

# Machine learning enables an 'almost perfect' diagnosis of an elusive global killer

October 24 2022

---



Credit: Pixabay/CC0 Public Domain

Sepsis, the overreaction of the immune system in response to an infection, causes an estimated 20% of deaths globally and as many as 20 to 50% of U.S. hospital deaths each year. Despite its prevalence and

severity, however, the condition is difficult to diagnose and treat effectively.

The disease can cause decreased blood flow to vital organs, inflammation throughout the body, and abnormal blood clotting. Therefore, if [sepsis](#) isn't recognized and treated quickly, it can lead to shock, organ failure, and death. But it can be difficult to identify which pathogen is causing sepsis, or whether an infection is in the [bloodstream](#) or elsewhere in the body. And in many patients with symptoms that resemble sepsis, it can be challenging to determine whether they truly have an infection at all.

Now, researchers at the Chan Zuckerberg Biohub (CZ Biohub), the Chan Zuckerberg Initiative (CZI), and UC San Francisco (UCSF) have developed a new diagnostic method that applies machine learning to advanced genomics data from both microbe and host—to identify and predict sepsis cases. As reported on October 20, 2022 in *Nature Microbiology*, the approach is surprisingly accurate, and has the potential to far exceed current diagnostic capabilities.

"Sepsis is one of the top 10 public health issues facing humanity," said senior author Chaz Langelier, M.D., Ph.D., an associate professor of medicine in UCSF's Division of Infectious Diseases and a CZ Biohub Investigator. "One of the key challenges with sepsis is diagnosis. Existing [diagnostic tests](#) are not able to capture the dual-sided nature of the disease—the infection itself and the host's immune response to the infection."

Current sepsis diagnostics focus on detecting bacteria by growing them in culture, a process that is "essential for appropriate antibiotic therapy, which is critical for sepsis survival," according to the researchers behind the new method. But culturing these pathogens is time-consuming and doesn't always correctly identify the bacterium that is causing the

infection. Similarly for viruses, PCR tests can detect that viruses are infecting a patient but don't always identify the particular virus that's causing sepsis.

"This results in clinicians being unable to identify the cause of sepsis in an estimated 30 to 50% of cases," Langelier said. "This also leads to a mismatch in terms of the antibiotic treatment and the pathogen causing the problem."

In the absence of a definitive diagnosis, doctors often prescribe a cocktail of antibiotics in an effort to stop the infection, but the overuse of antibiotics has led to increased antibiotic resistance worldwide. "As physicians, we never want to miss a case of infection," said Carolyn Calfee, M.D., M.A.S., a professor of medicine and anesthesia at UCSF and co-senior author of the new study. "But if we had a test that could help us accurately determine who doesn't have an infection, then that could help us limit [antibiotic use](#) in those cases, which would be really good for all of us."

### **Eliminating ambiguity**

The researchers analyzed whole blood and [plasma samples](#) from more than 350 critically ill patients who had been admitted to UCSF Medical Center or the Zuckerberg San Francisco General Hospital between 2010 and 2018.

But rather than relying on cultures to identify pathogens in these samples, a team led by CZ Biohub scientists Norma Neff, Ph.D., and Angela Pisco, Ph.D., instead used metagenomic next-generation sequencing (mNGS). This method identifies all the nucleic acids or genetic data present in a sample, then compares those data to reference genomes to identify the microbial organisms present. This technique allows scientists to identify genetic material from entirely different

kingdoms of organisms—whether bacteria, viruses, or fungi—that are present in the same sample.

However, detecting and identifying the presence of a pathogen alone isn't enough for accurate sepsis diagnosis, so the Biohub researchers also performed transcriptional profiling—which quantifies gene expression—to capture the patient's response to infection.

Next they applied machine learning to the mNGS and transcriptional data to distinguish between sepsis and other critical illnesses and thus confirm the diagnosis. Katrina Kalantar, Ph.D., a lead computational biologist at CZI and co-first author of the study, created an integrated host-microbe model trained on data from patients in whom either sepsis or non-infectious systemic inflammatory illnesses had been established, which enabled sepsis diagnosis with very high accuracy.

"We developed the model by looking at a bunch of metagenomics data alongside results from traditional clinical tests," Kalantar explained. To start, the researchers identified changes in gene expression between patients with confirmed sepsis and non-infectious systemic inflammatory conditions that appear clinically similar, then used machine learning to identify the genes that could best predict those changes.

The researchers found that when traditional bacterial culture identified a sepsis-causing pathogen, there was usually an overabundance of genetic material from that pathogen in the corresponding plasma sample analyzed by mNGS. With that in mind, Kalantar programmed the model to identify organisms present in disproportionately high abundance compared to other microbes in the sample, and to then compare those to a reference index of well-known sepsis-causing microbes.

"In addition to that, we also noted any viruses that were detected, even if

they were at lower levels, because those really shouldn't be there," Kalantar explained. "With this relatively straightforward set of rules, we were able to do pretty well."

### **'Almost perfect' performance**

The researchers found that the mNGS method and their corresponding model worked better than expected: They were able to identify 99% of confirmed bacterial sepsis cases, 92% of confirmed viral sepsis cases, and were able to predict sepsis in 74% of clinically suspected cases that hadn't been definitively diagnosed.

"We were expecting good performance, or even great performance, but this was almost perfect," said Lucile Neyton, Ph.D., a postdoctoral researcher in the Calfee lab and co-first author of the study. "By using this approach, we get a pretty good idea of what is causing the disease, and we know with relatively high confidence if a patient has sepsis or not."

The team was also excited to discover that they could use this combined host-response and microbe detection method to diagnose sepsis using plasma samples, which are routinely collected from most patients as part of standard clinical care. "The fact that you can actually identify sepsis patients from this widely available, easy-to-collect sample type has big implications in terms of practical utility," Langelier said.

The idea for the work stemmed from previous research by Langelier, Kalantar, Calfee, UCSF researcher and CZ Biohub President Joe DeRisi, Ph.D., and their colleagues, in which they used mNGS to effectively diagnose lower respiratory tract infections in critically ill patients. Because the method worked so well, "we wanted to see if the same type of approach could work in the context of sepsis," said Kalantar.

## Broader implications

The team hopes to build upon this successful [diagnostic technique](#) by developing a model that can also predict antibiotic resistance from pathogens detected with this method. "We've had some success doing that for respiratory infections, but no one has come up with a good approach for sepsis," Langelier said.

Furthermore, the researchers hope to eventually be able to predict outcomes of patients with sepsis, "such as mortality or length of stay in the hospital, which would provide key information that would allow clinicians to better care for their patients and match resources to the patients who need them the most," Langelier said.

"There's a lot of potential for novel sequencing approaches such as this to help us more precisely identify the causes of a patient's critical illness," added Calfee. "If we can do that, it's the first step towards precision medicine and understanding what's going on at an individual patient level."

**More information:** Katrina L. Kalantar et al, Integrated host-microbe plasma metagenomics for sepsis diagnosis in a prospective cohort of critically ill adults, *Nature Microbiology* (2022). [DOI: 10.1038/s41564-022-01237-2](#)

Provided by Chan Zuckerberg Biohub

Citation: Machine learning enables an 'almost perfect' diagnosis of an elusive global killer (2022, October 24) retrieved 25 December 2022 from <https://medicalxpress.com/news/2022-10-machine-enables-diagnosis-elusive-global.html>

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.