YALE CANCER ANSWERS

Cancer Research: Fact vs. Fiction

Hosted by: Howard Hochster, MD

Guest: Shuangge Steven Ma, PhD

September 17, 2017

Welcome to Yale Cancer Answers with doctors Howard Hochster, Anees Chagpar and Steven Gore. I am Bruce Barber. Yale Cancer Answers is our way of providing you with the most up-to-date information on cancer care by welcoming oncologists and specialists who are on the forefront of the battle to fight cancer. This week it's a conversation about biostatistics in cancer with Dr. Steven Ma. Dr. Ma is a Professor of Biostatistics at the Yale School of Public Health and Director of the Yale Cancer Center Biostatistics Shared Resource. Dr. Hochster is a Professor of Medicine and Medical Oncology and Associate Director for Clinical Sciences at Yale Cancer Center.

Hochster	I personally have always regarded statistics as kind of one of the things that has made clinical medical research possible. When there is just so much variation in how things come out, we need statisticians to help us decide if it is for real or just by chance. So, can you start out by telling us a little bit about what the role is of statistics in biomedical research today.
Ma	I think I completely agree with you, especially for complex diseasse like cancer and maybe cardiovascular diseases, mental disorders, there are just way too many variations we need to somehow control and we need statistics to guide us to really understand what is going on under those diseases and what is kind of the best of prediction models – so, if we have a patient coming in, what we can tell the patient and what is going to happen next and also to set out the best treatment strategy. So, pretty much in all stages of research and treatment of those complex diseases including cancer, we need statisticians to really tell us what is going on and what our strategy should be.
Hochster	And so, we go out with the cancer center biostatistics shared resource to describe to develop clinical research and understand how to do the clinical trials so we get the right answer. And for example, if I give somebody a treatment, they get better and so I think it is from the treatment, but that is not always the case, they might have gotten better anyway, right? So, we need to design these trials with the right statistics to have confidence that what we observe is actually due to the treatment and not just random chance.
Ma	Right. So, while you mentioned the clinical trials, that is definitely one area where statistics and statisticians play a very critical role. As you mentioned, we actually see a lot of variations in clinical and observational status, so we really need statistics to identify what is the truth and what just happened by chance or was going to happen anyway, so in statistics, we call those variables as confounders, so basically what we do as we need hardcore statistics to really separate those confounding facts from the treatment effects that we are actually interested in. So, that is actually one good example where statistics play an important role in cancer and also other disease research.

00:03:58 into MP3: https://ysm-websites-live-prod.azureedge.net/cancer/2017-YCA-0917-Podcast-Ma_315286_5_v1.mp3

Hochster   So, when I want to test a new drug for the treatment of colon cancer for example, which is what I do, I go to my statistician, can you tell us like what the role of a statistician in developing and designing the clinical trials?

Ma   Basically, we are, I mean statisticians, are involved in pretty much every single step of clinical research, clinical trials. So, the first thing we do is we sit down with clinicians, like you, to really design the study. So, we need to identify the right population we want to target and very important task is to decide how many samples we need, so how many patients we want to recruit and when it is feasible to recruit based on for example our hospital, studies and also we need to work on the study recruitment plan, like how long I want to keep the study recruitment and what kind of strategies we need. So, that is kind of the first step in the study design. And also, when the study actually starts, we also need to be involved, we need to make sure all data are collected in an unbiased way so basically, there will be no human selection bias and also we do a lot of interim analysis, so pretty much we keep an eye on the data, on the study, execution at every single minute. So, we need to make sure there is no harm to the patients and one we actual trim the study either with design, and finally when the study ends, we need to conduct unbiased analysis and rigorously justify if the treatment has any beneficial treatment effect.

Hochster   So, as you were saying, the first thing might be to decide on what the study is – is it just the treatment alone, is it the treatment compared to a control arm or whatever, and that will help us to determine the size, so the size could be 30 and it could be 3000 depending on what you are looking for. And then, you talked about – you look at the data and what we call interim analysis to make sure that it is safe to continue the study. Can you tell us a little bit about how that works because sometimes we hear our study was stopped early because it was not positive, or sometimes the study is so positive that we do not need to finish it.

Ma   Right. So, before we start a trial, we have a hypothesis and usually hypothesis is the new trial is going to do no harm and actually going to be better than existing treatment. Unfortunately, that hypothesis is not always true. So, what kind of makes clinical trials interesting and also very complicated and very challenging is the subjects are actually human being, while the patients they are like us and they are humans, so we want it different from animal studies. So, if we find somewhere in the middle of the study the drug is not doing what we expect, if the trial actually harm patients, we actually need to stop the study before the planned end, and so in the news we sometimes see this kind of bad news on those new trials got stopped early on purely because of bad effects – either there is no treatment effects or actually harmful effects. Another possibility is the drug is so good, we before the study ends we have already accumulated enough evidence to give a positive conclusion. So, in this case, we do not need to wait until the end of the study and we can terminate early so this drug can potentially get approved early on and benefit other patients early on.

