



Special Podcast:

N. Tolan *et al.*

Q&A: "Big Data" in Laboratory Medicine.

Clin Chem 2015;61:1433-1440.

<http://www.clinchem.org/content/61/12/1433.extract>

Guests: Nicole Tolan, Harvard Medical School; Dan Holmes, University of British Columbia; Laura Parnas, Livermore, California, and Rajeev Kumar, Stanford.

Bob Barrett:

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

Big data. Depending upon who you ask, you may get different answers to the question of, what does big data mean to you? Big data is classically defined as datasets so large or complex that traditional data processing applications are inadequate.

In the December 2015 issue of *Clinical Chemistry* in a Q&A feature, eight experts, with different backgrounds in laboratory medicine and patient care, describe what they consider to be big data.

These experts detailed a number of diverse clinical environments for big data and review various IT solutions to improve efficiencies in their analysis.

Each highlight the need to ensure high quality data goes in so that the conclusions translate to effective clinical management.

In today's podcast we are joined by four participants who provide their own insights to big data and how to cope and how it can help.

And we will start with one of the moderators, Dr. Nicole Tolan; she is the Associate Director of Clinical Chemistry and Medical Director of Point-Of-Care-Testing at Beth Israel Deaconess Medical Center and an instructor in pathology at Harvard Medical Center.

So Dr. Tolan, big data is often considered to be those data obtained through high resolution -omic studies but this *Clinical Chemistry* article focuses on how laboratory professionals and clinical care providers are actually dealing with big data on a day-to-day basis. Now, can you elaborate on this?

Dr. Nicole Tolan: Sure, I think big data really goes beyond high-resolution -omic data and in medicine it generally refers to all of the information captured in the Electronic Health Record. So more specifically this includes all of the data generated in clinical laboratory testing.

Recently, there has been efforts to also now include real-time data that's obtained from all of our personal electronic devices and other personal use meters, like glucose meters for instance, in an effort to really leverage this big data towards achieving optimal health beyond a single measurement obtained during the doctor's visit.

So specifically in laboratory medicine we can extrapolate from this big data to routinely perform what Steve Master fondly refers to as medium data analytics.

This clinical data can be assessed in terms of assay performance. For instance, Dan Holmes has built a number of very beautiful automated analytics for continuous quality assessment and improvement and we can monitor select patient populations in an effort to not only improve laboratory practice but overall clinical care, and of course we can identify areas where physician education may be necessary in terms of improving test utilization.

I think clinical chemists are really vital in these processes to ensure that data analysis is accurate and it's not fraught with other systematic errors or biases. We have all of this clinical data at our fingertips, and with it the clinical knowledge to really properly analyze this data.

In the end, if we put garbage into our sophisticated data analytics software we're still going to get garbage out. This starts right from the beginning when we query the medical record and obtain this "medium data" amongst all of the other big data that's captured.

Bob Barrett: How did you arrive at this topic for discussion?

Dr. Nicole Tolan: I think over time there's been a growing international group of clinical chemists focused on effective information management.

In particular, at the 2015 SYCL Workshop at the last AACC Annual Meeting, we organized the series of informational talks and interactive sessions, really focusing on information technology in laboratory medicine.

And this was really well attended and from it we realized the great interest of clinical chemists to learn more about these information technology tools for clinical data analysis.

We are really working to ramp up the dialogues amongst the clinical chemists to highlight that these strategies are all within our reach.

We have the opportunity to improve quality assessment methods. For instance, to gain efficiencies in the clinical lab by leveraging this data towards effective middleware rules and allow us to focus on the exceptions that need intervention to really improve laboratory and clinical practice.

Bob Barrett: Dr. Tolan, how do you deal with big data in your own practice, and what have you taken from these experts to put in place in your day-to-day responsibilities?

Dr. Nicole Tolan: So I had the good fortune to learn from a number of these experts along the way and we've incorporated various tools and big data analytics strategies to mainly assess quality and improve efficiencies.

