Bob Barrett: This is a podcast from *Clinical Chemistry*, sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I am Bob Barrett.

Big data now seems to be everywhere. Google search queries are used to predict flu outbreaks and airline statistics or mined to determine the best hour of which day to book plane tickets.

Clinical laboratories can also benefit from these trends. Given the emerging importance of reproducible data analysis pipelines, one could think of a special group of trained big data experts that could be available to regulatory agencies and to clinical laboratories for review and auditing of their processes and practices, and who may be able to offer suggestions for improvements.

That's the topic of an article in the February 2015 issue of *Clinical Chemistry* titled "Learning from Our Mistakes: The Future of Validating Complex Diagnostics." The lead author of that opinion piece is Dr. Stephen Master. He is the Chief of Clinical Chemistry and Director of the Central Laboratory at the Weill Cornell Medical Center in New York City, where he also served as a member of their Institute for Precision Medicine. He is our guest in today's podcast.

Dr. Master, the latest buzzword seems to be big data. What exactly is big data?

Dr. Stephen Master: Well, big data is a term that really reflects a fairly recent and rapid increase in the pure amount of information that's being collected, and specifically that's being collected in a way that can be easily analyzed using computational tools.

Maybe the most well-known example of big data comes from companies like Google or Amazon or Netflix, who can use the data that they gather to predict what you want to see or buy. So in the case of Amazon for example, when you see the books that are recommended for you that list has been

produced by an algorithm based on what Amazon knows about you and your shopping habits, as well as on the aggregated shopping habits of a much larger customer base.

Or in the case of Google the information that's available from your search history or your Gmail or whatever else Google knows about you can be computationally analyzed using machine learning approaches and used to predict what ads you should be shown.

So at the commercial level big data approaches have been extremely powerful. So if big data can be used to predict something like consumer behavior, then it's certainly exciting to try and apply the same techniques to predict the outcome of disease and to provide better patient care.

I think it's probably worth highlighting at least two ways in which this is happening. The first is that, as electronic medical records improve, a number of groups are trying to use big data strategies to learn about statistical relationships that will help to guide medical management of patients. Of course, the data that are generated from clinical lab are in many ways perfect for this application since they already exist often in a numeric form that can be easily adapted for the kinds of data mining purposes we care about. So that's one fairly obvious and potentially powerful way to apply big data in medical care.

The second way is also interesting, though, because it shows that the big data that is useful to predict disease can sometimes comes from unexpected places. In 2009, a group from Google published a paper in *Nature* in which they show that they could predict influenza outbreaks in the US by tracking search terms that Google gathered.

At first glance this doesn't seem like medical data, but it turned out that this project which was called Google Flu Trends, was able to provide accurate results one to two weeks earlier than the CDC. So it held out the prospect of significant public health benefit.

Bob Barrett: What do you think that initiatives like the Google Flu Trends project can teach us about testing in the clinical lab?

Dr. Stephen Master: Well, on the one hand as I've indicated, Google Flu Trends was a great example of using big data in healthcare. Google's process, which included mining 50 million search terms in an unguided way and coming up with a predictive diagnostic, is a great example of how machine learning can work. And those of us who are in the clinical lab who also increasingly managed large data sets coming from both the fact that we see a large numbers of patients and also from the fact that we measure more and more things about those

patients, you can just think for example of the marked increase in data that have come from the widespread adoption of genomic testing for example, so when you think about the analogy between say Google and the clinical lab, you can see why there is a lot of excitement about using similar algorithms to analyze lab data.

Now I should say though that there is a catch. The flip side of the story is that although Google Flu Trends did a great job in the 2007-2008 flu seasons, by 2012 it was not performing as well. It turns out that changes in the usage of terms by Google users among other things were leading to poorer performance.

So when Google went back and incorporated data from 2010 and 2011 into their model, their 2012 predictions improved. Okay, so what does this tell us about the clinical lab? We face an analogous situation, as you know. For example, when treatments change or patient populations change, it may be that we will see changes in the performance of complex algorithms that we may develop for diagnostics.

So in this article we argue that it is very important not to lock ourselves into static models, but rather to build in statistical approaches, such as Bayesian Statistical Approaches for example, that can be dynamically refined over time, and this is what we mean by learning from our mistakes.

Bob Barrett: The March 2014 issue of *Science* had an article that stated that Google was guilty of big data hubris. What do you think of that?

Dr. Stephen Master: Well, as I said, there are a number of things that have been noted about Google Flu Trends subsequent to its original publication in 2009. Some of that had to do with the problems that arose from not incorporating the 2010, 2011 search term changes, and so certainly one could take this too far and say that Google has decided that they can predict the entire world simply based on search terms.