Hochster   So, when we come up with a study like this, we kind of write out the whole program for the study and something called protocol or clinical trial protocol and that is approved by our local IRB, that is the people who tell us it is okay to do clinical research in human beings, human experimentation, and in that you will pre-specify we are going to do an interim analysis at some point, and what you are saying is that sometimes the interim analysis, most of the time you are going to say go ahead and finish the study, but sometimes it is so good, you can stop before you finish if the effect is maybe twice as much as one expected or if it really is so ineffective then you will stop the trial early, so more people do not have to be in the trial and get exposed to something that they do not need.

Ma   Yes, that is true. I mean usually will have the protocol saying we are going to do this study for a certain time and we are going to recruit a certain number of patients, but as we have just discussed, this does not have to be exactly what is going to happen, so we can actually conduct those interim analysis and terminate the study either making a very positive conclusion or making a negative conclusion. So, it actually goes both ways and can happen before the planned end of the study.

Hochster   Right, so that is very helpful and a way to protect our patients from being in trials that are not going to give us a positive answer or if it already is very positive. So, you are in the school of public health, that means you look at like a lot of effects out there in normal people, like there is another kind of completely different research that statisticians are involved in looking at, that public health or populations looking for effects like does coffee cause pancreatic cancer or something, can you tell us a little bit about that kind of work.

Ma   So, I mean, we have just discussed clinical trials, but honestly if you look at what we do every day, we do not have that many clinical trials, I mean clinical trials are very expensive, not that common. Actually, we perhaps spend more time doing this kind of population research. So, we are not trying to test a specific new drug, rather we are trying to understand the health at a population level, so as you mentioned something can be as common as like what kind of good or bad effects of coffee or water or apple or banana have on our health. So, what we do is we conduct those observational studies so we do not really conduct the clinical trial on the treatment effects of coffee. Instead what we do, we collect data retrospectively or prospectively, so basically we try to collect information on what ordinary people like you and me do every day and we also try to collect data on our health conditions. So, we are trying to identify the situation or even better effects between what we do every day and our health condition.

Hochster   Like how do you do that, I mean I do not remember what I eat, I mean yesterday maybe, but last week forget it.

00:12:10 into MP3: https://ysm-websites-live-prod.azureedge.net/cancer/2017-YCA-0917-Podcast-Ma_315286_5_v1.mp3

Ma	So, one way of doing study is actually cheaper as to do a retrospective study, so basically what we do is we ask you to fill in questionnaires on questions like what you did yesterday or months ago and as you mentioned, if you have to recall from a long time ago, there is definitely a risk of recall bias. So, another way, which can be more accurate but yes more expensive is to do a perspective study. So, we recruit people like you and we ask you to keep track of what you do tomorrow and the day after. So, basically you keep track of things like what you eat, how long you exercise, when you go to sleep, this kind of information and after a while, like a couple of months or couple of weeks, we contact you again, we get our data back. So, this kind of perspective study is as I mentioned is more accurate, but it is definitely more time consuming and also more expensive.
Hochster	And there are some big studies out there, thousands, tens of thousands of people who have been in these longitudinal studies.
Ma	Yes. So, kind of best way is not to collect cross-sectional data, rather to follow people for a long period of time. So, those studies will be organized mostly by government funding agencies like ASA or FDA. They follow a large cohort of people for really long period of years or even 10 years or even longer, and we have a lot of rich data and very interesting findings from those longitudinal studies.
Hochster	Well, thank you very much for that interesting description, Dr. Ma. We are going to take a short break for a medical minute. Please stay tuned to learn more information about biostatistics and cancer with Dr. Steven Ma.

Medical Minute Support for Yale Cancer Answers is provided by AstraZeneca, committed to providing targeted cancer medicines for patients. When it comes to cancer treatment, one size does not fit all. More information at astrazeneca-us.com.

Breast cancer is the most common cancer in women, and in Connecticut alone, approximately 3000 women will be diagnosed with breast cancer this year, but thanks to earlier detection, noninvasive treatments and novel therapies, there are more options for patients to fight breast cancer than ever before. Women should schedule a baseline mammogram beginning at age 40 or earlier if they have risk factors associated with breast cancer. Digital breast tomosynthesis or 3D mammography is transforming breast screening by significantly reducing unnecessary procedures while picking up more cancers and eliminating some of the fear and anxiety many women experience. More information is available at YaleCancerCenter.org. You are listening to WNPR, Connecticut's public media source for news and ideas.