For example, we've recently analyzed how our point-of-care glucose meters are performing in the ICU. With the help of a medical resident, Dr. Dan Schultze, we have automated the assessment of the discrepancies between point-of-care lab and lab glucose results but we're looking at it from the clinical context rather than the arbitrary accuracy cut-off of say 10% that the FDA is proposing.

So working within R, which is the statistical programming language, we've matched glucose meter results to laboratory results that have been collected within 10 minutes of each other. But we're also minimizing the distance in time between these data pairs.

These points were then plotted by both the Clarke and Parke Error Grids in order to assess the clinical impact of these discrepancies. And we've automated this by meter serial number for ongoing assessment of meter performance.

Now, while Dan has a background in computer programming and his knowledge of R made this very efficient, these analyses are something that other laboratories can easily perform.

In fact these consensus plots or error grids are available on the package within the open-source, the Comprehensive R Network or CRAN and it's listed under EGA or Error Grid Analysis.

Bob Barrett: Thanks Dr. Tolan! And we will get back to you in a bit but since you were talking about Dr. Holmes, let's turn to Dr. Dan Holmes; he is Division Head of Clinical Chemistry in the Department of Pathology and Laboratory Medicine at St.

Paul's Hospital and University of British Columbia in Vancouver.

Doctor, in this Q&A Article, you've outlined that big data is somewhat jokingly the amount of data a laboratorian may want to process that they cannot do using standard programs such as Excel.

Now there is a lot of truth to this, so can you detail for us how an interested laboratorian would go about transitioning into this method of big data analysis?

Dr. Dan Holmes: Well, if somebody wants to get into big data analysis, there are a number of tools available for them on the market. The Business Intelligence Industry has a well-developed set of tools that have come about, that are driven obviously by profits and these are spilling over into lab medicine.

An example tool might be something like Tableau, and Tableau is a drag-and-drop programming tool. So you don't actually write any code, you drag widgets into their place and then reports are generated or your dashboards are generated based on the decisions you've made in the drag-and-drop process. There is programming underneath there but it's hidden from you.

There are other tools like Tableau, I won't name them, but if you go and look for these tools online you will find them. For me personally, this was a little too confining.

I wanted to use something where I completely control what's being displayed. So I used the R Programming Language, which is an open-source language and it's free unlike Tableau which is a commercial product.

Those who know me well know that I'm kind of an R evangelist, if you like, that is I promote the use of R in lab medicine. R allows you to manipulate the data anyway you like, clean-up your data and control how it's displayed and the nice thing is its open-source and it's free.

R isn't the only programming tool you could use. There are other approaches. You could use Python, or something called D3.js but I decided to focus on one tool that I've kind become accustomed to.

So when it comes to learning R, the place to start, the book of course, and then once you get into it you'll discover that there are many problems that you've been thinking about how to solve and suddenly you have a way to solve them.

And I think the root of solving problems is the way to learn. You pick a small problem, you work on it, you solve it, and

then you move to larger problems and move to creating tools that are more automated.

Bob Barrett: Well doctor, you've talked about the software tools and the resources available to get started. How does one gain access to their clinical laboratory data?

Dr. Dan Holmes: It's a surprise to me that this is a problem everywhere I go. I talk to people and they can't gain access to the raw data in their lab information system very easily. That's because the lab information system data is not an open database so we can freely make queries to pull up massive amounts of data at once.

We are quite fortunate in our situation that we have a mirror database specifically for making queries. So we are on Sunquest for our lab information system. That's a hierarchical database and there's a mirror database which is a relational database that mirrors all of the lab information system data.

Our programmers have made a web interface to that mirror database and so we can query that mirror database whenever we like. We're very fortunate in that regard.

If you don't have access to some kind of mirror database that you have the freedom to query then you have to formulate a relationship with someone lab information system IT to help you formulate queries, make sure you're getting the data that you want, and pull it out in a way that your program, whatever program you're using, happens to be able to digest it.

