developing and refining this algorithm to extract case count information from sentences that have been identified to contain case count information (**Figure 4**).

**Figure 4: Case count extraction**

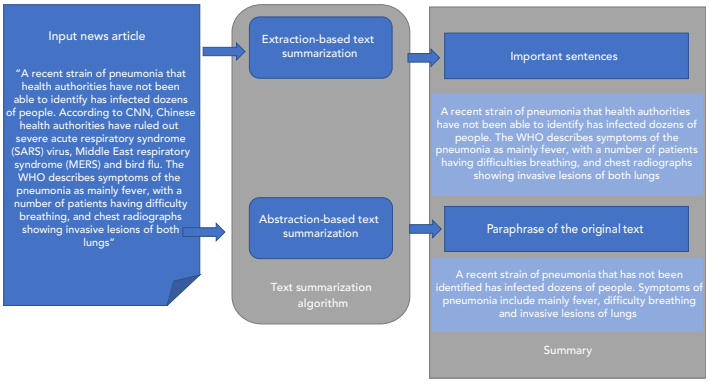



## 5. Automatic text summarization

The goal of text summarization is to quickly and accurately create a concise summary that retains the essential information in the original text. Text summarization in EBS systems would increase the number of articles that can be scanned for threat detection by reducing the volume of text that needs to be read. There are two main types of text summarization: extraction-based and abstraction-based. Extraction-based summarization involves identifying the most important key words and phrases from the text and combining them verbatim to produce a summary. Abstraction-based summarization uses a more sophisticated technique that involves paraphrasing the original text to write new text, thus mimicking human text summarization.

Text summarization in NLP is normally developed using supervised ML models trained on corpora. For both extraction-based and abstraction-based summarization, key phrases are extracted from the source document using methods including part-of-speech tagging, word sequences or other linguistic pattern recognition (32). Abstraction-based summarization goes a step further and attempts to create new phrases and sentences from the extracted key phrases. A number of techniques are used to improve the level of abstraction including deep learning techniques and pre-trained language models (33) (**Figure 5**).

**Figure 5: Automatic text summarization**




Abbreviation: WHO, World Health Organization

## Discussion

NLP has a huge number of potential applications in health care because of the omnipresence of text data. Electronic health records are an obvious source of data for NLP application, but text relevant to health care extends far beyond health records; it includes traditional and social media sources, which are the main sources of data for EBS systems, in addition to official government reports and documents.

As NLP algorithms can interpret text and extract critical information from such diverse sources of data, they will continue to play a growing role in the monitoring and detection of emerging infectious diseases. The current COVID-19 pandemic is an example of where NLP algorithms could be used for the surveillance of public health crises. (This is, in fact, something several co-authors of this article are currently developing).

While NLP algorithms are powerful, they are not perfect. Current key challenges involve grouping multiple sources referring to the same event together and dealing with imperfections in the accuracy of information extraction due to nuances in human languages. Next-generation information extraction NLP research that can improve these challenges include event resolution (deduplication and linkage of the same events together) (34) and advancements in neural NLP approaches such as transformers networks (35), attention mechanism (36) and large-scale language models such as ELMo (37), BERT (38) and XLNet (39) to improve on the current performance of algorithms.

### Conclusion

We have discussed several common NLP extraction algorithms for EBS systems: article classification, which can identify articles that contain crucial information about the spread of infectious diseases; geolocation, which identifies where a new case of the disease has occurred; temporal extraction, which identifies when a new case occurred; case count extraction, which identifies how many cases occurred; and article summarization, which can greatly reduce the amount of text for a human to read.

Although the field of NLP for information extraction is well established, there are many existing and emerging developments relevant to public health surveillance on the horizon. If capitalized, these developments could translate to earlier detection of emerging health threats with an immense impact on Canadians and the world.

### Conflict of interest

None.

## Funding

EE Rees and V Ng are currently co-Principal Investigators for a Canadian Safety and Security Program (CSSP), a federally-funded program from the Department of National Defence, the grant is a three-year grant titled 'Incorporating Advanced Data Analytics into a Health Intelligence Surveillance System' – CSSP-2018-CP-2334.

## References

1. Bates M. Models of natural language understanding. Proc Natl Acad Sci USA 1995 Oct;92(22):9977–82. DOI PubMed

2. Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, Levy O, Bowman S. SuperGLUE: a stickier benchmark for general-purpose language understanding systems. Ithaca (NY): arXiv; 2019 (revised 2020-02-13; accessed 2020-02-24). https://arxiv.org/abs/1905.00537

3. Lott JP, Boudreau DM, Barnhill RL, Weinstock MA, Knopp E, Piepkorn MW, Elder DE, Knezevich SR, Baer A, Tosteson AN, Elmore JG. Population-based analysis of histologically confirmed melanocytic proliferations using natural language processing. JAMA Dermatol 2018 Jan;154(1):24–9. DOI PubMed

4. Nguyen AN, Truran D, Kemp M, Koopman B, Conlan D, O'Dwyer J, Zhang M, Karimi S, Hassanzadeh H, Lawley MJ, Green D. Computer-Assisted Diagnostic Coding: effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings. AMIA Annu Symp Proc 2018 Dec;2018:807–16. PubMed

5. McCoy TH Jr, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH. Stratifying risk for dementia onset using large-scale electronic health record data: A retrospective cohort study. Alzheimers Dement 2020 Mar;16(3):531–40. DOI PubMed

6. Wheater E, Mair G, Sudlow C, Alex B, Grover C, Whiteley W. A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. BMC Med Inform Decis Mak 2019 Sep;19(1):184. DOI PubMed

7. Karlekar S, Niu T, Bansal M. Detecting linguistic characteristics of Alzheimer's Dementia by interpreting neural models. In: Proceedings of NAACL-HLT 2018. New Orleans (LA). June 1–6, 2018.

8. World Health Organization. A guide to establishing event-based surveillance. Manila (PH): WHO Regional Office for the Western Pacific; 2008.

9. Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). Can Commun Dis Rep 2015 Sep;41(9):209–14. DOI PubMed

10. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. J Am Med Inform Assoc 2008 Mar-Apr;15(2):150–7. DOI PubMed

11. World Health Organization. Epidemic Intelligence from Open Sources (EIOS): saving lives through early detection. Geneva (CH): World Health Organization; 2020 (accessed 2020-01-24). https://www.who.int/eios

12. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P; Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in

the early alerting and reporting project for the detection of A/H5N1 influenza events. PLoS One 2013;8(3):e57252. DOI PubMed

13. Nadeau D, Sekine S. A survey of named entity recognition and classification. In: Nadeau D, Sekine S, editors. Named entities: recognition, classification and use. Lingvisticae Investigationes 2007;30(1):3–26.

14. Sekine S, Nabota C. Definition, dictionaries and tagger for extended named entity hierarchy. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004). Lisbon (PT). 26–28 May 2004.

15. Princeton University. WordNet: an electronic lexical database. Princeton (NJ): Princeton University; 2010 (accessed 2020-01-24). https://wordnet.princeton.edu/

16. Miller GA. WordNet: a lexical database for English. Commun ACM 1995;38(11):39–41.

17. GeoNames [database]. 2020. https://www.geonames.org/

18. Santos J, Anastácio I, Martins B. Using machine learning methods for disambiguating place references in textual documents. GeoJournal 2015;80(3):375–92. DOI

19. Gritta M, Pilehvar MT, Limsopatham N, Collier N. What's missing in geographical parsing? Lang Resour Eval 2018;52(2):603–23. DOI PubMed

20. DeLozier G, Baldridge J, London L. Gazetteer-independent toponym resolution using geographic word profiles. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin (TX): AAAI Press; 2015. p. 2382–8.

21. Allen JF. Maintaining knowledge about temporal intervals. In: Weld DS, De Kleer J. Readings in qualitative reasoning about physical systems. 1990, Elsevier; pp. 361–72. DOI

22. Pustejovsky J, Ingria R, Sauri R, Castano JM, Littman J, Gaizauskas RJ, Setzer A, Katz G, Mani I. The specification language TimeML. In: Mani I, Pustejovsky J, Gaizauskas R, editors. The language of time: a reader. 2005. p. 545–58.

23. Pustejovsky J, Kiyong L, Bunt H, Romary L. ISO-TimeML: an international standard for semantic annotation. In: Proceeding from the International Conference on Language Resources and Evaluation 2010. La Valette (MT). 2010 May.

24. Styler WF 4th, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, Erickson B, Miller T, Lin C, Savova G, Pustejovsky J. Temporal annotation in the clinical domain. Trans Assoc Comput Linguist 2014 Apr;2:143–54. DOI PubMed

25. Chambers N. Navytime: event and time ordering from raw text. In: Second Joint Conference on Lexical and Computational Semantics (*SEM). Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Annapolis (MD): Naval Academy; 2013.

26. Lee HJ, Xu H, Wang J, Zhang Y, Moon S, Xu J, Wu Y. UTHealth at SemEval-2016 Task 12: an end-to-end system for temporal information extraction from clinical notes. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). San Diego (CA): Association for Computational Linguistics; 2016. DOI

27. Strötgen J, Zell J, Gertz M. HeidelTime: tuning English and developing Spanish resources for TempEval-3. In: Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM). Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Atlanta (GA);2013.

28. Lin C, Miller T, Dligach D, Bethard S, Savova G. A BERT-based universal model for both within- and cross-sentence clinical temporal relation extraction. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. Minneapolis (MN);2019.

29. Niu J, Ng V, Penn G, Rees E. Temporal histories of epidemic events (THEE): a case study in temporal annotation for public health. In: Proceedings of the International Conference on Language Resources and Evaluation. Marseille (FR);2020.

30. Fundel K, Küffner R, Zimmer R. RelEx--relation extraction using dependency parse trees. Bioinformatics 2007 Feb;23(3):365–71. DOI PubMed

31. Nasheri N, Vester A, Petronella N. Foodborne viral outbreaks associated with frozen produce. Epidemiol Infect 2019 Oct;147:e291. DOI PubMed

32. Aries A, Eddine ZD, Hidouci WK. Automatic text summarization: what has been done and what has to be done. arXiv:1904.00688. https://arxiv.org/abs/1904.00688

33. Kryscinski W, Paulus R, Xiong C, Socher R. Improving abstraction in text summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels (BE): Association for Computational Linguistics; 2018. DOI

34. Petroni F, Raman N, Nugent T, Nourbakhsh A, Panic Z, Shah S, Leidner J. An extensive event extraction system with cross-media event resolution. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser L, Polosukhim I. Attention is all you need. In: Proceedings from the 31st Conference on Neural Information Processing Systems (NIPS 2017). Long Beach (CA);2017.

36. Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling. Proc Interspeech 2016;685–9. DOI

37. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. Deep contextualized word representations. In: Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) 2018. New Orleans (LA). DOI

38. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR 2018;abs/1810.04805. https://arxiv.org/abs/1810.04805

39. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le Q. XLNet: generalized autoregressive pretraining for language understanding. arXiv 2019;1906.08237. https://arxiv.org/abs/1906.08237

# Good times bad times: Automated forecasting of seasonal cryptosporidiosis in Ontario using machine learning

Olaf Berke[1]*, Lise Trotz-Williams[1,2], Simon de Montigny[3,4]

## Abstract

**Background:** The rise of big data and related predictive modelling based on machine learning algorithms over the last two decades have provided new opportunities for disease surveillance and public health preparedness. Big data come with the promise of faster generation of and access to more precise information, potentially facilitating predictive precision in public health ("precision public health"). As an example, we considered forecasting of the future course of the monthly cryptosporidiosis incidence in Ontario.

**Methods:** The traditional statistical approach to forecasting is the seasonal autoregressive integrated moving-average (SARIMA) model. We applied SARIMA and an artificial neural network (ANN) approach, specifically a feed-forward neural network, to predict monthly cryptosporidiosis incidence in Ontario in 2017 using 2005–2016 data as a training set. Both forecasting approaches are automated to make them relevant in a disease surveillance context. We compared the resulting forecasts using the root mean squared error (RMSE) and mean absolute error (MAE) as measures of predictive accuracy.

**Results:** Cryptosporidiosis is a seasonal disease, which peaks in Ontario in late summer. In this study, the SARIMA model and ANN forecasting approaches captured the seasonal pattern of cryptosporidiosis well. Contrary to similar studies reported in the literature, the ANN forecasts of cryptosporidiosis were slightly less accurate than the SARIMA model forecasts.

**Conclusion:** The ANN and SARIMA approaches are suitable for automated forecasting of public health time series data from surveillance systems. Future studies should employ additional algorithms (e.g. random forests) and assess accuracy by using alternative diseases for case studies and conducting rigorous simulation studies. Difference between the forecasts from the machine learning algorithm, that is, the ANN, and the statistical learning model, that is, the SARIMA, should be considered with respect to philosophical differences between the two approaches.



**Affiliations**

[1] Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, ON

[2] Wellington-Dufferin Guelph Public Health, Guelph, ON

[3] École de santé publique - Département de médecine sociale et préventive, Université de Montréal, Montréal, QC

[4] Centre de recherche du CHU Sainte-Justine, Montréal, QC

***Correspondence:**
oberke@uoguelph.ca

***Suggested citation:*** Berke O, Trotz-Williams L, de Montigny S. Good times bad times: Automated forecasting of seasonal cryptosporidiosis in Ontario using machine learning. Can Commun Dis Rep 2020;46(6):192–7. https://doi.org/10.14745/ccdr.v46i06a07

***Keywords:*** disease surveillance, machine learning, statistical learning, cryptosporidiosis, artificial neural network, SARIMA, forecasting, seasonal time series

## Introduction

Cryptosporidiosis is a potentially lethal diarrheal disease that affects humans and animals. It is caused by the protozoan parasite *Cryptosporidium* spp. (1). Some 20 of the known 26 species have been associated with human infections (2). The majority of human infections are caused by *C. hominis* and *C. parvum,* which are mostly related to anthropogenic and zoonotic transmissions, respectively (3). The main infection route for humans is through consumption (including while swimming) of water contaminated with the parasites' oocysts.

Cryptosporidiosis is often asymptomatic but can result in mild-to-severe gastrointestinal disease and even mortality. Human infection prevalence in North America ranges between 1% and 4% annually, but can be up to 20% elsewhere (4). While

cryptosporidiosis is likely underreported, it is known to occur more frequently in children and immunocompromised people. No prophylactic treatment is available, making public health preparedness based on surveillance an important preventive option.

New opportunities for statistics, epidemiology and disease surveillance in public health have emerged over the last two decades since the advent of big data (5,6). Eysenbach introduced the term "infodemiology" for the use of big data (and specifically social media use and behaviour data) in health surveillance (7). A prominent example of infodemiology is the Google Flu Trends project, which predicted regional outbreaks of influenza 7 to 10 days ahead of conventional surveillance methods by the Center for Disease Control and Prevention (CDC) but was grossly overestimating influenza prevalence (8). That project is a valuable example of the opportunities as well as the risks of big data, termed "big data hubris" (8,9).

