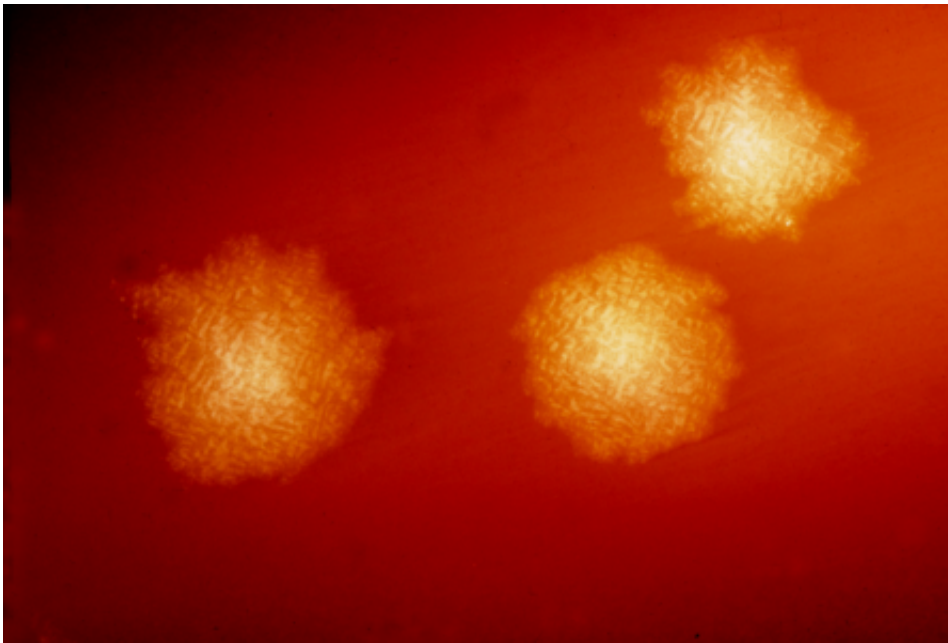


# Machine learning model provides rapid prediction of *C. difficile* infection risk

March 26 2018

---



This photograph depicts *Clostridium difficile* colonies after 48hrs growth on a blood agar plate; Magnified 4.8X. *C. difficile*, an anaerobic gram-positive rod, is the most frequently identified cause of antibiotic-associated diarrhea (AAD). It accounts for approximately 15–25% of all episodes of AAD. Credit: CDC

Every year nearly 30,000 Americans die from an aggressive, gut-infecting bacteria called *Clostridium difficile* (*C. difficile*), which is resistant to many common antibiotics and can flourish when antibiotic treatment kills off beneficial bacteria that normally keep it at bay. Investigators from Massachusetts General Hospital (MGH), the

University of Michigan (U-M) and Massachusetts Institute of Technology (MIT) now have developed investigational "machine learning" models, specifically tailored to individual institutions, that can predict a patient's risk of developing *C. difficile* much earlier than it would be diagnosed with current methods. Preliminary data from their study, which is being published today in *Infection Control and Hospital Epidemiology*, were presented last October at the ID Week 2017 conference.

"Despite substantial efforts to prevent *C. difficile* infection and to institute early treatment upon diagnosis, rates of infection continue to increase," says Erica Shenoy, MD, PhD, of the [MGH Division of Infectious Diseases](#), co-senior author of the study and assistant professor of Medicine at Harvard Medical School. "We need better tools to identify the highest risk patients so that we can target both prevention and treatment interventions to reduce further transmission and improve patient outcomes."

The authors note that most previous models of *C. difficile* infection risk were designed as "one size fits all" approaches and included only a few risk factors, which limited their usefulness. Co-lead authors Jeeheh Oh, a U-M graduate student in Computer Science and Engineering, and Maggie Makar, MS, of MIT's Computer Science and Artificial Intelligence Laboratory and their colleagues took a "big data" approach that analyzed the whole electronic health record (EHR) to predict a patient's *C. difficile* risk throughout the course of hospitalization. Their method allows the development of institution-specific models that could accommodate different patient populations, different EHR systems and factors specific to each institution.

"When data are simply pooled into a one-size-fits-all [model](#), institutional differences in patient populations, hospital layouts, testing and treatment protocols, or even in the way staff interact with the EHR can lead to

differences in the underlying data distributions and ultimately to poor performance of such a model," says Jenna Wiens, PhD, assistant professor of Computer Science and Engineering at U-M and co-senior author of the study. "To mitigate these issues, we take a hospital-specific approach, training a model tailored to each institution."

Using their machine-learning-based model, the investigators analyzed de-identified data - including individual patient demographics and medical history, details of their admission and daily hospitalization, and the likelihood of exposure to *C. difficile* - from the EHRs of almost 257,000 patients admitted to either MGH or to Michigan Medicine - U-M's academic medical center - over periods of two years and six years, respectively. The model generated daily risk scores for each individual patient that, when a set threshold is exceeded, classify patients as at high risk.

Overall, the models were highly successful at predicting which patients would ultimately be diagnosed with *C. difficile*. In half of those who were infected, accurate predictions could have been made at least five days before diagnostic samples were collected, which would allow highest-risk patients to be the focus of targeted antimicrobial interventions. If validated in prospective studies, the risk prediction score could guide early screening for *C. difficile*. For patients diagnosed earlier in the course of disease, initiation of treatment could limit the severity of the illness, and patients with confirmed *C. difficile* could be isolated and contact precautions instituted to prevent transmission to other patients.

The research team has made the algorithm code freely available [here](#) for others to review and adapt for their individual institutions. Shenoy notes that facilities that explore applying similar algorithms to their own institutions will need to assemble the appropriate local subject-matter experts and validate the performance of the models in their institutions.

Study co-author Vincent Young, MD, PhD, the William Henry Fitzbutler Professor in the Department of Internal Medicine at U-M, adds, "This represents a potentially significant advance in our ability to identify and ultimately act to prevent infection with *C. difficile*. The ability to identify [patients](#) at greatest risk could allow us to focus expensive and potentially limited prevention methods on those who would gain the greatest potential benefit. I think that this project is a great example of a 'team science' approach to addressing complex biomedical questions to improve healthcare, which I expect to see more of as we enter the era of precision health."

**More information:** *Infection Control and Hospital Epidemiology* (2018). [DOI: 10.1017/ice.2018.16](https://doi.org/10.1017/ice.2018.16)

Provided by Massachusetts General Hospital

Citation: Machine learning model provides rapid prediction of *C. difficile* infection risk (2018, March 26) retrieved 26 April 2024 from <https://medicalxpress.com/news/2018-03-machine-rapid-difficile-infection.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--