Bob Barrett: So doctor what are some examples of how the R Statistical Programming Language can be used for more efficient even automated day-to-day responsibilities like quality assurance measures. Talk about what you have done.

Dr. Dan Holmes: We've used R in number of different ways. If you can dream it you can do it. It's a programming language and that's what Steven Master and I say in the course that we teach about R that it's a -if you can dream it, you can do it. You can do just about anything you like.

Here are some examples of some of the things that we've done. We wrote QC report generator so every Monday morning the QC database is tapped for the QC results for the whole -- of the last say either one month or three month depending on frequently the test is done.

The data is processed and generated into Levey Jennings plot, unique to each instrument. Those Levey Jennings plots are put into a PDF with which QC level is being evaluated

and what the target range is, what the expected and the observed ranges are, and then that's automatically e-mailed to all the concerned parties early in the morning on Monday.

So when I arrive in to work there are six e-mails in my inbox, one for each area of the clinical chemistry lab and I can buzz through the QC results quickly by looking at these PDFs and that's an expandable service we could use. We could do that for other hospitals that are using the same lab information system as us.

We have also made automated quality reports. We generate KPI reports -- that is Key Performance Indicator Reports for turnaround times on potassium, on hemoglobin, on troponin. Those reports are tracking things like turnaround time as a function of time of day, day to week, tracking the pre-analytical and an analytical process.

Those reports can be auto generated using a pipeline that I like very much. The language ultimately at the bottom of all this is R, and we use a markup language called R Markdown. R Markdown uses a tool called Knitr to produce a PDF that shows all of the things that I had generated in my R code in a single report.

So in that way I can generate a turnaround time report for my hospital or for other hospitals. It takes about two minutes to run, and each month I feed it new data and it can make statistical inferences and do all the tasks that I have programmed in. So it's a nice way of producing something that gives me useful data but doesn't require the tedium of going into Excel.

The other nice thing about an automated quality report like that written in R, is that I know exactly what I did and if someone else wants to see what I did to generate this report, they can look at what data was excluded based on what the code shows.

When you do something in a graphical user interface on the other hand, you don't really track the mouse moves of the person, you don't know exactly what they did and all of us who had the experience of finishing an analysis in a graphical user interface program like Excel, realizing we have made a mistake and having to go back to the beginning with our data, when you program it re-running from the beginning, is a matter of clicking.

So it's more work up front but ultimately it's more transparent, more repeatable and if you need to make some changes those changes are quite easy.

Another thing we do is we can in that report is we can do low tracking, we can see where the specimens are coming from, we can see when there are spikes in the number of testing ordered from the emerger elsewhere.

Finally another example of something we have done is we have written middleware for our mass spectrometer. It takes the raw data file off the mass spectrometer and prepares it for uploading to the lab information system.

So that is kind of a rudimentary interfacing if you like. It's not a true interfacing. It's what we call a flat file interfacing but R is very suited to do that kind of flat file processing and it's quite easy.

Bob Barrett: We know that machine learning tools and novel quality tools have come out of the big data arena, well what barriers exist for having these tools operate in real time in the lab?

Dr. Dan Holmes: I mean that's a really interesting question and as the few colleagues of mine have written software to operate machine learning algorithms on routine lab data, Steven Master has done some of that and Chris McCudden and Matt Henderson at the University of Ottawa have also done something similar.

The challenge, if you want to implement something like this in real time, is that you would have to pipe the data off the instrument into your homebrew middleware and then pipe it back to the lab information system.

This is something that's likely to be met with the eye of incredulity by the IT folks unless you have a very collaborative and research focused group in IT.

Now that may be the case that your institution. I don't think it would be met with enthusiasm at a lot of institutions. The other approach would be to do it in a quasi real-time approach where you're not piping the data into your Blackbox and then piping it back to lab information system.

You are letting your results file, and then you're applying your machine learning algorithm to the file results and then appending a comment or interpretation as appropriate.