Big data are often characterized by the five V's: volume, variety, velocity, veracity and value (9). Big data hubris refers to the veracity or truthfulness of the data. The promise of big data is that vast amounts of data (volume) of different types and from different sources (variety) provide a more complete and precise representation of reality, hence leading to "precision public health" (10). However, when the data are not representative of the population of interest, predictive inferences are biased.

Disease surveillance results in a big data situation due to data velocity and volume: data are constantly updated and growing in size. The dynamic nature of disease surveillance data requires an automated approach to analysis and forecasting. The traditional statistical time series modelling approach is the seasonal autoregressive integrated moving-average model (SARIMA) proposed by Box and Jenkins (11). A widely used machine learning algorithm for time series forecasting is the (feed-forward) artificial neural network (ANN) (12). We applied both forecasting approaches to predict monthly cryptosporidiosis incidence in Ontario in 2017 using 2005–2016 data as a training set. We compared these forecasting approaches using the 2017 incidence as test data, with the root mean squared prediction error (RMSE) and the mean absolute prediction error (MAE) as measures of accuracy.

Similar comparisons have been reported in the literature. Zhang and Qi (13) compared SARIMA and ANN using simulations and showed that the ANN is consistently better at forecasting than the SARIMA model, when data are appropriately preprocessed. Kane *et al*. (14) compared forecasts of avian influenza H5N1 outbreaks by the SARIMA model to those from the random forest algorithm and concluded that machine learning provides enhanced predictive ability over the time series modelling. Similarly, in a study of typhoid fever incidence in China, Zhang *et al.* compared SARIMA modelling to three different ANN architectures; the researchers concluded that all three neural network algorithms outperform the statistical model (15).

The goal of this study is to compare the two approaches to automating forecasting of monthly incidence rates of cryptosporidiosis in Ontario for the year 2017. The specific objectives were (1) to compare the accuracy of forecasts using the RMSE; (2) to compare forecasts using the MAE; and (3) to visually compare the forecasted incidence rates to the observed time series.

## Methods

The data we used were monthly incidence counts of cryptosporidiosis in Ontario for the years 2005 to 2017 as reported to Public Health Ontario and available from the respective homepages (16). For analysis, we split the dataset into training data (monthly incidences in 2005 to 2016) and test data (monthly incidences in 2017).

For exploration purposes, we reported ranges of annual and monthly mean incidence in the training data and inspected the data with the seasonal and trend decomposition using Loess (STL) method (17). The seasonal component was assumed to be time invariant or periodic, while the trend component was found using a moving window of length 73 months, or six years plus one month.

A SARIMA model (11) is a data-generating model that includes seasonal and trend components. It is used to describe autocorrelations within a time series and to predict future values. It is described by the order of filters applied to remove seasonal and trend components as well as by the order of lagged correlations in the filtered series. The filtered series is assumed to be stationary and Gaussian. A brief description of the SARIMA model is: SARIMA$(p,d,q)(P,D,Q)_S$, where $S$ denotes the length of the season (here 12 months), $d$ and $D$ denote nonseasonal and seasonal difference filters to remove trend and seasonal components, respectively. Furthermore, $p$ and $P$ are orders of the nonseasonal and seasonal autocorrelation parameters, respectively. Finally, $q$ and $Q$ denote the nonseasonal and seasonal order of moving-average parameters. The SARIMA modelling approach was automated by using maximum likelihood estimation and stepwise backward model selection with the Bayesian information criterion (BIC). The SARIMA model as fit to the 2005–2016 training data was then used to forecast monthly incidences for 2017 test data.

The ANN is a data-driven and automated algorithm to forecasting time series data. While a variety of ANN architectures exist (18,19), we applied the staple feed-forward multilayer neural network with a single hidden layer in this study (12). More specifically, the ANN is described as ANN$(p,P,k)_S$, where $p$, $P$ and $S$ have the same meanings as in the SARIMA model, and $k$ denotes the number of nodes in the hidden layer. Automatic selection of the ANN's order values was as follows: $S$=12 is known; $k$ was the rounded value of $(p+P+1)/2$, where $P$ was set to $P$=1 to accommodate linear seasonality; and $p$ was selected as

the optimal order of an autoregressive model fit to the remainder of term of the STL decomposed series.

We applied the ANN algorithm as follows: linear combinations of input data were subjected to the nonlinear sigmoid activation function $1/(1+exp(-z))$ as output from a hidden layer, and the output from the hidden layer was then aggregated in the form of linear combinations, which resulted in the final output. The ANN was trained using 100 repetitions, that is, 100 different random starting values for the weight parameters of the linear combinations between input and hidden layer as well as the hidden and output layers. Furthermore, the input series (i.e. the 2005–2016 data) was preprocessed using an automatic selection of the Box–Cox transformation parameter (by the Guerrero method (12)) followed by studentizing (i.e. centring and scaling). For each repetition, the algorithm was trained by an iterative experimental process of optimizing a loss function. The resulting set of forecasts, or ensemble, was averaged over all iterations.

Both forecasting approaches provide prediction intervals. The SARIMA prediction interval was based on estimated model parameter. The ANN prediction interval was based on 1,000 bootstrapped sample paths (12), that is, using resampled past residuals. In addition, both forecasting approaches were compared by their accuracy measures (RMSE and MAE) for the monthly forecasts and the observed test data of the year 2017.

All data analysis was performed in R (20) and RStudio (21) using the "forecast" package (12).

## Results

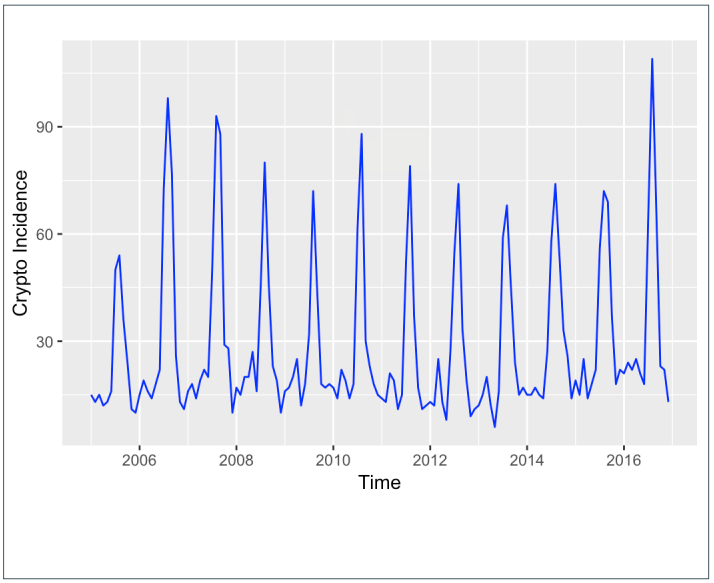
The time series of monthly reported cryptosporidiosis incidences in Ontario for the years 2005 to 2016 is dominated by a seasonal component, with summer peaks and only a weak – if any – upward trend (**Figure 1**). The STL decomposition in **Figure 2** confirms this impression. The training data time series is relatively short with 12 years comprising 4,152 cases, or an annual average of 346 cases, which is equivalent to about 2.57 annual cases per 100,000 population at risk. The average number of monthly cases was 29, ranging from 6 to 109 cases over 2005 to 2016.

The stepwise automated model selection resulted in a SARIMA$(1,0,0)(1,1,0)_{12}$ model with model parameter estimates being first order autoregressive parameter AR(1)=0.41 (standard error [SE]=0.08) and first order seasonal autoregressive parameter SAR(1)=−0.35 (SE=0.10) (**Figure 3**).

The automatically selected ANN is of order ANN$(11,1,6)_{12}$, that is, the last 11 observations plus the first seasonal observation are linearly combined into six nodes of a single hidden layer. The input series was Box–Cox transformed with an automatically chosen parameter $\lambda$=−0.21. The forecasts from the ANN are visualized together with 80% and 95% prediction intervals in **Figure 4.**

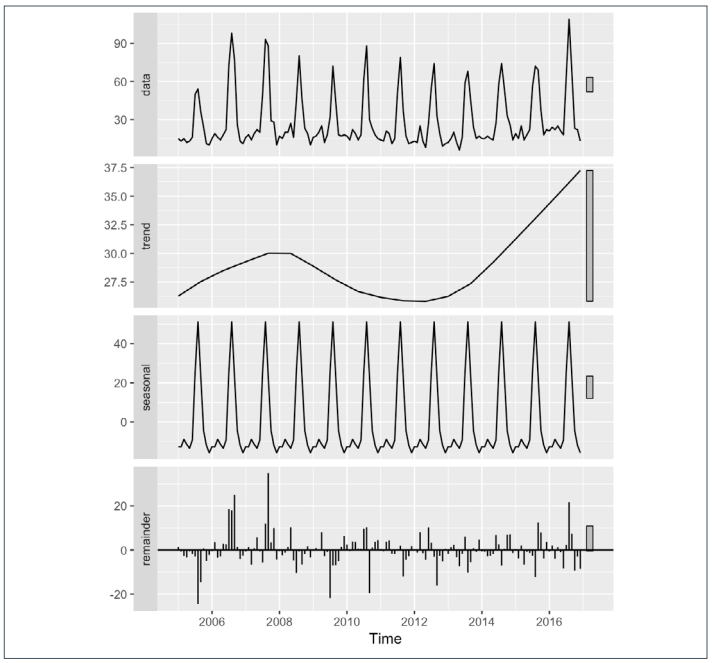
The observed monthly incidences and rounded forecasts are presented in **Table 1** and **Figure 5** for both models. **Table 2** shows the summaries of the RMSE and MAE from the 2017 forecasts for both approaches.

**Figure 1: Time series plot of the monthly incidence of cryptosporidiosis in Ontario during the years 2005 to 2016**



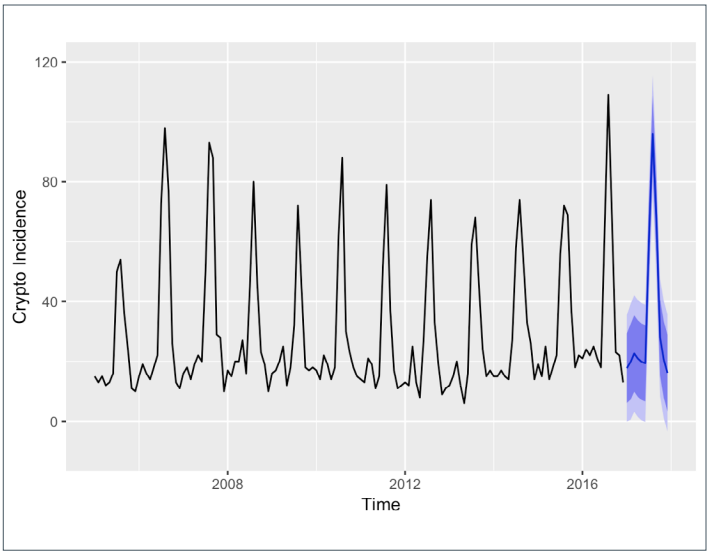

Abbreviation: Crypto, cryptosporidiosis

**Figure 2: The cryptosporidiosis time series of monthly incidences from 2005 to 2016**[a]



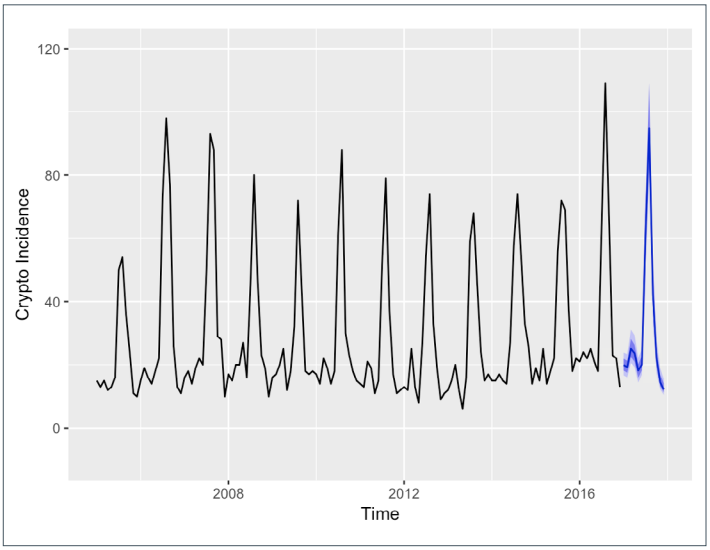

[a] Seasonal and trend decomposition based on Loess procedure (STL) plot of the training dataset (17)

**Figure 3: Forecasts for 2017 monthly cryptosporidiosis incidences with 80% and 95% prediction intervals from a SARIMA(1,0,0)(1,1,0)12 model**



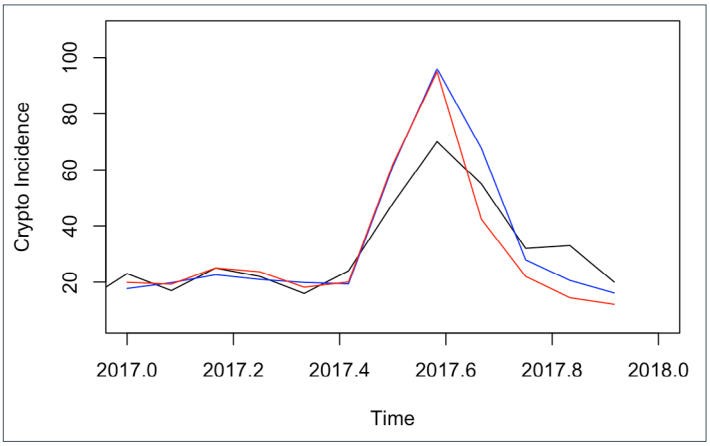

Abbreviations: Crypto, cryptosporidiosis; SARIMA, seasonal autoregressive integrated moving-average

**Figure 4: Forecasts for 2017 monthly cryptosporidiosis incidences with bootstrapped 80% and 95% prediction intervals from an ANN(11,1,6)$_{12}$ network**




Abbreviations: ANN, artificial neural network; Crypto, cryptosporidiosis

**Figure 5: Time series plot of the observed monthly cryptosporidiosis incidences for 2017 and the forecasts from SARIMA and ANN approaches**




Abbreviations: ANN, artificial neural network (red line); Crypto, cryptosporidiosis (black line); SARIMA, seasonal autoregressive integrated moving-average (blue line)

**Table 2: Predictive performance measures for the SARIMA and ANN approaches**

| Model | RMSE | MAE |
|---|---|---|
| SARIMA(1,0,0)(1,1,0)$_{12}$ | 10.3 | 7.7 |
| ANN(11,1,6)$_{12}$ | 11.2 | 8.4 |

Abbreviations: ANN, artificial neural network; MAE, mean absolute error; RMSE, root mean squared error; SARIMA, seasonal autoregressive integrated moving-average

**Table 1: Observed cryptosporidiosis incidence rates for 2017 and rounded forecasts from SARIMA and ANN approaches[a]**

| Incidence rate per month | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| Observed | 23 | 17 | 25 | 22 | 16 | 24 | 48 | 70 | 55 | 32 | 33 | 20 |
| SARIMA | 18 | 20 | 23 | **21** | 20 | 19 | **61** | 96 | **68** | **28** | **21** | **16** |
| ANN | **20** | **19** | **25** | 24 | **18** | **20** | 62 | **95** | **42** | 22 | 15 | 12 |

Abbreviations: ANN, artificial neural network; SARIMA, seasonal autoregressive integrated moving-average
[a] For each month, the forecast closer to the observed incidence are in bold

## Discussion

The monthly cryptosporidiosis incidence in Ontario is characterized by a dominant seasonal pattern that generally peaks in August. The short peak in incidence may support the concept of human behaviour as a main driver for infection since environmental conditions (e.g. ambient temperature) do not vary in a pattern similar to the incidence. No increasing trend was identified, meaning that the incidence is not emerging.