But nonetheless, it is the case that Google Flu Trends does a remarkable job at predicting the flu, given the caveats about retraining the algorithm. And that's really one of the things that was an impetus for us writing this article, was to say that in the same way we don't want to be caught in the clinical laboratory community with static algorithms that do not reflect changes that occur in our underlying patient populations or in their underlying treatment.

Bob Barrett: What are some of the risks involved in using complex algorithms to interpret laboratory data?

Dr. Stephen Master: Well, the main difficulty with using complex algorithms is that they can hide many other mistakes in medical research, whether those are errors in study design, or problems with analytical variation, or issues with the data management itself. There have been a number of examples in the last decade or so where complex biological data have been interpreted using an algorithm in a way that is just frankly incorrect, we now know. And there is at least one publicized case in which this led to patients being placed in the inappropriate arms of a clinical oncology trial. So clearly this is something that needs to be very carefully considered. So how can we draft this?

One solution has been to try and understand the underlying biology that drives the algorithm. Another strategy has been to make sure that there is appropriate regulatory oversight, and this is certainly something that the FDA, for example, has been very aware of.

Now third response came from what was really a landmark report released in 2012 by the Institute of Medicine, that recommended that these kinds of complex algorithms, specifically when applied to -omics assay such as genomics, proteomics or metabolomics, should be completely specified in a final form prior to rigorous clinical testing, in order to ensure that their behavior is well understood.

Bob Barrett: And how important is it for us to understand the biological basis of these algorithms?

Dr. Stephen Master: Well, that's a very interesting question. It would be tempting to say that the algorithms we develop using big data approaches should always have a clear biological basis because this provides an independent rationale for their use, and certainly anything that's discovered in the process of mining big data should absolutely serve as an impetus to drive research. If we find something that's reproducible we want to understand it, and that's all as it should be.

However, one of the points that we make in the paper is that it isn't always good to wait for this level of scientific understanding before we start making use of the results.

One good historical example of this is a Hungarian physician in the mid-19th century named Ignaz Semmelweis who showed a correlation between hand washing and a decreased risk of puerperal fever, and of course today this is still a mainstay of hospital infection control.

He knew that the correlation existed, but he didn't know the mechanism, and this led to delays in the implementation. And with the benefit of hindsight I think we would all now say that it's clearly not a good idea to wait for that full

scientific understanding if you have a clear statistical association.

We would argue that the same is true for the results of big data in clinical diagnostics. Although it should be said following that cautionary tail of the last 15 years, it's clearly critical that good study design, data collection practices, and bioinformatics, are used to ensure the reliability of the results.

Bob Barrett: Now finally doctor, you mentioned earlier that Google Flu Trends demonstrated the value of moving beyond static models; would this approach have any implications for the way that diagnostic tests are regulated?

Dr. Stephen Master: Yes, I am very glad you brought that up. In this paper we argue that one lesson of Google Flu Trends is that you can't just mine big data, create a great algorithm, and then stop. As treatment paradigms or patient populations change, it will be necessary to update our models based on new data that comes in.

Now the way that we as a field currently handle this from a regulatory perspective is to say, you've created an assay with a static diagnostic algorithm, we've tested that static algorithm, and if things stop working down the road, then we will start the process all over again.

What we are proposing is an alternative where we say that it's possible to learn from our mistakes using a more dynamic approach while still being rigorous using tools such as Bayesian Statistics, to continually update our algorithms based on new data that we receive.

But, of course, this raises the critical question, if laboratories start building dynamic analyses into complex data, do that into their diagnostics, how can a consumer of data or regulatory agency know that the results are reliable and clinically valid? And this is obviously very important since we do not want to propose any changes that would put the field at the risk of some of the problems in the last 10-15 years. We do think that there is an answer though, which is to rigorously validate the dynamic updating process itself.

So rather than saying that we approve a fixed algorithm with fixed parameters, we might envision a regulatory approval of a well-defined data gathering process that is used to alter the algorithm in a specified statistically valid way. In fact, my co-author has proposed that there would be a role for specially trained big data experts that we could call Algorithmists, who would be able to review and audit and validate these types of approaches for the laboratory and for regulatory agencies.

So we do think that there are ways to more efficiently have our diagnostic algorithm learn from their mistakes in a way that maintains the safety and validity of laboratory testing.

Bob Barrett: Dr. Steven Master is Chief of Clinical Chemistry and Director of the Central Laboratory at the Weill Cornell Medical Center in New York City, and a member of their Institute for Precision Medicine.

He's been our guest in this podcast from *Clinical Chemistry* on the potential use of big data for assay validation.

I am Bob Barrett. Thanks for listening!