Obviously that's not a true middleware solution, but it would be a valid thing to do if you had a working tool that was clinically validated and you want to implement it in the most painless way possible, if you like.

Bob Barrett: Finally can you talk a little more about how these data science tools create a paradigm for reproducible report generation?

Dr. Dan Holmes: Sure. The data science tools that are programmatic are fundamentally different than the graphical user interface tool. In a graphical user interface tool, you make mouse moves, you cut columns, you remove data, chop it, etcetera, to produce your final data set, massage it however you like, remove missing values or do something with them, remove less than and greater than signs.

Do what you need to do and finally generate an image or statistical analysis. But there is no tracking of exactly how you processed the data and how you created your final report or images.

The difference in using a programming language like R, or Python, what have you, is that what you have done is clearly laid out in your code.

If you decided that you were going to remove all values that were less than the lower limit of detection, there will be a line in your code that shows that you have done that.

Now if someone wants to come to you and say, how did you do this analysis? You can go back and say, well here is the code by which this analysis was done, you can walk them through what that means, you can tell them how it accomplished, what you have accomplished.

When you do it that way, you have a totally transparent report and if you want to make changes to it, it's very easy. Once the code is written, you can execute it again very quickly after the necessary changes have been introduced and those changes themselves are going to be trackable.

The really nice thing to do is that you can actually create versioning for your document or your code and every time you make a change you can track that change with versioning software and you can see what you have done all the way along as your report generator code becomes more sophisticated over time.

Bob Barrett: Thanks so much Dr. Holmes. Now another moderator of the Q&A session in *Clinical Chemistry* was Laura Parnas. She is the Technical and Scientific Director for Clinical Chemistry at Sutter Health Shared Laboratory in Livermore, California.

So Dr. Parnas, the digital health revolution is amongst us with patients increasingly involved in monitoring their physiological functions and continuously gathering information about their bodies, and healthcare teams working to leverage patient engagement with the potential for managing health and chronic diseases in real time.

As large amounts of data are generated by continuous monitoring devices, what types of tools were available to make the connection between patients and healthcare providers?

Laura Parnas: Well most of the biggest technology companies including Google, Apple, Intel, etcetera, and some other smaller, lesser known startups have demonstrated interest and are very focused in the wearable devices or mobile health or what we also know as Digital Health Medicine Market.

The field, right now it is in inception but its growing very, very fast and it's getting a lot of attention.

We as healthcare providers need to be ready for what's in the pipeline as we evolve toward a personalized medicine environment. There is no denial that this is in our near future and we need to be ready for it.

Apple, for example, has designed a framework called HealthKit with a goal to centralize data storage of personal health information allowing apps that provide health and fitness services to share their data with each other and with Apple's health app.

The idea with this is to use the health app as a data repository that can be connected with the electronic health record. The EHR will then receive the selected data and post it to the patient's chart where the physician will be able to evaluate this data together with the rest of the patient's health information and act as needed.

Many large healthcare institutions in the U.S. have started the pilot Proof of Concept projects using Apple's HealthKit to monitor in real time a variety of measurements including blood pressure and laboratory test results.

The ultimate goal of this is to obtain a more complete picture of the patient's health and to identify and act up on changes and conditions as soon as they happen.

Two of the experts in this Q&A have Proof of Concept projects on their way utilizing Apple's HealthKit at their institutions and with us they will share invaluable insights into their experience with the platform as well as how it has impacted their practice of medicine.

Bob Barrett: Thank you Dr. Parnas. That's all we have got for you but we will get back to that Apple HealthKit a bit. We are joined now by Rajiv Kumar. He is Medical Director of Clinical Informatics in the Departments of Pediatrics and Clinical Informatics at Stanford School of Medicine and Stanford Children's Hospital in Palo Alto, California.

So, Dr. Kumar, patients with Type I diabetes control their blood sugar with multiple injections of insulin throughout the day. They use their own data, including current blood sugar measurements and the amount of carbohydrates they anticipate to eat, in formulas determined by their healthcare provider to calculate each dose of insulin. Each dose may be adjusted by variables, including activity.