Neither the machine learning algorithm (i.e. the ANN) nor the statistical learning method (i.e. SARIMA) were found to have a superior performance in predicting monthly cryptosporidiosis incidence. While the ANN forecasts were closer to the observations for six months, the SARIMA performed better for a different group of five months; both methods were tied for the month of September of 2017 (see Table 1). However, the accuracy measures RMSE and MAE indicate a slight advantage for the SARIMA forecasts: the ANN's RMSE and MAE were higher by 0.9 and 0.7 units, respectively (see Table 2).

This slight advantage for the SARIMA is interpreted as follows: the SARIMA forecasts are, on average, almost one case per month more accurate than ANN forecasts. Although this result is unexpected with respect to the cited reports (13–15), it is in line with a systematic review (22) that found no evidence for more accurate predictions from machine learning alternatives to statistical logistic regression modelling. However, it should be noted that this is a case study and results are specific to this example. While the SARIMA model assumes white noise residuals and an additive seasonal component, this was not checked here using the automated modelling approach. Similarly, the ANN is optimized using backpropagation, which is known to have difficulties finding the optimal parameter estimates (19). Therefore, the ANN employs ensemble forecasting to guard against individual erroneous forecasts.

Proper data preprocessing is important for machine learning algorithms (23). This means a time series needs to be scaled and centred (i.e. studentized or normalized) prior to analysis. Data preprocessing is a natural part of the autoregressive integrated moving-average modelling approach, as trend and seasonal effects are filtered out before the model is fitted to the time series. In our study, stepwise model selection led to filtering out a seasonal effect, but a trend effect was neither identified nor removed. The ANN was preprocessed by a Box–Cox transformation, followed by centring and scaling.

Big data analysis is often presented together with machine learning algorithms for inference, that is, predictive modelling. The reason for doing so might originate from the impression that traditional statistical methods are inappropriate for the challenges of big data. For example, the variety of data expressed by the number of covariates could render traditional statistical inference less attractive and impractical. On the other hand, machine learning algorithms are designed around modern statistical methods for dimension reduction and regularization (e.g. Lasso regression). The training of machine learning algorithms is what is otherwise known as parameter estimation in statistical modelling and is no different from statistical learning methods, being based on cross-validation and bootstrapping.

In summary, to a certain degree statistical learning and machine learning do not differ. However, in public health, applications of big data analysis, namely predictive modelling including time series forecasting, differ from traditional biostatistical data analysis in terms of risk factor identification and assessment. Breiman distinguished this as "the two cultures" of statistical modelling: the data modelling culture and the algorithmic modelling culture (24). He argued that statistical theory is irrelevant if modelling assumptions are not met in real-data situations. However, he also admitted that machine learning algorithms are often based on little theory, and modelling assumptions are replaced by properties of the algorithms, that is, whether these converge and deliver good predictions.

From a philosophical point of view, machine learning is based on a "black box" that is not open to interpretations or explanations. In the current example, the ANN$(11,1,6)_{12}$ algorithm included a nonlinear combination of the time series data and 85 parameters (23). On the other hand, the SARIMA model describes how past observations affect the future course of a process; this characteristic might propose causal hypotheses (25). Therefore, it is not entirely correct to simply compare the forecasting methods by their predicted values or accuracy measures as the approaches are philosophically different and not entirely comparable: the ANN is a predictive algorithm, while the SARIMA is a descriptive and predictive model.

### Limitations

A limitation of this study is the lack of adjustment for the population at risk. Indeed the Ontario population is steadily increasing, but at an annual rate below 0.5%, which is negligible in this context, where underreporting is of greater concern. No trend in the monthly cryptosporidiosis incidence rates was indicated by either the SARIMA or ANN approaches.

### Conclusion

Cryptosporidiosis is a strongly seasonal disease, leading to good times and bad times of varying caseloads for public health. Machine learning methods suitable for forecasting of public health time series data from surveillance systems are becoming more popular; they have been demonstrated to be more accurate than traditional statistical methods. However, in this particular case study, the SARIMA model resulted in slightly lower RMSE and MAE and thus greater accuracy than the ANN. Both forecasting approaches captured the seasonal pattern of cryptosporidiosis well.

Future studies should employ additional algorithms (e.g. random forests) and assess accuracy in different setting, either by using alternative diseases for case studies or employing a more systematic approach and conducting simulation studies.

## Authors' statement

OB conceived the study, collected the data and performed the data analysis
OB, LTW and SDM all wrote and approved the manuscript

### Conflict of interest

The authors declare having no conflict of interest.

## References


1. Abeywardena H, Jex AR, Gasser RB. A perspective on Cryptosporidium and Giardia, with an emphasis on bovines and recent epidemiological findings. Adv Parasitol 2015:88:243–301. DOI PubMed

2. Ryan U, Fayer R, Xiao L. Cryptosporidium species in humans and animals: current understanding and research needs. Parasitology 2014:141(13):1667–85. DOI PubMed

3. Nwosu A, Berke O, Pearl DL, Trotz-Williams LA . Exploring the geographical distribution of cryptosporidiosis in the cattle population of Southern Ontario, Canada, 2011-2014. Geospat Health 2019;14(2);236–46. DOI

4. Medema G, Teunis P, Blokker M, Deere D, Davison A, Charles P, Loret JF. Risk assessment of Cryptosporidium in drinking water. Geneva (CH): World Health Organization; 2009. pp. 143.

5. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning and privacy. Annu Rev Public Health 2018;39:95–112. DOI PubMed

6. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epi-demiologists. Am J Epidemiol 2019;188(12):2222–39. DOI PubMed

7. Eysenbach G. Infodemiology: the epidemiology of (mis) information. Am J Med 2002;113(9):763–5. DOI PubMed

8. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar;343(6176):1203–5. DOI PubMed

9. Fuller D, Buote R, Stanley K. A glossary for big data in population and public health: dis-cussion and commentary on terminology and research methods. J Epidemiol Community Health 2017 Nov;71:1113–7. DOI PubMed

10. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nature 2016;540:189–91. DOI

11. Box G, Jenkins G. Time series analysis: forecasting and control. San Francisco: Holden-Day; 1970.

12. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice (2nd ed). Melbourne (AU): OTexts; 2018.

13. Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. Eur J Oper Res 2005;160(2):501–14. DOI

14. Kane MJ, Price N, Scotch M, Rabinowitz P. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics 2014;15:276. DOI PubMed

15. Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X. Comparative study of four time series methods in forecasting typhoid fever incidence in China. PLoS One 2013 May;8(5):e63116. DOI PubMed

16. Public Health Ontario. Infectious Disease Trends in Ontario. Toronto (ON): Public Health Ontario; 2020 (accessed 2020-01-20). https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/reportable-disease-trends-annually#/14

17. Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: a seasonal-trend decomposi-tion procedure based on Loess. J Off Stat 1990;6(1):3–73.

18. Jain AK, Mao J, Mohluddin KM. Artificial neural networks: a tutorial. Computer 1996;29(3):31–4. DOI

19. Warner B, Misra M. Understanding neural networks as statistical tools. Am Stat 1996;50(4):284–92. DOI

20. R Core Team. R: A language and environment for statistical computing. Vienna (AT): R Foundation for Statistical Computing; 2019.

21. RStudio Team. RStudio: integrated development for R. Boston (MA): RStudio, Inc.; 2018.

22. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019 Jun;110:12–22. DOI PubMed

23. Kuhn M, Johnson K. Applied predictive modeling. New York (NY): Springer; 2013. pp. 1–600. DOI

24. Breiman L. Statistical modeling: the two cultures. Stat Sci 2001;16(3):199–231. DOI

25. Shmueli G. To explain or to predict? Stat Sci 2010;5(3):299–310. DOI

# Modelling scenarios of the epidemic of COVID-19 in Canada

Nick H Ogden[1]*, Aamir Fazil[1], Julien Arino[2], Philippe Berthiaume[1], David N Fisman[3], Amy L Greer[4], Antoinette Ludwig[1], Victoria Ng[1], Ashleigh R Tuite[3], Patricia Turgeon[1], Lisa A Waddell[1], Jianhong Wu[5,6]

## Abstract

**Background:** Severe acute respiratory syndrome virus 2 (SARS-CoV-2), likely a bat-origin coronavirus, spilled over from wildlife to humans in China in late 2019, manifesting as a respiratory disease. Coronavirus disease 2019 (COVID-19) spread initially within China and then globally, resulting in a pandemic.

**Objective:** This article describes predictive modelling of COVID-19 in general, and efforts within the Public Health Agency of Canada to model the effects of non-pharmaceutical interventions (NPIs) on transmission of SARS-CoV-2 in the Canadian population to support public health decisions.

**Methods:** The broad objectives of two modelling approaches, 1) an agent-based model and 2) a deterministic compartmental model, are described and a synopsis of studies is illustrated using a model developed in Analytica 5.3 software.

**Results:** Without intervention, more than 70% of the Canadian population may become infected. Non-pharmaceutical interventions, applied with an intensity insufficient to cause the epidemic to die out, reduce the attack rate to 50% or less, and the epidemic is longer with a lower peak. If NPIs are lifted early, the epidemic may rebound, resulting in high percentages (more than 70%) of the population affected. If NPIs are applied with intensity high enough to cause the epidemic to die out, the attack rate can be reduced to between 1% and 25% of the population.

**Conclusion:** Applying NPIs with intensity high enough to cause the epidemic to die out would seem to be the preferred choice. Lifting disruptive NPIs such as shut-downs must be accompanied by enhancements to other NPIs to prevent new introductions and to identify and control any new transmission chains.



**Affiliations**

[1] Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, St. Hyacinthe, QC and Guelph, ON

[2] Department of Mathematics & Data Science NEXUS, University of Manitoba, Winnipeg, MB

[3] Dalla Lana School of Public Health, University of Toronto, Toronto, ON

[4] Department of Population Medicine, University of Guelph, Guelph, ON

[5] Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON

[6] Fields-CQAM Laboratory of Mathematics for Public Health, York University, Toronto, ON

*Correspondence:
nicholas.ogden@canada.ca

***Suggested citation:*** Ogden NH, Fazil A, Arino J, Berthiaume P, Fisman DN, Greer AL, Ludwig A, Ng V, Tuite AR, Turgeon P, Waddell LA, Wu J. Modelling scenarios of the epidemic of COVID-19 in Canada. Can Commun Dis Rep 2020;46(6):198–204. https://doi.org/10.14745/ccdr.v46i06a08

***Keywords:*** modelling, COVID-19, Canada, non-pharmaceutical interventions

## Introduction

In this article, we review efforts within the Public Health Agency of Canada (PHAC) to model transmission of severe acute respiratory syndrome virus 2 (SARS-CoV-2), the agent of coronavirus disease 2019 (COVID-19), to support public health decisions. The COVID-19 pandemic has emerged at remarkable speed. The SARS-CoV-2 is likely a bat-origin coronavirus (1) that may have "spilled over" into humans from an intermediary animal reservoir not yet identified. The first detected human exposure event was linked to a "wet market" in the city of Wuhan, the capital city of the province of Hubei, China some time during late 2019 (2). The virus was likely already capable of human-to-human transmission, but evolved more efficient transmissibility during late 2019 (3). Human-to-human transmission was officially recognised by the global public health community in mid-January 2020 (4). Shortly after this, spatial spread in China was reported. Intense public health measures (or non-pharmaceutical interventions; NPIs) of case detection, contact tracing and quarantine, and

social distancing were implemented within the province of Hubei, and the region was isolated from the rest of China by travel restrictions (5). International travel restrictions to and from China were introduced but cases had already escaped outside of China (6), and the consequent global spread resulted in declaration of a pandemic (4). The first travel-related cases were identified in Canada in January 2020 and by April 2020, community transmission was occurring in all provinces with the possible exception of Prince Edward Island. Community transmission has yet to be reported from the territories. The majority of cases and deaths have been reported from the four largest provinces (British Columbia, Alberta, Ontario and Quebec). Physical distancing (including school, college, university and "non-essential" business closures) was implemented from mid-March 2020 in Canada and subsequent reductions in disease transmission suggest that these and other NPIs (detailed below) are slowing the epidemic (7).

## Evolution of COVID-19 modelling

At first, modelling studies focused on the epidemic in China, and particularly on the dynamics of the epidemic in the city of Wuhan and throughout the province of Hubei. At this early stage, there was much effort to analyse surveillance data from China to obtain parameter estimates such as the basic reproduction number ($R_0$), case fatality rate and incubation period (8). For the first attempts at Susceptible-Exposed-Infectious-Recovered (SEIR) type dynamic models, parameter estimates were "borrowed" from what was known about other coronaviruses (SARS-CoV and MERS-CoV) (6) and/or obtained by fitting the models to surveillance data (9,10). As more data on SARS-CoV-2 virus transmission and the course of infection in humans have become available, models have become increasingly parameterised using SARS-CoV-2-specific data (11). With global spread of the disease, and with a vaccine likely more than a year away, modelling efforts turned to assessing the possible extent of the epidemic in countries outside China, and the impact of different NPIs (11,12). Emerging science has revealed that SARS-CoV-2 virus is highly transmissible by respiratory and possibly fecal-oral routes, is transmitted before symptoms appear and some cases may be entirely asymptomatic (13,14). The virus can be highly pathogenic for older people and some younger people, particularly those with co-morbidities (14). Presymptomatic transmission, mild symptoms (particularly in younger age groups) and asymptomatic cases all hinder detection of infective cases (in contrast to SARS) (15), making control difficult. Modelling efforts to date have illuminated the magnitude of the challenge we face: 1) the global population is entirely immunologically naive, 2) the virus is very highly transmissible ($R_0$ values may be greater than five in some settings) (16) and 3) the level of pathogenicity of SARS-CoV-2 means even the most advanced healthcare systems in the world may be completely overwhelmed if the virus is allowed to spread without introducing NPIs. At the same time, and in contrast to pandemic influenza, there are no known effective antivirals.

