

Strengthening Public Health with Data: Empowering Epidemiologists with Data Science

[Case Studies Data Science](#)[See All News](#)

Epidemiologists swim in data every day, but the value of that work shines brightest during times of crisis. The COVID-19 pandemic spotlighted the role of data in protecting public health — and what is possible with more resources dedicated to data modernization.

In some communities, state and city officials, school districts, and employers could access near real-time dashboards to see how the virus was spreading and make informed decisions to protect communities. In many others, health departments, government leaders, and other decision makers were stymied by outdated technologies, inefficient systems, and processes for collecting data that didn't scale well to the enormity of the pandemic. It opened many eyes to the importance of modernized data systems and the possibilities of using data to make a real difference in public health outcomes for countless diseases and conditions.

That's why most public health departments know it's time to boost their data science capabilities. Using data science, epidemiologists can evaluate and translate qualitative public health issues into quantitative problems, analyze them using machine learning models, and take action. Through this data-driven approach, they create solutions that improve individual and community outcomes, enhance operational efficiency, and advance health research.

These are laudable goals, but how do you set up your data science program to achieve them? For the data scientists at Ruvos, getting started always begins with the first step in the scientific process: asking a question.

Getting Started: Define the Problem

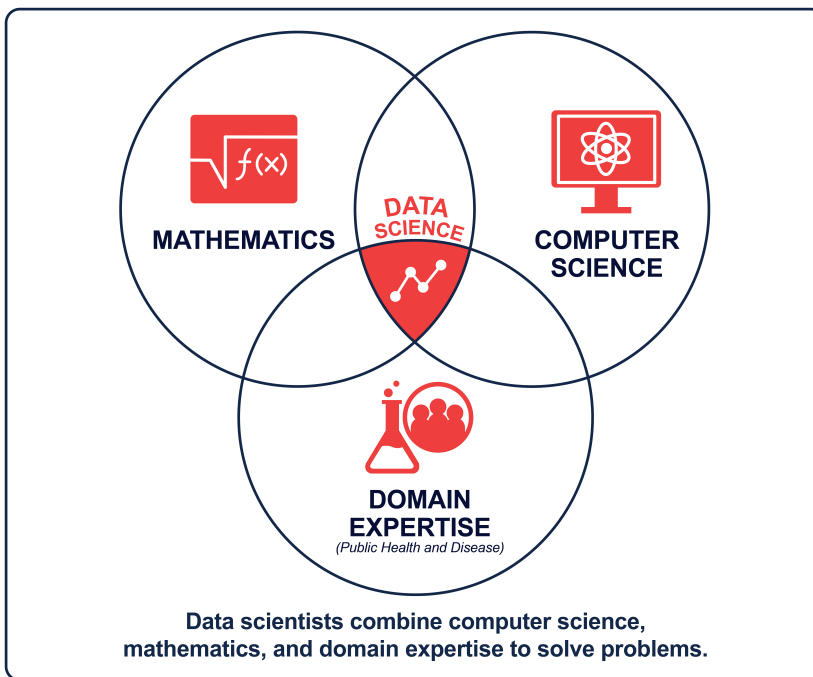
One of the most common misperceptions about data science is that you need to start with the right technology tools. However, a cutting-edge artificial intelligence (AI) model or sleek dashboard is only useful if it produces the actionable insights you need to solve your specific challenges.

A good data scientist starts by defining the problem that needs to be solved. Ruvos' data scientists did just that when they launched an initiative to study PASC/long COVID. They wanted to answer the question, "How do we predict a patient's risk of developing long COVID?" With that question in mind, they determined how to quantify the risk and then built a model to do just that. Only after developing answerable questions can you determine what data you need to test your hypothesis — and what technology and tools you need to collect, analyze, and interpret that data.

Creating a Well-Oiled Data Machine

Collecting and analyzing data can feel overwhelming when you start thinking about the common issues surrounding public health data. Too often, data is incomplete, outdated, siloed, or reported in inconsistent formats, making it difficult to draw valid conclusions and take action. The problems are so pervasive that some epidemiologists report spending 80% of their time wrangling data before they can begin analyzing it.

Luckily, today's technology tools can drastically reduce the time epidemiologists spend collecting and cleaning up data. It's possible to automate most of the mundane tasks associated with data collection and integration, freeing up time to focus on the jobs only humans can do — deriving actionable insights and developing strategies to improve public health. That's what one state discovered when it set out to clean up its COVID-19 data.



Data Science in Action: How a State Department of Health Improved Lab Data Accuracy

During the pandemic, public health officials in a US state began noticing irregularities in the COVID-19 labs they were receiving from testing labs. The number of test results they received from each lab on a given day could vary between 10,000 and 100,000.

Naturally, that vast range raised questions about the accuracy of the data. Were labs resubmitting the same data, making case numbers appear too high? Were test results getting “clogged” somewhere, and the outbreak was actually more severe than it seemed?

The agency needed to get to the bottom of their data issue to then identify where outbreaks were actually happening and develop mitigation strategies.

Combing through hundreds of thousands of lab reports daily to identify irregularities wasn’t an option for the state’s overextended epidemiologists. Instead, they turned to Ruvos to create an algorithm that would detect anomalies in the data flows. If a lab sent more or fewer results than were expected, the system flagged its data for closer inspection.

By automating this simple — but time-consuming — task, epidemiologists had greater confidence in the data and could focus on determining where and how to allocate resources. Improving the quality

of this data pipeline is the first step in unlocking the power of data science to aid epidemiologists and state health officials.

Ruvos is now building automated analytical models and dashboards to enable the state's Department of Health to forecast disease spread. With thoughtful leadership and continued investment in modernizing data systems, building analytic tools, and training public health practitioners to use and interpret data models, Ruvos envisions a future where predicting outbreaks and disease spread becomes as reliable as predicting the weather — so that pandemics like COVID-19 can be prevented early.

Embarking on Your Data Science Journey

While most public health agencies embrace the potential of data science, funding and talent shortages have hindered many organizations from building their capabilities. Limited technology curriculum in traditional epidemiology programs and scarce computer science graduates interested in public health have led to a weak talent pipeline. Lacking staff with the right skill sets, organizations face challenges in evaluating and implementing the tools needed to break through data roadblocks, especially with today's rapid pace of technological change.

The right partner can make all the difference. The Ruvos data science team brings together computer science, mathematics, and epidemiology experts to partner with public health agencies in building tools and capacities to get the most from their data. We work with public health agencies nationwide to understand and solve their data challenges, focusing on making data actionable and finding efficient, cost-effective solutions.

Ready to embark on your journey? To discuss how your organization can harness data science to drive actions that enhance public health outcomes, contact [Ruvos](#) today.