Now, in general, healthcare providers review retrospective blood sugar data and adjust those dosing formulas accordingly at the quarterly visits, but their patients, particularly growing pediatric patients, may benefit from adjustments more frequently.

So, Dr. Parnas was telling us about Apple's HealthKit. Can you tell us about your experience with Apple's HealthKit to monitor your patients' glucose levels between clinical visits?

Dr. Rajiv B. Kumar: Sure. I just want to highlight a specific point to your question is that diabetes is a big data disease, right? Our patients do a lot of work recognizing the variables that you mentioned, but also trying to identify trends and respond to trends and really own their data and learn from their data. And for our pediatrics patients, that's even more challenging, because their bodies, their physiology are constantly in flux, constantly changing.

Every time I see a patient, they are always bigger than the last time I saw them. Sometimes they are going through puberty, and they are playing sports on different days of the week. They have fevers and colds and going to grandparent's home and all sorts of variables can affect their blood sugar control and how much insulin they need.

So lots and lots of data to review, but in reality we only see our patients once every three months, and at those visits, it's just impossible to look at their three months of data comprehensively. So usually, we look at the most recent two maybe four weeks, and often again, that data is skewed. So we ask our patients in between visits to communicate data to us just so we can stay caught up to date, but that's easier said than done.

It's already such a difficult disease, and then to ask our patients and their parents to do more work and convert their data into a secure method and send it to us electronically or by fax or by written logs, and all these old school methods just really aren't efficient. So most patient don't actually do it and we kind of play catchup every three months.

So then we learned about Apple's HealthKit and we recognized it as an opportunity to passively gather data, and so to gather data in between visits without increasing work for patients and their families, and without increasing work for the healthcare providers who are also stretched pretty thin.

So let me just give a bit of background about HealthKit and continuous glucose monitors and our electronic health record and all that bit. So, I'll start with continuous glucose monitors. These are sensors that a patient wears on their body. They are inserted at home. They just wear them right in their subcutaneous fat, which mean they put it right into their fat and it sticks to their skin and it's an easy procedure to do at home.

And these take a blood sugar estimate every five minutes and send it to a wireless receiver that's nearby the patient, and with Bluetooth's capacity that receiver can send data to a mobile device. So the continuous glucose monitor we use is called a Dexcom G4, and that data goes to an app on an Apple iPhone or an iPod called Shared2.

Once data reaches that phone, it's on the phone. HealthKit is another app on the phone and MyChart is yet another app on the phone. MyChart is our patient-facing portal, so it's a way patients can access their health records.

I think of all these apps as like a train station, so data can get on, data can get off at each stop, but the union station of all these train stations is HealthKit. So if data get on one app and you wanted to get it to another app, so you can go through union station and send it to where you wanted to go.

And you do so securely, because you are going to from one app to another app on the mobile device. You are not going to the Apple servers, you are not going to the cloud, you are literally from one app to another app to another app, and so it's very secure.

So the idea here is that a patient who wears continuous glucose monitor, a blood sugar is assessed every five minutes, it's sent to their phone through HealthKit it gets sent over to the MyChart app, and the MyChart app shares the database with our electronic health record, which means if I am in a patient's chart, I can see their blood sugars.

Completely passive; once we set this up, it's completely passive, so patients are constantly sending data and we have access to it whenever we need.

So it's a really efficient system, and the key aspect here again is that we are getting 288 blood sugars readings a day passively, and we are getting this data in between visits. We will talk a little bit more about how we use that, but a novel way of getting this data.

Bob Barrett: You emphasized that a patient's health data should be unified in one database. Why is that so important?

Dr. Rajiv B. Kumar: Yeah, it's the really important question. So, all the data I have for a patient lives in my electronic health record, except for the one outcome measure I care the most about in Type I diabetes and Type II diabetes, which is the home blood sugar readings, right?