## COVID-19 modelling in Canada

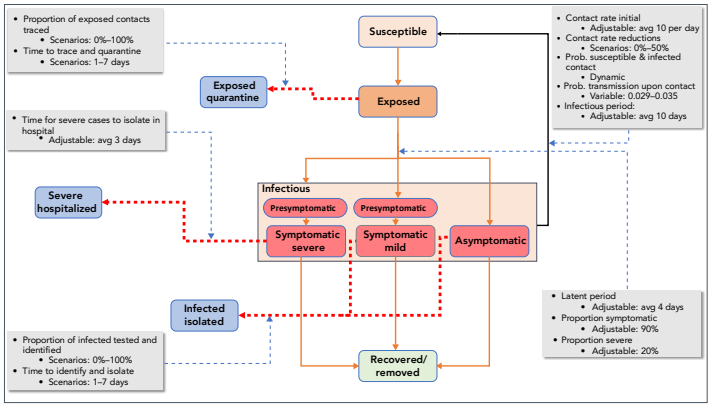
Predictive modelling of COVID-19 by Canadian scientists is a field of study that was approximately three months old at the time of writing (early May 2020), but there is extensive skill in Canada in modelling the transmission of infectious diseases. Some previously-developed models that investigated interventions to control H1N1 and other influenzas (17,18) have been adapted to assess the transmission of SARS-CoV-2 and the impacts of different NPIs (9,19,20). Tang *et al*. (9) and Li *et al*. (21) have developed SEIR-type models to explore transmission and learn from NPIs implemented in China and South Korea. An Expert Modelling Group, comprising more than 50 federal, provincial, territorial and university-based modellers and epidemiologists, has been assembled by PHAC to develop a Canadian COVID-19 modelling network that supports decision-making. Similar groups have been developed in other countries, with liaison between them facilitated by a World Health Organization modelling group.

## COVID-19 modelling at PHAC

In January 2020, a modelling group was convened, and development of two complementary modelling approaches was initiated. The prime objective of this modelling was to assess the impact of different NPIs and levels of efficacy needed to control the epidemic in Canada. In the absence of a vaccine or treatment, the NPIs available to control the epidemic, which were explored in modelling, are 1) physical distancing (closure of schools, colleges and universities, meeting places, gatherings and personal distancing) that reduces the rates of contact between members of society (including those who may be infected), 2) detection of cases by surveillance, and their isolation to prevent them from transmitting infection and 3) tracing and quarantine of people who have had contacts with cases.

The two modelling approaches used will be published as separate, more detailed articles, and are termed "approaches" as the models themselves have evolved with evolving knowledge. What follows is a broad description of these approaches, which are based on 1) an agent-based model (22) and 2) a deterministic compartment model (23). The agent-based approach was developed *de novo* using AnyLogic© software, while the deterministic model runs in R (24). Initial versions of the deterministic model were adapted from Tang *et al*. (9). Both are SEIR-type models with elements to model SARS-CoV-2 and impacts of NPIs, with more realism (**Figure 1**). These elements include compartments for isolated cases and quarantined "exposed" contacts from which onward transmission to susceptible people is limited or absent, compartments for

**Figure 1: A schematic of the Analytica 5.3 compartment model showing the flow of populations between compartments**




Abbreviations: avg, average; prob, probable
The black arrow reflects the transmission between infected and susceptible populations, the red dotted lines reflect the removal of infectious or exposed individuals as a result of non-pharmaceutical interventions

asymptomatic cases that may or may not be detected by surveillance, as well as flows to "isolation" and "quarantine" compartments that allow variation according to different levels of public health effort. Parameters in the models are calibrated according to values obtained by literature searches, which are conducted each day to ensure evolving knowledge is captured by the models.

In the deterministic modelling approach, effects of physical distancing are modelled in a simplistic way by reducing daily per capita contact rates informed by a number of metrics including cellphone data (25). The agent-based model approach, which contains stochastic elements, uses a more detailed representation of communities, including homes and communal meeting spaces that may represent workplaces, schools, malls and restaurants etc., so this model is capable of modelling effects of closures in more detail and predicting different realizations of disease spread in the community. The agent-based approach includes contact rates within and between age groups that are (20) based on the POLYMOD study (26). Contact data specific for the United Kingdom (UK) and European countries were used as a surrogate for contact rates in Canada given that similar studies have not been conducted in Canada. The deterministic approach has used homogenous mixing but is also being modified to include contact rates that vary within and between age groups. For illustrative purposes, a deterministic compartment model has been developed in Analytica 5.3 (Lumina Inc.) using the knowledge of COVID-19 transmission elucidated by the two modelling approaches. Code for this model is available upon request with instructions on how to explore the model and generate results.

Key outputs from the models are 1) under what circumstances the NPIs cause the epidemic to die out by reducing the effective reproduction rate ($R_e$) below unity (i.e. one infected person infects fewer than one other person, on average) and 2) the final attack rate (i.e. the total percentage of the population infected) and 3) the approximate duration of the epidemic.

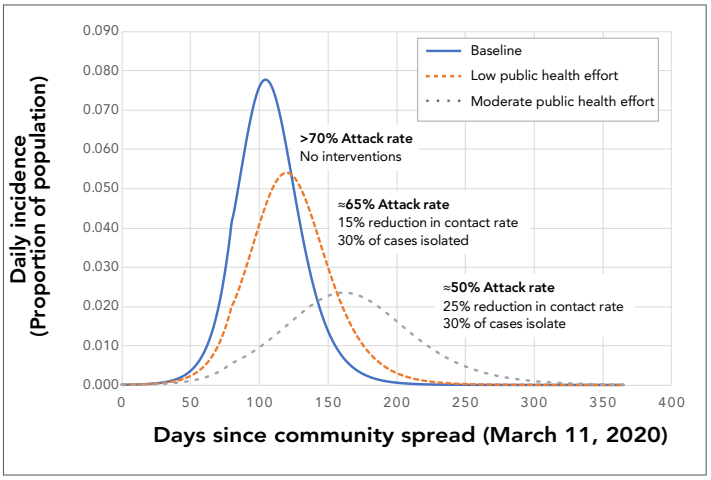
## Synopsis of findings of modelling conducted within and outside Public Health Agency of Canada

The following synopsis of results of modelling studies includes those conducted by PHAC with reference to studies conducted by modelling groups outside the Government of Canada. The outputs are illustrated here using graphs generated by the model developed in Analytica 5.3.

1. What happens in the absence of NPIs?

The baseline to compare impacts of NPIs is a scenario where there is no effort to control disease spread. In this case, the attack rate is predicted to be greater than 70%, and the epidemic lasts approximately one year (**Figure 2**). These findings are consistent with studies estimating impacts on UK and United States populations (11).

**Figure 2: Impacts of partially-effective NPIs on the epidemic compared to the baseline with no control efforts**



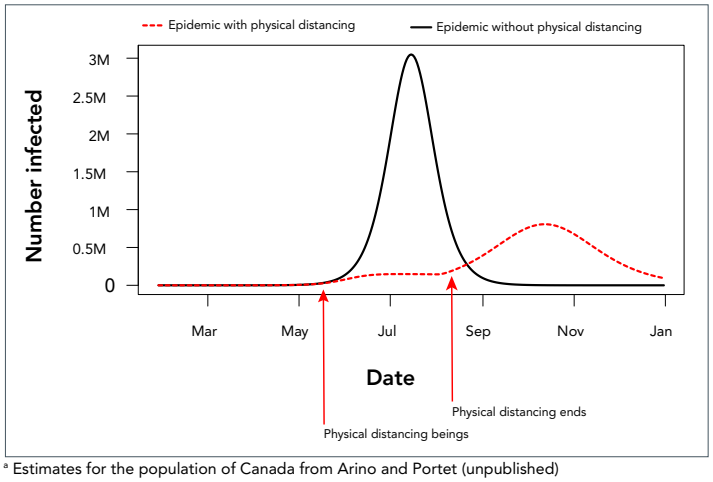

Values by each curve are the final estimated attack rate and the non-pharmaceutical interventions (NPIs) employed. It is assumed that NPIs are in effect for the duration of the epidemic

2. What happens when NPIs are partially effective?

If NPIs, maintained throughout the epidemic, are partially effective (i.e. they have impacts on the epidemic but do not cause it to die out), the main effects are as follows: the epidemic is prolonged, the peak is reduced, the epidemiological curve is flattened and the attack rate is reduced to approximately 50% (to 25% in some models) (20) (Figure 2). This finding is consistent with equivalent modelling studies (11,20). This scenario has been termed "delay and reduce."

If the NPIs do not cause the epidemic to die out and are lifted before the epidemic is over, the epidemic is predicted to rebound and the attack rate can be as high as without NPIs because the majority of the population remains naive (**Figure 3**) (11,20).

**Figure 3: The effect of initiating partially-effective non-pharmaceutical interventions (in this case physical distancing), and then removing these interventions before the epidemic ends**[a]




[a] Estimates for the population of Canada from Arino and Portet (unpublished)

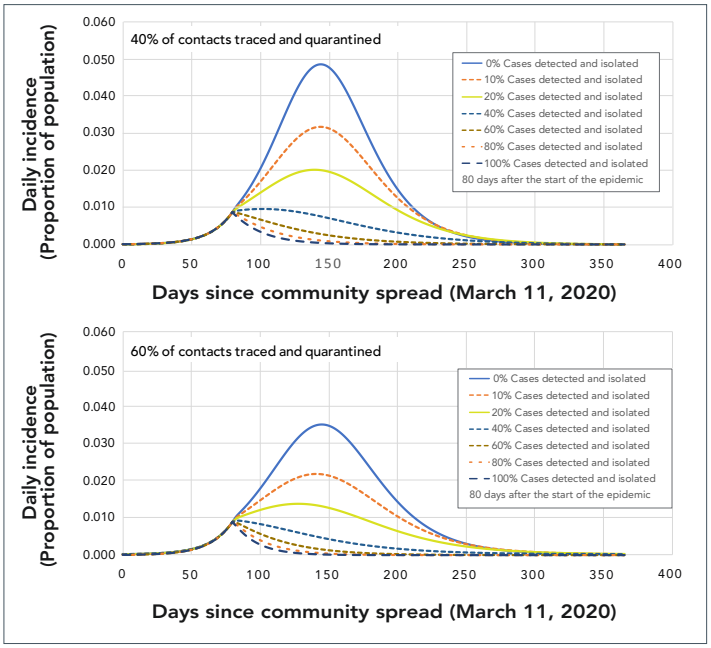
3. What happens when NPIs are highly effective?

When NPIs are highly effective, $R_e$ falls below unity, the epidemic dies out and does not rebound if NPIs are lifted (also referred to as epidemic control). How soon that happens, and thus the final attack rate (which may be anywhere between less than 1% and 25%), depends on a number of factors including at what point in the epidemic NPIs are implemented, the intensity with which NPIs are implemented, the duration of implementation and the level of compliance (**Figure 4**) (12).

## Assessing hospitalisation and mortality rates from attack rates

The main objective of the modelling approaches was to compare the impacts of different NPIs. There remains much uncertainty in some model parameters and their distributions, including the duration of the latent period, the proportion of cases that are asymptomatic and the duration of infectivity. The strength of these models is their ability to compare amongst different NPIs using current best estimates of parameter values. However, in order to ensure health care capacity is sufficient to respond to the pandemic, planners need to have a range of estimates for the expected numbers of cases, hospitalizations, cases needing care in intensive care units (ICU) and fatalities. Initial modelling approaches focused on estimating total attack rates, and were not designed to estimate hospitalizations, cases

**Figure 4: Effects of different levels and combinations of non-pharmaceutical interventions on whether or not, and how quickly, epidemic control is reached**




needing ICU care and fatalities. To obtain these estimates from total attack rates, age-specific severity estimates from analysis of international surveillance data (27) were used to assess the proportion of cases in Canada that would be mild or asymptomatic, require hospitalization or ICU treatment, and may die, according to the demography of the Canadian population as a whole (28). The estimates per one million population are shown in **Table 1**.

These estimates are crude and more precise health care needs estimates should be calculated at the community level (e.g. catchment area for a hospital or group of hospitals) so the specific age structure and co-morbidities of the community under evaluation can be accounted for in the model (19).

## Observations from Canada and elsewhere in the world

The outputs of modelling studies are theoretical, but their insights and policy implications have been bolstered by real-world evidence. Epidemic control has been realised in Singapore, China and South Korea, with $R_e$ falling below unity by application of a prompt and intense level of NPIs (5,21,29). In contrast, in Europe, NPIs to date do not seem to have brought $R_e$ below unity (30). At the time of writing, in Canada the epidemic is geographically heterogeneous, but unpublished estimates suggest that in some jurisdictions $R_e$ may be below unity, while in others this state has not been reached (*unpublished; Dr. Ashleigh Tuite and Dr. David Champredon*). At the time of writing, the observed case fatality rate is higher than that predicted using

**Table 1: Estimates per million population of hospitalisations, patients needing intensive care unit care and fatalities according to different scenarios for levels of control[a]**

| Level of epidemic control | "Delay and reduce" | | | Epidemic controlled | | |
|---|---|---|---|---|---|---|
| Attack rate | 50% | 25% | 10% | 5% | 2.5% | 1% |
| All cases | 500,000 | 250,000 | 100,000 | 50,000 | 25,000 | 10,000 |
| Mild (89.5%) | 450,000 | 225,000 | 90,000 | 45,000 | 22,000 | 9,000 |
| Hospitalised—not ICU (8%) | 39,000 | 19,000 | 7,800 | 3,900 | 2,000 | 800 |
| Hospitalised— ICU (2.5%) | 12,000 | 6,000 | 2,400 | 1,200 | 600 | 200 |
| Fatalities (1.2% all cases) | 6,000 | 3,000 | 1,200 | 600 | 300 | 100 |

Abbreviation: ICU, intensive care unit
[a] Figures greater than 1,000 were rounded to the nearest 1,000

methods described above because of extensive transmission in long-term care and seniors facilities. In these facilities, contact rates are likely very high (31) and the population very vulnerable to COVID-19.

## Conclusion

The modelling studies described here provide information for planning of public health policies to combat the unprecedented risk of COVID-19 to the health and well-being of Canadians and the rest of the world. These studies underline that without NPIs, the majority of Canadians would acquire infection in a relatively short period of time, and the health care system would most likely be overwhelmed, resulting in a higher case fatality rate, particularly in the most vulnerable age groups. The intensity of the NPIs, and the compliance of the public, will determine whether the epidemic is brought under control, or delayed and reduced. The former would seem the preferred objective as the numbers of Canadians affected would be minimized. However, this will require a very high degree of public health effort and public buy-in and, if successful, will require a high level of vigilance to identify imported cases and any transmission chains that may result, because the Canadian population remains largely infection naive. If transmission in Canada is not completely extinguished, strong NPIs will have to remain in place or the epidemic will rebound. Any lifting of physical distancing, which appears to be bearing fruit at present, will have to be matched by increased efforts to detect cases by surveillance and to trace and quarantine contacts.

Modelling studies are not predictions, they present plausible outcomes with different levels of NPIs, given our current knowledge of the virus and its transmission, and can be used to support planning, particularly in evidence-limited situations, such as emerging infectious disease epidemics. Our knowledge is constantly evolving, and the models and their outcomes will evolve accordingly. The models provide information that is useful for decision-making, but they do not make decisions. Decisions on public health programs to control COVID-19 in Canada will be made accounting for a range of additional factors that include (but are not limited to) economic impacts, ethical and legal concerns, and the negative health impacts of aggressive physical distancing.

## Funding

This work was funded by the Public Health Agency of Canada. JA and JW's work has also been funded by the Canadian Institute of Health Research (CIHR) 2019 Novel Coronavirus (COVID-19) rapid research program.

## References

1. Lau SK, Luk HK, Wong AC, Li KS, Zhu L, He Z, Fung J, Chan TT, Fung KS, Woo PC. Possible Bat Origin of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020;26(7). DOI [Epub ahead of print] PubMed

2. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr Biol 2020 Apr;30(7):1346–1351.e2. DOI PubMed

3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020;26(4):450–2. DOI PubMed

4. World Health Organization. WHO Timeline – COVID-19. Geneva (CH): WHO; April 27, 2020. https://www.who.int/news-room/detail/08-04-2020-who-timeline---covid-19

5. Wang C, Liu L, Hao X, Guo H, Wang Q, Huang J, He N, Yu H, Lin X, Pan A, Wei S, Wu T. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA 2020;323(19):1915-23. DOI

6. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395(10225):689–97. DOI PubMed

7. Government of Canada. Coronavirus disease (COVID-19). Ottawa (ON); Government of Canada; modified May 15, 2020. https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html

8. Li J, Wang Y, Gilmour S, Wang M, Yoneoka D, Wang Y, You X, Gu J, Hao C, Peng L, Du Z, Xu DR, Hao Y. Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study. MedRXiv: 2020a https://www.medrxiv.org/content/10.1101/2020.02.18.20024315v1.full.pdf

9. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, Wu J. Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions. J Clin Med 2020a;9(2):E462. DOI PubMed

10. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM; Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020;20(5)552-8. DOI [Epub ahead of print]

11. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunubá Z, Cuomo-Dannenburg G, Dighe A, Dorigatti I, Fu H, Gaythorpe K, Green W, Hamlet A, Hinsley W, Okell LC, van Elsland S, Thompson H, Verity R, Volz E, Wang H, Wang Y, Walker PG, Walters C, Winskill P, Whittaker C, Donnelly CA, Riley S, Ghani AC. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. London (UK): Imperial College; 2020. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf

12. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Munday JD, Kucharski AJ, Edmunds WJ, Funk S, Eggo RM; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020;8(4):e488–96. DOI PubMed

13. Wei WE, Li Z, Chiew CJ, Yong SE, Toh MP, Lee VJ. Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020;69(14):411–5. DOI PubMed

14. Vetter P, Vu DL, L'Huillier AG, Schibler M, Kaiser L, Jacquerioz F. Clinical features of covid-19. BMJ 2020;369:m1470. DOI PubMed

15. Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis 2020 May;20(5):e102–7. DOI PubMed

16. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020;26(7). DOI [Epub ahead of print] PubMed

17. Arino J, Brauer F, van den Driessche P, Watmough J, Wu J. A model for influenza with vaccination and antiviral treatment. J Theor Biol 2008;253(1):118–30. DOI PubMed

18. Tuite A, Fisman DN, Kwong JC, Greer A. Optimal pandemic influenza vaccine allocation strategies for the canadian population. PLoS Curr 2010;2:RRN1144. DOI PubMed

19. Arino J, Portet S. A simple model for COVID-19. Inf Dis Modelling. 2020:5: 309-15. PubMed

20. Tuite AR, Fisman DN, Greer AL. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. CMAJ 2020;192(19):E497–505. DOI PubMed

21. Li J, Yuan P, Heffernan J, Zheng T, Ogden N, Sander B, Li J, Li Q, Bélair J, Dzevela Kong J, Aruffo E, Tan Y, Jin Z, Yu Y, Fan M, Cui J, Teng Z, Zhu H. Observation wards and control of the transmission of COVID-19 in Wuhan. Bull World Health Organ. DOI

22. Ng V, Fazil A, Waddell S, Bancej C, Turgeon P, Ogden NH. An agent-based model of COVID-19 transmission in Canada: forecasting impacts and de-escalation of non-pharmaceutical public health interventions [submitted for publication].

23. Ludwig A, Berthiaume P, Orpana H, Swerdfeger H, Otten A, Statistics Canada team, Ogden NH. A dynamic compartmental model of national COVID-19 transmission and the impact of varying levels of case identification and contact tracing in Canada [submitted for publication].

24. R Core Team. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2020. https://www.r-project.org/

25. Google. https://www.google.com/covid19/mobility/

26. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008;5(3):e74. DOI PubMed

27. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PGT, Fu H, Dighe A, Griffin JT, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, FitzJohn R, Gaythorpe K, Green W, Hamlet A, Hinsley W, Laydon D, Nedjati-Gilani G, Riley S, van Elsland S, Volz E, Wang H, Wang Y, Xi X, Donnelly CA, Ghani AC, Ferguson NM. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020:pii:S1473-3099(20)30243-7. DOI [Epub ahead of print]

28. Statistics Canada. Population estimates on July 1st, by age and sex. Table 17-10-0005-01. Ottawa (ON): StatsCan; modified May 18, 2020. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501

29. Tang B, Xia F, Bragazzi NL, McCarthy Z, Wang X, He S, Sun X, Tang S, Xiao Y, Wu J. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. Bull World Health Organ 2020. DOI [Epub ahead of print]

30. Flaxman S, Mishra S, Gandy A, Unwin HJ, Coupland H, Mellan TA, Zhu H, Berah T, Eaton JW, Guzman PN, Schmit N, Callizo L, Ainslie KE, Baguelin M, Blake I, Boonyasiri A, Boyd O, Cattarino L, Ciavarella C, Cooper L, Cucunubá Z, Cuomo-Dannenburg G, Dighe A, Djaafara B, Dorigatti I, van Elsland S, FitzJohn R, Fu H, Gaythorpe K, Geidelberg L, Grassly N, Green W, Hallett T, Hamlet A, Hinsley W, Jeffrey B, Jorgensen D, Knock E, Laydon D, Nedjati-Gilani G, Nouvellet P, Parag K, Siveroni I, Thompson H, Verity R, Volz E, Walters C, Wang H, Wang Y, Watson O, Winskill P, Xi X, Whittaker C, Walker PG, Ghani A, Donnelly CA, Riley S, Okell LC, Vollmer MA, Ferguson NM, Bhatt S. Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. London (UK): Imperial College; 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/

31. Najafi M, Laskowski M, de Boer PT, Williams E, Chit A, Moghadas SM. The effect of individual movements and interventions on the spread of influenza in long-term care facilities. Med Decis Making 2017;37(8):871–81. DOI PubMed

# The Canadian Public Health Laboratory Network protocol for microbiological investigations of emerging respiratory pathogens, including severe acute respiratory infections

Respiratory Virus Infections Working Group[1]

***Suggested citation:*** Respiratory Virus Infections Working Group. The Canadian Public Health Laboratory Network protocol for microbiological investigations of emerging respiratory pathogens, including severe acute respiratory infections. Can Commun Dis Rep 2020;46(6):205–9. https://doi.org/10.14745/ccdr.v46i06a09
***Keywords:*** protocol, COVID-19, emerging respiratory pathogens, SARI, Canada, CPHLN



**Affiliation**

[1] Canadian Public Health Laboratory Network, Winnipeg, MB

***Correspondence:**
dionne.marcino@canada.ca

## Introduction

A protocol for severe acute respiratory infections (SARI) was initially developed as a response to the 2003 severe acute respiratory syndrome (SARS) outbreak (1). The protocol's intended use was to facilitate the diagnosis of novel and emerging respiratory infections, including SARI, due to both unknown and known respiratory pathogens that have the potential for large-scale epidemics. With both the Middle East respiratory syndrome coronavirus (MERS-CoV) and the influenza A(H7N9) virus, a key factor in diagnosis is the determination of risk based on epidemiologic factors, which, in turn, is related to exposure in an "area of concern". With the more highly transmissible severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of the coronavirus disease 2019 (COVID-19) pandemic, testing guidelines have changed as the pandemic has developed. Initially, testing focused on those with travel-associated risk factors; with a shift to broader testing once more cases were locally acquired (within Canada). This initial risk assessment must be done in concert with the local Ministry of Health. Signals of novel and emerging respiratory infections, including SARI alerts, should trigger clinicians to "think, tell and test":

- **Think** about the possibility of an emerging respiratory infection (e.g. novel influenza A virus)
- **Tell** the local medical officer of health or local public health official; notify your local laboratory and provincial public health laboratory (PPHL) that you suspect a novel pathogen
- **Test** for pathogen based on clinical symptoms and only after appropriate consultation

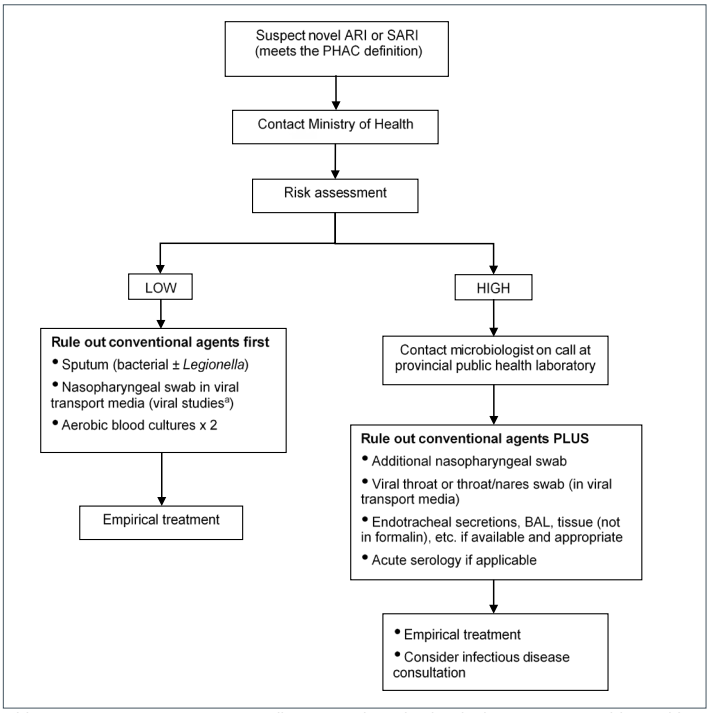
## Laboratory protocol

### When to test

Guidance on when to test for novel or emerging pathogens is influenced by many epidemiological factors. At the time of the emergence of a novel pathogen, before widespread human infection, the probability that a SARI is due to a novel pathogen is extremely low. Therefore, in patients with no epidemiological risk factors, the most common pathogens should be ruled out before considering a novel, unusual or more highly virulent pathogen. When appropriate risk factors exist, novel pathogens should still be ruled out regardless of whether another pathogen is detected, as viral co-infections have been well documented with novel viral respiratory pathogens (e.g. MERS-CoV, influenza A(H7N9) and SARS-CoV-2) (**Figure 1**).

Although testing is initially focused on individuals with epidemiologic links, such as those who have travelled to a region where the pathogen is circulating, once there is widespread activity, such as in the COVID-19 pandemic, healthcare providers should have a low threshold to consider testing for the novel pathogen when reviewing patients with acute respiratory illness (ARI). This threshold for testing will continue to be influenced by the epidemiology of the infection such as 1) the stage of the pandemic wave that the jurisdiction is in, 2) whether there is local or widespread activity and 3) whether the response is in a containment or mitigation phase. Specific testing guidelines are developed at the provincial level, and will vary across Canada. Other factors that may influence approaches to testing include the availability of testing supplies and reagents,

**Figure 1: Laboratory protocol**




Abbreviations: ARI, acute respiratory illness; BAL, bronchoalveolar lavage; PHAC, Public Health Agency of Canada; SARI, sudden acute respiratory infection
[a] Contact microbiologist on call for guidance regarding appropriate test

which may be in short supply at various times during a pandemic. This testing may be done at the local laboratory or the PPHL depending on local capacity and expertise.

## Specimens to collect

Until the ideal specimen to detect an emerging pathogen has been identified, a broad range of specimens should be collected including nasopharyngeal swab (NPS), throat or combined throat/nares swab, bronchoalveolar lavage (BAL), endotracheal secretions, and sputum. For pediatric patients, a nasopharyngeal aspirate is a suitable replacement to a NPS, although it is an aerosol generating medical procedure which requires airborne precautions whereas NPS does not. Although saliva has been suggested for detection of some emerging pathogens like SARS-CoV-2, it requires more validation prior to being recommended as the sole, specimen for collection.

For patients not admitted to hospital, including those in emergency room settings, a single upper respiratory tract specimen is usually sufficient for testing emerging respiratory virial pathogens (such as SARS-CoV-2 or H7N9). Upper respiratory tract specimens include a NPS, throat OR combined throat/nares swab collected in universal transport medium. An NPS is the preferred specimen due to possible increased sensitivity in comparison with a throat swab. A combined throat/nares specimen may also be collected, provided the testing laboratory has approved the submission of this combined specimen source type.

For hospitalized patients, in particular those with SARI, submission of both upper and lower respiratory tract specimens is recommended when possible. As above, NPS is the preferred upper respiratory tract specimen. A throat or combined throat/nares swab collected in universal transport medium may be submitted as an additional upper tract specimen. Lower respiratory tract specimens should also be submitted when possible. These lower tract specimens include sputum, endotracheal aspirates and BAL. For a number of emerging pathogens, including avian influenza and novel coronaviruses, there have been reports of patients who were found to be negative on upper respiratory tract testing but positive on lower respiratory tract testing.