Everything I do, every dose I change, every medicine I prescribe, every bit of information or advice that I give is to gain better control of home blood glucose readings, but I don't have access to that data.

People can send me PDFs, they can text numbers, they can fax numbers, but it doesn't live where all the other variables live.

So when I am trying to use analytics or run reports to compare against age, gender, brand of insulin prescribed, type of insulin prescribing habits, other lab tests that live in the chart, growth spurts, growth chart, puberty data, I can't compare against the one thing that I care about the most.

So that has always been very, very frustrating, as I can't learn from my patients not only individually but also the population.

So as our electronic health records analytic tools are improving, we now have the ability to look within our own patient population and say, who has the best blood sugar control now that we have the blood sugars where they belong, in the electronic health record database, and why, why do they have the best control, what's different about them compared to other patients who don't have great control.

And as these tools evolve, we'd be able to become more interoperable with other health systems and compare data here in Northern California to Southern California and to Chicago and to Florida and to every place, and say who is gaining the best control, who has the best trends, and what are they doing differently and how can we learn from each other.

So right now, we are totally blind. I am not aware of another hospital in the world that has the raw blood sugar

data within the charts, and so we can't assess anything unless we leave our chart, go to third-party software that doesn't have all the other data.

So as we are evolving here I think it's really important that we at least have our data in one place or that all the places we have data are interoperable so we can compare and learn from our trends.

Bob Barrett: What were the major barriers that your team faced and what challenges do you predict for others who are looking to replicate this work?

Dr. Rajiv B. Kumar: So I think the first barrier is getting my colleagues comfortable with this idea, right? So our patients are doing a lot of work as it is, and it's difficult for them to send us data between visits before we had this new method.

And for our colleagues, my colleagues are also stretched quite thin. So when I go to them and say, I can get you an extra 288 data points a day, I am not always greeted with smiles and laughs, and the worry that we are just going to be overwhelmed with data, and so that was a big challenge. So we really needed to define our workflow to make sure that we weren't interrupting the system or overwhelming the system.

And so the idea here is that raw data is coming into electronic health record, but we don't look at the raw data. We build analytic tools in the background to recognize who is having trends that are actionable, who is having too many low blood sugars overnight, who is having too many low blood sugars overall, who is having too many high blood sugars et cetera.

And then we look at our population as a whole, then we can run a report and say, of all of your patients, these are the ones you really need to contact today or these are the ones that really have an actionable trend, right?

So by doing this with everyone, let's say, we had everyone signed up for this system, we don't necessarily have to look at everyone every day or all the time. The electronic health record sorts our patients out for us and tell us who we need to be worried about.

Then when we find patients we need to interact with and learn the trends and give advice to improve those trends, we don't want to log into their chart and see 288 blood sugars readings a day. It's just too much.

So we built a viewer, a method of looking into all this data in a way that very intuitive. So literally, within our chart, we

click a button, and that button turns all their blood sugars all the dates all the times into a figure called Modal Day, which is intuitive to physicians who take care of diabetes or healthcare providers who take care of diabetes, and shows their trends by hour of the day and allows you to manipulate what you call low blood sugar, what you call high blood sugar, what you call day, what you call night, and identifies trends really quickly.

And then once you see the trend, you call the patient up or you message them through the patient portal and say, hey I see this trend, what do you think about it.

That's a very fun part of it, because not only are we able to give advice and interact more frequently, we are able to assess really what's going on, because the patient knows exactly what causes their trends usually. They may not put it together and recognize the pattern, but they are able to identify variables that I couldn't have otherwise guessed, and they become part of the team and it empowers them and it empowers us, so it's a really good method.

I think another concern from my coworkers is patient dissatisfaction. So people have tried home blood sugar monitoring in the past and published articles, and have pretty good outcome results, but also noting that there were patients that were just irritated that they weren't contacted in real time when they had a low blood sugar or high blood sugar.