## Recommended pathogens and specimens to test

At the time of the emergence of a novel pathogen, before widespread human infection, the probability that a severe acute respiratory illness is due to the novel pathogen is extremely low. Therefore, in patients with no epidemiological risk factors the most common pathogens should be ruled out before considering an unusual or more highly virulent pathogen. This includes the following:

**Conventional bacteria**

- Sputum for routine bacterial Gram stain and culture

**Atypical bacteria**

- *Legionella*—sputum, BAL, endotracheal aspirate, lung tissue for polymerase chain reaction (PCR) and/or culture
- Urine for *Legionella* urinary antigen testing
- *Mycoplasma/Chlamydia*—NPS, throat swab, and/or lower tract specimen for PCR and/or culture

**Conventional respiratory viruses**

- Including human influenza, parainfluenza, respiratory syncytial virus, adenovirus, human metapneumovirus, rhinovirus/enterovirus, coronavirus
- Specimens—NPS, endotracheal secretions, BAL, with or without throat swab (or combined throat/nares swab) and sputum
- NPS is the primary specimen type for respiratory viruses including seasonal influenza. However, based on our experience with pandemic H1N1, deeper specimens, such as endotracheal secretions or BAL, must be collected in cases of severe respiratory infection with negative NPS
- A number of avian influenza A viruses, including H7N9, has been detected in throat swabs. Influenza A (H7N9) was only detectable in sputum specimen in one of four patients. While sputum and throat swabs are not ideal for most influenza viruses, multiple specimens types should be considered in patients suspected of having avian influenza A viruses

## Testing methods

Testing should be conducted using assays validated for the specific pathogen:

- SARS-CoV-2 should be tested by real-time reverse transcription-PCR (rRT-PCR) (see above)
- The primary method for detection of influenza A and B (24 hour turnaround time) should be rRT-PCR with subtyping (H3N2 or H1N1). Preferred protocols for detection of novel influenza viruses are those developed by the United States Centers for Disease Control
- For parainfluenza, human metapneumovirus, coronavirus, rhinovirus/enterovirus and adenovirus, respiratory multiplex RT-PCR should be done on all specimens if possible; or on influenza-negative specimens when there is a clinical indication to detect non-influenza viruses
- Rapid influenza diagnostic tests should not be used to rule out influenza A. The sensitivity of currently available rapid influenza diagnostic tests for human influenza strains is suboptimal. The sensitivity of currently available commercial tests for detection of H7N9 is poor and should not be used for clinical testing
- SARS-CoV-2, novel influenza A viruses and MERS-CoV are classified as RG3 pathogens. Routine culturing of specimens from suspect patients should only be considered in PPHLs with containment level (CL) 3 facilities. Virus culture in a CL2 laboratory may be considered if the specimen has been tested for the relevant emerging pathogens and is negative by rRT-PCR

If more invasive samples are collected they should be processed for a wide range of pathogens:

- BAL for testing for a broad range of pathogens (bacteria, viruses, mycobacteria, fungi)
- Open lung biopsy—bacterial, mycobacterial and fungal, cultures, RT-PCR and histology (ensure specimen is NOT PUT IN FORMALIN

## When to suspect SARS-CoV-2 virus

During the early phases of the pandemic, which began in Wuhan, China in December 2019, only persons who returned from Wuhan, then the province of Hubei, China, with ARI were considered for testing. With the progression of the epidemic, testing of those with ARI after return from travel to countries with COVID-19 activity was indicated.

Following the evolution of COVID-19 to a pandemic and local transmission in most jurisdictions in Canada, testing approaches were broadened, with an initial focus on case identification for contact tracing and testing to support the containment strategy. Testing focuses on the following groups:

- Persons with ARI who are travelers returning from areas with local COVID-19 activity
- Hospitalized persons; contacts of outbreak cases
- Institutionalized persons
- Healthcare workers
- Remote, isolated and/or Indigenous communities
- Vulnerable populations

Once the case numbers increase, with more extensive community transmission and pressures on testing resources, the goal of testing may need to be prioritized to support the mitigation strategy including the following:

- Testing persons at risk for serious disease
- Those likely to transmit virus within a healthcare facility or vulnerable community setting
- Those from whom COVID-19 disease would have an impact on delivering healthcare or critical infrastructure
- Those for whom exposure would put them at risk of testing positive

Additional groups may be considered for testing, depending on the stage of the pandemic, local policy and availability of reagents.

The COVID-19 testing should be completed for patients who meet testing criteria regardless of whether another pathogen is identified. Early data suggests that up to 30% of patients with COVID-19 can have co-infection with other respiratory viruses.

Further information on laboratory testing for COVID-19 is available from the Canadian Public Health Laboratory Network (CPHLN) COVID-19 Best Practices document (2).

## When to suspect the Middle East respiratory syndrome coronavirus

Limited data suggest that MERS-CoV can present as a co-infection with other viral pathogens. As such, in addition to specimens that are negative for conventional pathogens, those that do have other identified pathogens **but are consistent with suspect cases of MERS-CoV based on the Public Health Agency of Canada (PHAC) case definition, or alternatively provincial testing guidelines** should be tested. The details regarding testing and some control materials for method development are available from the National Microbiology Laboratory (NML). To date only a few PPHLs have developed the capacity to test for this pathogen in-house; all other PPHLs should forward the suspect specimens to the NML for further testing.

## When to suspect a novel influenza virus (including H7N9)

Influenza viruses that are positive on the initial influenza identification test but cannot be subtyped using RT-PCR should be further characterized. Laboratories that have the capacity to further characterize the specimens by novel subtyping PCRs or sequencing methods (e.g. sequence the HA, N, M or other

genes) to determine the subtype of the virus should do so. Those that lack this capacity should rely on the NML for further characterization. However, given that subtyping assays are usually less sensitive than the identification assays, weak positive results may not be typable. Based on local experience, each laboratory should evaluate these on a case-by-case basis, in concert with their local clinicians and public health colleagues.

Influenza-positive specimens outside the influenza season or obtained from patients with a history of exposure to animals (e.g. pigs or chickens) should be routinely submitted to the PPHL and/or NML for characterization.

Note: While initial analysis of the in-house assays used by many laboratories suggests they should be effective in identifying H7N9, it is difficult to determine their effect on the sensitivity of testing. This is particularly true of the performance of commercial assays whose primer sequences are not known. It is important for laboratories to have vendors supply information about the ability of their assays to detect novel influenza viruses. Laboratories using Level of Detection Tests should monitor viral sequences and their matches to the primers and probes in the assays.

### If a front line laboratory suspects a novel/emerging respiratory pathogen

The initial tests (as outlined above) would be similar but supplemental testing will be required at the PPHL. The laboratory should communicate with the clinician to ensure that the following specimens are collected:

- A second NPS/endotracheal aspirate or BAL—to be used for confirmation by the NML
- A viral throat swab (in viral transport media)—a number of avian influenza A viruses, including H7N9, have been detected in throat swabs. Multiple specimen types should be collected when novel influenza viruses are considered and, when possible, include both upper and lower respiratory tract specimens
- Acute and convalescent sera collection may be appropriate, depending on the specific virus suspected, and advice from NML and PPHLs. Serology is not recommended for patients suspected of influenza A(H7N9) or MERS-CoV infection. Some SARS-CoV-2 serology assays have been developed by several commercial providers and are being evaluated by NML and some PPHLs. These include ELISA-based and immunochromatographic point-of-care tests. Their role in clinical testing and public health has yet to be clarified, as insufficient data are available on sensitivity, specificity and positive and negative predictive values. Testing guidelines will be developed once assay performance characteristics have been elucidated and assays are validated for clinical testing

### If a provincial public health laboratory suspects a novel respiratory pathogen

- The PPHL should notify the patient's healthcare provider, local public health unit and Ministry of Health immediately when a suspect specimen is identified
- All specimens with suspected novel respiratory pathogens (as outlined below) must be forwarded to the NML for confirmatory testing. If a novel respiratory pathogen causes an epidemic or pandemic, with local transmission, only early specimens will be sent to NML for confirmatory testing. In addition, testing may be implemented at hospital or community laboratories, as has occurred during the SARS-CoV-2 pandemic
- Specimens suspected to contain a novel respiratory virus should be handled using CL2 with enhanced personal protective equipment **if manipulated outside a biosafety cabinet**

Note: Virus culture should not be conducted on respiratory specimens in a CL2 laboratory when a novel or emerging pathogen is suspected, as they are RG3 pathogens. Virus culture, if required, may be considered in a CL2 setting if the specimen has been tested for these pathogens and is negative by RT-PCR.

## Specimen transport

Specimens should be transported to the laboratory as soon as possible, preferably within 72 hours on ice packs. If a longer delay is anticipated, specimens should be frozen at -70°C, and transported on dry ice. However, specimens should not be frozen at -20°C, as this may affect the recovery of the virus if culture is required. If -70°C/dry ice is not available specimens should remain at 4°C and shipped as soon as possible. Specimens should be transported as diagnostic specimens per the usual practice for seasonal influenza specimens, and no enhanced precautions are necessary. See the PHAC SARS-CoV-2 Biosafety Advisory for more information (3).

Specimen tubes must be labelled and requisition completed correctly and fully, with matching patient names, unique identifiers and relevant clinical information.

## Authors' statement

The Respiratory Virus Infection Working Group of the Canadian Public Health Laboratory Network (CPHLN) is dedicated to providing leadership and guidance on topics related to respiratory viral pathogens, including laboratory response to emerging respiratory viruses. The Respiratory Virus Infection Working Group is comprised of leaders from public health laboratories across Canada.

## Acknowledgements

The Respiratory Virus Infection Working Group would like to thank members of the Canadian Public Health Laboratory Network (CPHLN) Secretariat, S Radons Arneson and D Marcino, for coordination of document synthesis. We would also like to thank the Laboratory Directors Council of the CPHLN for review of the document.

## References

1. Public Health Agency of Canada. Protocol for Microbiological Investigations of Severe Acute Respiratory Infections (SARI). Ottawa (ON): PHAC. https://www.canada.ca/en/public-health/services/emerging-respiratory-pathogens/protocol-microbiological-investigations-severe-acute-respiratory-infections-sari.html

2. National Collaborating Centre for Infectious Diseases. The Canadian Public Health Laboratory Network. https://nccid.ca/cphln/?highlight=cphln#038;hilite=%27cphln%27

3. Public Health Agency of Canada. SARS-CoV-2 (Severe acute respiratory syndrome-related coronavirus 2). Biosafety Advisory. Ottawa (ON): PHAC; updated February 29, 2020. https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/biosafety-directives-advisories-notifications/novel-coronavirus-january-27.html

# Managing immunization stress-related response: A contributor to sustaining trust in vaccines

C Meghan McMurtry[1,2,3,4]*

## Abstract

Adverse events following immunizations (AEFI) are important to identify and manage effectively so as to sustain trust in vaccines and optimize health. The AEFI category related to "anxiety about the immunization" was considered problematic as it did not adequately capture the range of stress responses that can occur. The currently used term for this category, immunization stress-related responses (ISRR), is broader, including the full spectrum of signs and symptoms that can arise in response to stress. ISRR can include vasovagal reactions (fainting), hyperventilation and functional neurological symptoms (e.g. weakness, nonepileptic seizures). It is based on a biopsychosocial framework in which biological (e.g. age, sex), psychological (e.g. preparedness, previous experiences, anxiety) and social factors (e.g. response by others, social media) interact to create an individual's stress response to the immunization process.

New guidance is available on prevention, early detection and management of ISRRs which is summarized in the article.



**Affiliations**

[1] Department of Psychology, University of Guelph, Guelph, ON

[2] Pediatric Chronic Pain Program, McMaster University, Hamilton, ON

[3] Adjunct Research Professor, Department of Paediatrics, Western University, London, ON

[4] Associate Scientist, Children's Health Research Institute, London, ON

***Correspondence:**
cmcmurtr@uoguelph.ca

***Suggested citation:*** McMurtry CM. Managing immunization stress-related response: A contributor to sustaining trust in vaccines. Can Commun Dis Rep 2020;46(6):210–8. https://doi.org/10.14745/ccdr.v46i06a10
***Keywords:*** vaccine, immunization, stress, fear, pain, syncope, fainting

## Introduction

Vaccines are a clear public health success story, protecting people from a number of diseases. Adverse events following immunizations (AEFI) are important to identify and manage appropriately so as to sustain trust in vaccines and optimize health. In 2015, the Global Advisory Committee on Vaccine Safety of the World Health Organization (WHO) brought together an Expert Working Group to discuss what was previously known as an "AEFI arising from anxiety about the immunization." Following review by the Global Advisory Committee in 2017 and 2018 and endorsement by the Strategic Advisory Committee for Vaccine Safety in April 2019, a detailed guidance manual for healthcare professionals, *Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization*, was published (1) along with a synopsis (2) and a peer-reviewed publication (3).

The guidance manual provides details on understanding, preventing, identifying and managing what are now termed "immunization stress-related responses," or ISRR (1).

The objective of this paper is to briefly describe ISRR, direct readers to detailed guidance on the topic and provide an overview of prevention and management.

This is the fifth article produced by the Canadian Vaccination Evidence Resource and Exchange Centre (CANVax) in the CANVax Briefs series. Multidisciplinary professionals at CANVax identify and develop useful resources to foster vaccine uptake (4,5).

### ISRR as part of AEFI

The safety of vaccinations is monitored globally, and AEFI, including events that are seen as arising from "anxiety" about the immunization, are grouped into five different categories (6). Naming these "anxiety reactions" is problematic for two reasons: anxiety does not paint an accurate or complete picture of what can be quite complex; and this description is indicative of a biomedical lens that classifies physiological responses as "physical" versus "psychological," which does not take into account that each individual's mind and body are intricately connected (1).

The term *immunization stress-related responses* (ISRR) acknowledges the full spectrum of signs and symptoms experienced in response to stress: vasovagal reactions (fainting), hyperventilation, and functional neurological symptoms (e.g. weakness, nonepileptic seizures), among others. The biopsychosocial framework helps to understand that biological (e.g. age, sex), psychological (e.g. preparedness, previous

experiences, anxiety) and social (e.g. peer behaviour and experiences, social media, community trust in health care) factors interact to develop an individual's stress response and ISRR (1).

ISRR and other AEFI require different prevention and treatment responses. For example, it is important to distinguish ISRR from anaphylaxis, which is life-threatening and requires urgent recognition and a particular pharmacologic response (intramuscular epinephrine) and expert management. ISRR is neither life-threatening nor helped by epinephrine, and requires different management.

### The role of the immunization process

Immunizations are typically delivered through injections. The process has several characteristics that can increase distress: pain from the injection, fear, sight of a needle, sight of blood, prolonged standing and responses by others in the environment (1,7). Children and adolescents are particularly at risk as immunizations are common at this age. Pain and fear can go hand in hand: the more scared an individual is about a needle, the more pain they report feeling (7,8). Most people who have high levels of needle fear report a previous negative experience (7,9–11).

There are short and long-term consequences of needle fear. In the short-term, individuals may require longer procedure times, have an increased risk of fainting, try to run away and experience greater distress and pain (7,8,10,12). Fear of the procedure can develop in the long-term along with fear of healthcare professionals, avoidance of medical procedures, vaccine hesitancy and lack of benefit from traditional pain management techniques (7,8,13–15).

### Identification of ISRR: Timing and manifestation

Understanding and recognizing ISRRs is key to facilitating prevention and appropriate management of this category of AEFIs. While other AEFIs occur only after immunization, an ISRR can occur immediately before, during or after immunization (1,6). The manifestations are acute stress responses, vasovagal reactions or dissociative neurological symptom reactions (DNSRs) (1). An acute stress response ("fight–flight–freeze" response) can vary in severity, from "butterflies in the stomach" and low to moderate levels of worry to more severe responses including difficulty breathing or rapid breathing/hyperventilation with tingling in the fingers and toes and increased heart rate (1,16,17). A vasovagal reaction is a fainting response that can cause a range of effects, from feeling mildly dizzy to losing consciousness due to insufficient blood flow to the brain (18).

An acute stress response may be followed by a vasovagal reaction after a sudden decrease in heart rate and a drop in blood pressure. Headache and nausea can also accompany stress reactions (1). Symptoms of an acute stress response and vasovagal reaction can present before, during or immediately after immunization, usually within five minutes (1).

DNSRs are characterized by neurological symptoms with no physical findings, otherwise known as functional neurological symptoms (19,20). Symptoms can include difficulty walking or moving a limb, weakness, tingling sensations in the muscles and nonepileptic seizures. These symptoms are considered involuntary. DNSRs have not been well documented or reported in individuals following immunization, but there are reports of "masses" or "clusters" of these reactions in multiple people in close proximity (21). The current evidence suggests that DNSRs result from complex multifactorial etiologies (22). DNSRs most commonly occur independently of immunization; a DNSR that develops after an immunization is best understood using a biopsychosocial framework in which the immunization process is one of a number of contributing factors (1).

A variety of postimmunization events, syndromes and disorders have been reported that have no confirmed relationship with immunization (1). These include complex regional pain syndrome (CRPS) type 1 with delayed onset; chronic fatigue syndrome; postural orthostatic tachycardia syndrome (POTS); and dissociative neurological symptom disorders (also known as conversion disorders) with delayed onset. They are not considered to be ISRR (1).

Acute stress response, vasovagal reaction and DNSR in individuals or "clusters" of individuals can occur independent of immunization. They have also been reported after immunization. WHO uses a detailed causality process to determine whether there is any relation between the symptoms and the immunization (6); more details can be found in the WHO ISRR guidance manual (1). ISRR are not caused by the vaccine, a defect in vaccine quality or an error in the immunization program or process.

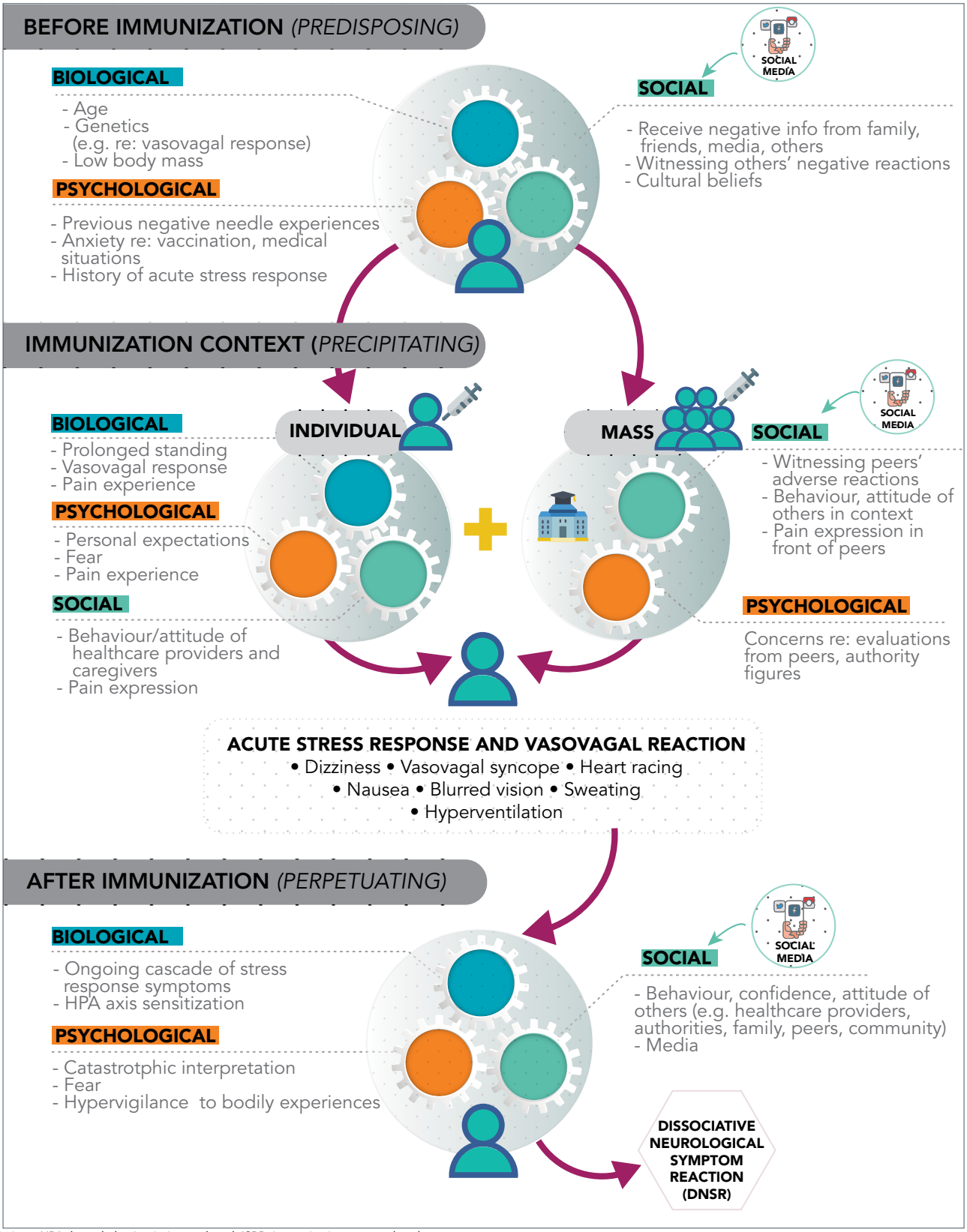
Each person who comes to be immunized has their own history, psychological strengths and vulnerabilities, and perceptions of the procedure and social context. Experiencing an ISRR is not the person's fault (1). **Figure 1** illustrates ISRR in individual and group contexts (1).

## Facilitating prevention and appropriate intervention of ISRR

### Prevention

Prevention relies on targeting predisposing risk factors. Clinicians should be educated on ISRRs, their prevention, screening and management (1). Brief reminders/educational materials on display in immunization clinics, for example, a poster describing the difference between anaphylaxis and ISRR, could be helpful. As social media can play a particularly strong negative role in mass immunization contexts, including school immunization programs, communication is important before, during and after

**Figure 1: ISRR in individual and group contexts**




Abbreviations: HPA, hypothalamic-pituitary-adrenal; ISRR, immunization stress-related response
There are three broad time points: before the immunization (historical, predisposing factors); in the immunization context (precipitating factors, initial response); and after immunization (delayed response influenced by perpetuating factors)
Risk factors: shapes with a patterned fill show examples of potential risk factors for an ISRR; gear shapes show the dynamic interactions between risk factors
Progression: the person being immunized is shown at different times with example risk factors leading to a cascade of symptoms (initial response, ongoing) consistent with ISRR. However, not everyone progresses step-by-step from one stage to the next. For example, a dissociative neurological symptom reaction (DNSR) does not need to follow an acute stress response
Social media's potential to provide negative information is highlighted
Source: Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization (1)

the immunization to reduce the risk of ISRR (23). Planning for mass immunizations should take into account existing rates of ISRR and vulnerability factors, that is, the age and sex of the recipients because adolescents and females are at greater risk of a vasovagal reaction (24). Therefore, planning for immunization clinics should include familiarizing healthcare providers with how to screen, prevent and manage vasovagal reactions (1). Targeted education sessions teaching coping strategies can also be helpful (25,26).

In every clinic, environmental strategies can reduce risk factors for ISRR. The immunization environment should be at a comfortable temperature (rather than overheated) and those at risk of an ISRR should be vaccinated in private (1). The flow of individuals through the clinic should be such that the waiting area has only a few people (i.e. should not be crowded) and no one should be waiting for long. Allowing space to sit rather than having to stand for a long time is helpful. To build trust, the healthcare team should be calm, confident and friendly and able to communicate well with the recipients and any caregivers; they will also need to address any caregivers who are nervous and exacerbating fear in the vaccine recipient (1).

## Screening

People at high risk for ISRR should be identified by screening for high levels of needle fear and previous negative experiences with needles, including fainting (1). For school vaccination campaigns, teachers, school nurses or other staff may be able to flag students at high risk for ISRR ahead of time, or individuals may self-identify. Late school-age and adolescent youth appear to be at higher risk of ISRR than other age groups. Individuals with a history of vasovagal reactions, including syncope and/or a high level of needle fear may be particularly at risk. Individuals with preexisting anxiety disorders and/or developmental disorders (including autism spectrum disorder) may also need extra time and care (1).

Each vaccine recipient should be asked if they have ever fainted (i.e. lost consciousness) and/or had prodromal symptoms (e.g. felt dizzy, nauseated and/or clammy and/or saw spots) before, during or after a needle procedure (1). Individuals who have a history of vasovagal reactions should be immunized in a seated or supine position and only move to sitting (from supine) or standing (from sitting) if there have no signs of a vasovagal reaction. Ideally the individual should stay seated for 15 to 30 minutes following the procedure, and the healthcare provider should monitor them for signs of a vasovagal reaction (1). In addition, the muscle tension technique can be taught to and used by the recipient (see Targeted Interventions for ISRR).

Although no current gold standard exists for screening for high levels of fear, it is recommended that healthcare providers ask vaccine recipients the questions shown in **Table 1** (27). Caregivers can be asked similar questions about their younger children.

**Table 1: Questions to ask to screen for high levels of needle fear**

| Age group, years | Question |
|---|---|
| 5–8 | 1) How afraid of needles are you? Not at all; a little bit; a medium amount; a lot; very, very much/most possible? |
| | 2) Do you try hard to miss having a needle because you are so scared? |
| Older than 8 | 1) How afraid of needles are you? Not afraid; a little bit; a moderate amount; a lot; or the most afraid possible? |
| | 2) Do you think this level is higher than it should be (or higher than that of most of your friends)? |
| | 3) Do you avoid getting needles because you are afraid? |

## Universal interventions

All recipients should be shown age-appropriate ways to manage pain and low to moderate fear (1,27). See **Table 2** for physical, psychological, procedural and pharmacological strategies recommended for different age groups. A supportive caregiver could also be present to help with coping strategies. For further details see *Reducing pain during vaccine injections: clinical practice guideline* (27).

## Targeted interventions for ISRR

If an individual is at elevated risk for an ISRR, additional measures need to be put in place, such as avoiding having them wait in the general waiting area, immunizing them at the beginning of the clinic and immunizing them in private (1). These strategies are designed to reduce contagion of fear and other negative emotions as well as to contain any negative effects of an ISRR, should one occur. These individuals may benefit from having a calm, supportive caregiver or friend with them; as noted above, a fearful caregiver or friend can exacerbate the situation and they need to be addressed immediately (1).

If an individual's level of fear is high but they are not avoiding the vaccination, two approaches could be used: first, identify what can be done in the immunization clinic to create a positive experience for the individual, for example, taking more time, making further environmental modifications, etc.; and second, determine whether treatment of the needle fear by a mental health professional outside of the immunization context is necessary before future immunizations (1,28). If high levels of needle fear and avoidance are present, consider delaying the needle to address these factors. For extreme fear, pharmacological strategies (e.g. anxiolytics, sedations (1)) could also be considered if the expertise is available.

If the individual is at risk for a vasovagal reaction, immunizing them in a reclining or supine position while they are using the muscle tension technique can be helpful (1). Muscle tension keeps an individual's blood pressure up and prevents the precipitous drop that can lead to a faint. This technique has been recommended for those aged seven years and older (adolescents are at greater risk for vasovagal syncope) (27,28). First, the

**Table 2: Strategies for managing vaccine-related pain and fear in different age groups**

| Type of strategy | Newborn | Infant (1–35 months) | Preschool (3–5 years) | School-aged (6–12 years) | Adolescent (13–18 years) | Adult (19 years+) |
|---|---|---|---|---|---|---|
| **Procedural** | | | | | | |
| Inject into anterolateral thigh | ✓ | ✓ (1–11 months) | – | – | – | – |
| No aspiration when injecting | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Give most painful vaccine last | ✓ | ✓ (or simultaneous injection 0–1 year) | ✓ | ✓ | ✓ | ✓ |
| **Physical** | | | | | | |
| Skin-to-skin (kangaroo care) before, during, after | ✓ | – | – | – | – | – |
| Cradled in parent's arms | ✓ | ✓ | – | – | – | – |
| Breastfeeding before, during, after OR sweet-tasting solutions before and/or nonnutritive sucking before, during, after | ✓ | ✓ | – | – | – | – |
| Seated upright[a] | – | – | ✓ | ✓ | ✓ | ✓ |
| External vibrating device with cold | – | – | ✓ | ✓ | ✓ | – |
| **Communication[b] and psychological** | | | | | | |
| Calm voice, simple language | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Don't say it won't hurt | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Use neutral words to signal procedure (e.g. "*1, 2, 3, here we go*") | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Avoid repeated excessive reassurance (e.g. "*it's okay, it's okay, it's okay*") before, during, after | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Talk about things other than the procedure (verbal distraction) before, during, after | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Distraction (age appropriate) | – | ✓ (e.g. toy, video with adult coaching to pay attention to distractor) | ✓ (e.g. blowing bubbles, toys, video, singing) | ✓ (e.g. video game, video, blowing bubbles, toys, music) | – | – |
| Breathing strategy | – | – | ✓ (breathing with toy) | ✓ (breathing with toy) | – | ✓ (cough, breath-hold) |
| **Pharmacological** | | | | | | |
| Topical anesthetic applied before (check product instructions for time)[c] | – | ✓ | ✓ | ✓ | ✓ | ✓ |
| Vapocoolant spray right before | – | – | – | – | – | ✓ |

Abbreviations: ✓, recommended strategy for a particular age group; –, strategy is not recommended for a particular age group
[a] Individuals with a history of vasovagal reaction should be immunized in a seated or supine position and only move to sitting (from supine) or standing (from sitting) if there have no signs of a vasovagal reaction
[b] The communication strategies for newborns and young infants are primarily directed to the caregiver
[c] Topical anesthetic should be used when feasible; for adolescents and adults, topical anesthetic may be used if resources are available and the person is at high risk for an immunization stress-related response (ISRR)

individual tenses their major muscle groups (e.g. abdomen, legs, contralateral arm to where the injection will be administered) for 15 to 30 seconds until they feel flushed or warm in the face. Next, they release the tension for 15 to 30 seconds, but they do not fully relax. They repeat these steps in cycles before, during and after the procedure until they have no prodromal symptoms.

It is critical to differentiate anaphylaxis from a DNSR (1). (The WHO ISRR guidance manual contains a table that can help

healthcare providers distinguish between anaphylaxis and an ISRR (29)). If an individual loses consciousness following a vaccination, it could be a result of vasovagal syncope or anaphylaxis. Anaphylaxis is potentially life-threatening and requires medication (30). While the recipient is in the recovery position (supine, on their side), a healthcare provider should monitor the recipient's pulse, respiration, blood pressure and peripheral circulation (1), watch their skin for rash or swelling and listen to their lungs for wheeze or stridor.

If an ISRR has been identified, it is important to communicate that the response is not due to a vaccine product or procedural error; that the response is a known event that staff resolve by following specific guidance; and that the response can resolve spontaneously without medication or hospitalization.