So we make it very clear to our patients that we just want to make their expectations appropriate. At this point, we are not able to watch all blood sugars in real-time. That's certainly our goal, but we are not there yet.

So we say, if you have something you are worried about, you have a trend, you're worried about, use your traditional methods. If it's something urgent, page us; if it's not urgent, send us a message, but now we can access all the data right then. You don't have to send us anything else. Just tell us your concerns and we will touch base and figure out what's going on.

But otherwise every patient that uses the system that had low blood sugars, had had high blood sugars and certainly we haven't called them with every low and high blood sugar, and no one has voiced any dissatisfaction.

In fact, patients are very happy with this overall so far. In fact, some patients forget they are even sending us data and they are really happy when we send them some suggestions. So that's been something that's worried my

colleagues but something that's proven to be not a major issue.

I think as healthcare is catching up with technology, right, as we advance in technology, I think there are other changes that need to happen.

We need to have more formal policy that can alleviate some concerns from some providers. Some are nervous about having access to all this data. Although it hasn't increased our workload terribly or at all, the goal is to be able to watch blood sugars in real time or to be able to more interactive with our patients.

I think we need to evolve as our healthcare model and including our billing models, right? It may be more important that we don't see patients every three months and bill for it at that time. Maybe we don't need to make patients drive to our clinic so frequently, but we can have more touch points in between visits.

But if we don't have billing policies where it all levels out at the end and we can't maintain the same staff that we have that are necessary, like diabetes educators and nurse practitioners and nutritionists and social workers and so I think policy building models, insurance models, the way we practice healthcare needs to adopt to our evolving technology and ability to assess patient generated health data.

Bob Barrett: Well, finally Dr. Kumar and not to sound too over the top of grandiose, but is this going revolutionize healthcare as we know it today?

Dr. Rajiv B. Kumar: Excellent question. So, you know, President Obama has pushed healthcare to escape this one-size-fits-all approach, and feels like healthcare really pushes forward to these buzz words we go around like precision medicine, enhanced population health, et cetera. Healthcare actually allows stuff like that to happen. Now I can passively get patient generated health data and it is not just data from the patients that have the fancy computers or have the best understanding that how to plug all these things in. It's literally just passively absorbed through a mobile device and sent to us and even our patients who have the lowest socioeconomic status are very likely to have a mobile device. That's the most common way they access the Internet rather than from a computer.

And so I think using the mobile device and using a platform that's passive on a mobile device that will revolutionize healthcare.

There is obviously more work to do getting all the databases to be interoperable, to look at the data and concert as a whole, both for populations and for nations and internationally there's a ton that can be learned literally within seconds if we had more data in the system.

So there is more work to happen before we revolutionize healthcare but we are certainly on that path with the tools that are evolving and it is going to be really important for chronic diseases like diabetes but others like obesity and asthma and cancers and heart disease and pretty much any chronic disease.

Bob Barrett: Okay. And thank you Dr. Kumar. Let's go back to the Nicole Tolan and give her the last word. Dr. Tolan, what are some of the challenges that you foresee, that will inhibit further improvements in clinical data analytics?

Dr. Nicole Tolan: Well, I think anything that really inhibits our ability to capture accurate clinical and laboratory data. Right now without a comprehensive global electronic health record, we're really missing some information that's recorded in systems that aren't integrated.

Going forward, not only do we need to stop manufacturers say from deliberately blocking information sharing, we need to really invest in a system or a number of systems that are harmonized if not standardized in order to really capture this data both in the clinical sense as well as in the laboratory and, well knowing that there is more and more data captured in the electronic health record each day.

Bob Barrett: That was Nicole Tolan, from the Harvard Medical School. She was joined by Dan Holmes from the University of British Columbia, Laura Parnas from Livermore, California, and Rajeev Kumar from Stanford in Palo Alto, California. They have been our guests in this big podcast on Big Data from *Clinical Chemistry*.

I am Bob Barrett, thanks for listening!