A DNSR that occurs following an immunization does not causally implicate either the immunization or the immunization process. A specific assessment is used to determine causality (6); the WHO ISRR guidance manual provides a list that can help to diagnose a DNSR. Examples include symptoms that are inconsistent with known disorders and inconsistent presentation of symptoms (e.g. that disappear inexplicably or do not respond typically to interventions) (31). Nonepileptic seizures are one example of a DNSR; these resemble epileptic seizures but do not have neural discharges in the way epileptic seizures do (epileptic seizures are differentiated from nonepileptic seizures in the WHO ISRR guidance manual (32)). Nonepileptic seizures are typically a diagnosis of exclusion (33). Although an electroencephalogram is the gold standard assessment for seizures, conducting one may not be practicable.

A DNSR may resolve spontaneously or may require the involvement of a multidisciplinary team including a mental health professional. The biopsychosocial framework that is used to understand ISRR should also guide treatment (1). Medical and psychological expertise is needed for further assessment and management to reduce functional disability. Treatment is specific to the presenting symptoms but may include physiotherapy, cognitive behavioural therapy and/or pharmacological strategies (34,35). In the short term, the healthcare providers at the immunization clinic should attempt to put the affected individual and others present at ease. They should note that anxiety about and fear of immunization are normal and can result in a bodily response that may seem extreme but can resolve spontaneously without any injury (1). The affected individual should be kept in a separate, calm, quiet space with only key people present. The healthcare providers can answer the questions raised by the affected individual and/or their caregiver(s). If the recipient and caregiver(s) are relatively calm, they may be able to be distracted by talking about something else or listening to music to further calm them. The goal is to encourage return to normal activity (1).

Similar or identical symptoms appearing in more than one person with no physiological cause have been the source of attention and curiosity for hundreds of years; the "spread" is thought to be due to shared beliefs and contagion of anxiety and fear (36–39). These "clusters" have been known as "mass psychogenic illness" or "mass hysteria" and have been reported inside and outside the immunization context (21,40). These terms can be inflammatory and demeaning to the affected individuals (1).

Trying to complete a mass vaccination campaign in a short amount of time is a risk factor for the development of ISRR (1). The biopsychosocial framework is used to understand clusters, with particular attention paid toward understanding and managing social factors. Known clusters in the immunization context have occurred in adolescents and adults but not in infants and young children (1). Anaphylaxis is rare and extremely unlikely to occur in clusters (1). If an ISRR cluster occurs, the affected individuals should be separated from others and each other to enable containment and appropriate management (1). The general strategies outlined above (i.e. training of staff, communication, environmental modifications, screening for people at risk for ISRR and targeted strategies such as privacy) are also critical in mass immunization contexts. Community leaders and healthcare providers who are known to the recipients can help keep them calm and comfortable. Educational materials such as posters differentiating anaphylaxis from ISRR and epileptic seizures from nonepileptic seizures could be designed and posted in the clinic (1).

## Conclusion

In summary, ISRRs are the redefined way to think about, identify and manage what was previously known as AEFI stemming from "anxiety" related to the immunization. ISRRs can occur before, during or after the immunization and are not due to the vaccine product itself or an error in the process. Prevention strategies include proactive communication, managing social media use and in-clinic environmental strategies. Screening can identify those with increased risk of an ISRR, including those with high levels of needle fear or with previous vasovagal reaction. Age-appropriate pain management strategies should be standard for all immunization recipients. Targeted interventions for those experiencing an ISRR include muscle tension for vasovagal reactions, reducing vaccine recipients' fear, increasing comfort and avoiding the contagion of fear and misinformation.

Understanding the nature of ISRRs and their occurrence provides an opportunity for their prevention and appropriate management, warding off future negative reactions towards immunization and health care in general, and contributing to sustaining trust in vaccines. The ISRR should be reported as part of AEFI surveillance.

See the WHO ISRR guidance manual for further information (1) and the CANVax website for updates.

## Author's statement

CMM — Conceptualization, writing – original draft, review, editing. This manuscript is based on *Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization*. The content of the World Health Organization (WHO) immunization stress-related response (ISRR) guidance manual was developed by M Gold, NE MacDonald, CM McMurtry, R Pless and U Heininger with coordination and supervision by MR Balakrishnan and P Zuber and support by L Menning and O Benes from WHO.

### Conflict of interest

CMM was a member of the immunization stress-related response (ISRR) expert working group. There are no other conflicts of interest.

## Acknowledgements

The author is grateful for the efforts of the other members of the immunization stress-related response (ISRR) working group (MR Balakrishnan, O Benes, M Gold, U Heininger, NE MacDonald, L Menning, R Pless and P Zuber). Thank you also to NE MacDonald for reviewing and commenting on an earlier version of this manuscript. Thank you to K Constantin for reviewing the French version of the article for accuracy.

## Funding

No external funding was received.

## References

1. World Health Organization. Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization. Geneva (CH): World Health Organization; 2019.

2. World Health Organization. Immunization stress-related responses: a synopsis of the manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization. Geneva (CH): World Health Organization; 2019.

3. Gold MS, MacDonald NE, McMurtry CM, Balakrishnan MR, Heininger U, Menning L, Benes O, Pless R, Zuber PL. Immunization stress-related response - Redefining immunization anxiety-related reaction as an adverse event following immunization. Vaccine 2020 Mar;38(14):3015–20. DOI PubMed

4. CANVax. The Canadian Vaccination Evidence Resource and Exchange Centre: Your online immunization resource centre. Ottawa (ON) (updated 2019; accessed 2020-03-17). https://www.canvax.ca/

5. MacDonald NE, Dubé E. A new resource to summarize evidence on immunization from the Canadian Vaccination Evidence Resource and Exchange Centre (CANVax). Can Commun Dis Rep 2020;46(1):16–9. DOI PubMed

6. World Health Organization. Causality assessment of an adverse event following immunization (AEFI): User manual for the revised WHO classification. Geneva (CH): World Health Organization; 2018.

7. McMurtry CM, Pillai Riddell R, Taddio A, Racine N, Asmundson GJ, Noel M, Chambers CT, Shah V; HELPinKids&Adults Team. HELPinKids&Adults Team. Far from "just a poke": common painful needle procedures and the development of needle fear. Clin J Pain 2015;31(10 Suppl):S3–11. DOI PubMed

8. Taddio A, Ipp M, Thivakaran S, Jamal A, Parikh C, Smart S, Sovran J, Stephens D, Katz J. Survey of the prevalence of immunization non-compliance due to needle fears in children and adults. Vaccine 2012;30(32):4807–12. DOI PubMed

9. Kleinknecht RA. Acquisition of blood, injury, and needle fears and phobias. Behav Res Ther 1994;32(8):817–23. DOI PubMed

10. Öst LG. Blood and injection phobia: background and cognitive, physiological, and behavioral variables. J orm Psychol 1992;101(1):68–74. DOI PubMed

11. Öst LG. Acquisition of blood and injection phobia and anxiety response patterns in clinical patients. Behav Res Ther 1991;29(4):323–32. DOI PubMed

12. Page AC. Blood-injury phobia. Clin Psychol Rev 1994;14(5):443–61. DOI

13. Noel M, Chambers CT, Petter M, McGrath PJ, Klein RM, Stewart SH. Pain is not over when the needle ends: a review and preliminary model of acute pain memory development in childhood. Pain Manag 2012;2(5):487–97. DOI PubMed

14. Gullone E. The development of normal fear: a century of research. Clin Psychol Rev 2000;20(4):429–51. DOI PubMed

15. Hamilton JG. Needle phobia: a neglected diagnosis. J Fam Pract 1995;41(2):169–75. PubMed

16. McEwen BS. Physiology and neurobiology of stress and adaptation: central role of the brain. Physiol Rev 2007;87(3):873–904. DOI PubMed

17. Selye H. Forty years of stress research: principal remaining problems and misconceptions. Can Med Assoc J 1976;115(1):53–6. PubMed

18. van Lieshout JJ, Wieling W, Karemaker JM, Eckberg DL. The vasovagal response. Clin Sci (Lond) 1991;81(5):575–86. DOI PubMed

19. World Health Organization. ICD-11 for Mortality and Morbidity Statistics (ICD-11 MMS): 2018 version. Geneva (CH): World Health Organization; 2018 (updated 2019-04; accessed 2020-03-17). https://icd.who.int/browse11/l-m/en

20. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fifth edition. Washington (DC): American Psychiatric Publishing; 2013. DOI

21. Loharikar A, Suragh TA, MacDonald NE, Balakrishnan MR, Benes O, Lamprianou S, Hyde TB, McNeil MM. Anxiety-related adverse events following immunization (AEFI): A systematic review of published clusters of illness. Vaccine 2018;36(2):299–305. DOI PubMed

22. Bodde NM, Brooks JL, Baker GA, Boon PA, Hendriksen JG, Mulder OG, Aldenkamp AP. Psychogenic non-epileptic seizures--definition, etiology, treatment and prognostic issues: a critical review. Seizure 2009;18(8):543–53. DOI PubMed

23. Suragh TA, Lamprianou S, MacDonald NE, Loharikar AR, Balakrishnan MR, Benes O, Hyde TB, McNeil MM. Cluster anxiety-related adverse events following immunization (AEFI): an assessment of reports detected in social media and those identified using an online search engine. Vaccine 2018;36(40):5949–54. DOI PubMed

24. Braun MM, Patriarca PA, Ellenberg SS. Syncope after immunization. Arch Pediatr Adolesc Med 1997;151(3):255–9. DOI PubMed

25. Freedman T, Taddio A, Alderman L, McDowall T, deVlaming-Kot C, McMurtry CM, MacDonald N, Alfieri-Maiolo A, Stephens D, Wong H, Boon H; Pain Pain Go Away Team. The CARD™ System for improving the vaccination experience at school: results of a small-scale implementation project on student symptoms. Paediatr Child Health 2019;24 Suppl 1:S42–53. DOI PubMed

26. Taddio A, Alderman L, Freedman T, McDowall T, McMurtry CM, MacDonald N, deVlaming-Kot C, Alfieri-Maiolo A; Pain Pain Go Away Team. The CARD™ System for improving the vaccination experience at school: results of a small-scale implementation project on program delivery. Paediatr Child Health 2019;24 Suppl 1:S54–67. DOI PubMed

27. Taddio A, McMurtry CM, Shah V, Riddell RP, Chambers CT, Noel M, MacDonald NE, Rogers J, Bucci LM, Mousmanis P, Lang E, Halperin SA, Bowles S, Halpert C, Ipp M, Asmundson GJ, Rieder MJ, Robson K, Uleryk E, Antony MM, Dubey V, Hanrahan A, Lockett D, Scott J, Bleeker EV, Team HA; HELPinKids&Adults. Reducing pain during vaccine injections: clinical practice guideline. CMAJ 2015;187(13):975–82. DOI PubMed

28. McMurtry CM, Taddio A, Noel M, Antony MM, Chambers CT, Asmundson GJ, Pillai Riddell R, Shah V, MacDonald NE, Rogers J, Bucci LM, Mousmanis P, Lang E, Halperin S, Bowles S, Halpert C, Ipp M, Rieder MJ, Robson K, Uleryk E, Votta Bleeker E, Dubey V, Hanrahan A, Lockett D, Scott J. Exposure-based Interventions for the management of individuals with high levels of needle fear across the lifespan: a clinical practice guideline and call for further research. Cogn Behav Ther 2016;45(3):217–35. DOI PubMed

29. World Health Organization. Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization. Geneva (CH): World Health Organization; 2019. Table 4.1, Differences between anaphylaxis, general acute stress response and vasovagal reaction with syncope. https://apps.who.int/iris/bitstream/handle/10665/330277/9789241515948-eng.pdf?sequence=1&isAllowed=y

30. Rüggeberg JU, Gold MS, Bayas JM, Blum MD, Bonhoeffer J, Friedlander S, de Souza Brito G, Heininger U, Imoukhuede B, Khamesipour A, Erlewyn-Lajeunesse M, Martin S, Mäkelä M, Nell P, Pool V, Simpson N; Brighton Collaboration Anaphylaxis Working Group. Anaphylaxis: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2007;25(31):5675–84. DOI PubMed

31. World Health Organization. Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization. Geneva (CH): World Health Organization; 2019. Table 4.2, Clues to diagnosis of a DNSR. https://apps.who.int/iris/bitstream/handle/10665/330277/9789241515948-eng.pdf?sequence=1&isAllowed=y

32. World Health Organization. Immunization stress-related response: a manual for program managers and health professionals to prevent, identify and respond to stress-related responses following immunization. Geneva (CH): World Health Organization; 2019. Table 4.3, Differentiating a non-epileptic seizure (subgroup of DNSR) from epilepsy. https://apps.who.int/iris/bitstream/handle/10665/330277/9789241515948-eng.pdf?sequence=1&isAllowed=y

33. Stone J, LaFrance WC Jr, Brown R, Spiegel D, Levenson JL, Sharpe M. Conversion disorder: current problems and potential solutions for DSM-5. J Psychosom Res 2011;71(6):369–76. DOI PubMed

34. Nielsen G, Stone J, Matthews A, Brown M, Sparkes C, Farmer R, Masterton L, Duncan L, Winters A, Daniell L, Lumsden C, Carson A, David AS, Edwards M. Physiotherapy for functional motor disorders: a consensus recommendation. J Neurol Neurosurg Psychiatry 2015;86:1113–9. DOI PubMed

35. FitzGerald TL, Southby AK, Haines TP, Hough JP, Skinner EH. Is physiotherapy effective in the management of child and adolescent conversion disorder? A systematic review. J Paediatr Child Health 2015;51(2):159–67. DOI PubMed

36. Jones TF, Craig AS, Hoy D, Gunter EW, Ashley DL, Barr DB, Brock JW, Schaffner W. Mass psychogenic illness attributed to toxic exposure at a high school. N Engl J Med 2000;342(2):96–100. DOI PubMed

37. Bauch CT, Galvani AP. Epidemiology. Social factors in epidemiology. Science 2013;342(6154):47–9. DOI PubMed

38. Bartholomew RE. 'Mystery illness' in Western New York: is social networking spreading mass hysteria? Skeptical Inquirer 2012;36(4):26–9.

39. Boss LP. Epidemic hysteria: a review of the published literature. Epidemiol Rev 1997;19(2):233–43. https://doi.org/10.1093/oxfordjournals.epirev.a017955 PubMed40. Clements CJ. Mass psychogenic illness after vaccination. Drug Saf 2003;26(9):599–604. DOI PubMed

40. Clements CJ. Mass psychogenic illness after vaccination. Drug Saf 2003;26(9):599–604. DOI PubMed

# CCDR

## CANADA COMMUNICABLE DISEASE REPORT

Public Health Agency of Canada
130 Colonnade Road
Address Locator 6503B
Ottawa, Ontario  K1A 0K9
phac.ccdr-rmtc.aspc@canada.ca

To promote and protect the health of Canadians through leadership, partnership, innovation and action in public health.

Public Health Agency of Canada

Published by authority of the Minister of Health.



This publication is also available online at

https://www.canada.ca/ccdr

Également disponible en français sous le titre :
**Relevé des maladies transmissibles au Canada**