Hochster	Welcome back to Yale Cancer Answers. This is Dr. Howard Hochster and I am joined tonight by my guest, Dr. Steven Ma, and we are discussing biostatistics and cancer. So, we were just talking about population-based observation studies. You said they are funded by government agencies, they may involve tens of thousands of patients for many years. So, that really gives you a lot of data to work with.

00:15:58 into MP3: https://ysm-websites-live-prod.azureedge.net/cancer/2017-YCA-0917-Podcast-Ma_315286_5_v1.mp3

Ma	Right. So, actually a lot of important findings came from those large studies. So, one example which is actually not very far here, not very far from Yale is a Framingham study. It has been going on for I think 70 or 80 years and is completely funded by the US government, and it has been collecting data on I think maybe over 10,000 subjects for over 80 years and a lot of cardiovascular study fundings actually came from this study.
Hochster	Right the Framingham study. So, you know, lot of times when we hear in the press about something is good for you and something is bad for you, it comes out of these studies, but it seems like it changes all the time. Like, for a while, they were telling everybody to eat more fiber to prevent colon cancer and then it was like well that did not really work. So, how do they come to these conclusions and what are some of the problems in coming up with the right conclusion.
Ma	So, fortunately or unfortunately, statisticians have been involved all those studies and a lot of those fundings actually came from statistical analysis, and problem with those observational studies or population-based studies is those studies were not conducted in well-controlled environment. So, it is not like clinical trials. In population-based studies, you cannot really select who you want to work with and exactly what you do. So, basically those subjects, those people, they just keep doing whatever they have been doing – their eating, their common ways, go to their jobs, so it is not a controlled experiment. So, that is why we have a lot of variations in those studies. And problem with statistical analysis which can also explain why we have those kind of mistaken or at least partial findings is this confounding facts. So, in statistical analysis, pretty much what we identify is association and not necessarily causation. So, we may observe association with fiber and colon cancer, but this is not necessarily causal effects. So, what we really need to do is in the first step, we need to identify those associations. So, those associations can suggest possible targets. The next step is really to conduct more experiments, collect more data and conduct more sophisticated analysis to really identify the causal effect. So, that was kind of interesting and that is what we not just the statisticians, but also medical researchers like yourself what we need to do to really identify the causal effects.
Hochster	So, a good example might be smoking. I mean one of the questions for a long time was and people denied the fact that smoking caused cancer, but if you have a population where people have cancer that smoke, you can show there is an association, but unless you have a control group, it is hard to show what you call causal effect. And you cannot really do a trial where you take people in randomly and say you smoke a pack of cigarettes a day and you do not smoke cigarettes, you cannot smoke, so we have to come up with these other kinds of ways of looking at, which we usually call case control studies.

00:19:49 into MP3: https://ysm-websites-live-prod.azureedge.net/cancer/2017-YCA-0917-Podcast-Ma_315286_5_v1.mp3

6

Ma   So, well there are actually a lot of things we can do, how we can have different designs, we have better control confounding facts. I mean we can do better statistics to only focus on the fact that how smoking trying to remove other possible effects, for example, if people are exposed to high level of pollution for a long time, they also have a higher risk of getting cancer, lung cancer. So, what we need to do is, we really need to control the effects of pollution. When we analyze the effects of smoking, so we need better study design, we need better statistical analysis and also eventually, we need to really understand the biology. So, what kind of biological mechanisms make smoking contributing to lung cancer. So, that is why we have a lot of colleagues in medical school and in school of public health, they do genetic or genomic studies trying to really understand as biology. So, that kind of study and that kind of research can really tell us whether smoking causes lung cancer and more importantly what underlying biology is.

Hochester   So, this kind of public health research has gone from just looking at food, habits, stuff and actually now looking at what is going at the molecular level, kind of molecular epidemiology so we can understand better what is happening to people with specific molecular changes.

Ma   Right. So, that is actually kind of most exciting and most important fact in the development in the past 10, maybe 20 years. The area of genetic epidemiology. So, in the past, we only looked at smoking, we looked at lung cancer, but now, we are really trying to add the genetic component so it is not just epidemiology or genetics, it is genetic epidemiology. So, we are trying to understand at molecular level why those risk factors can cause certain cancers.

Hochster   I see. And to do that, you need to get blood or tissue samples from people who are in studies.

Ma   Right. So, we need biopsy, we need samples from those patients, we may also need samples from healthy patients as a control, and we try to understand there are molecular differences. So, what kind of genetic mutations we see in lung cancer patients that we do not see in normal patients, and whether those kind of mutations are related to smoking or not.

Hochster   And there is a lot of that kind of research going on today at Yale?

Ma   Yes. It is actually not just Yale, it is nationwide and it is also in other countries. So, the research of biomedicine epidemiology has definitely moved into this era pretty much all cancer research we now have a component.

Hochster   That is really interesting. I want to focus and move a little bit over to another topic that I know you are very involved with and that is called bioinformatics. So, we have reached this age of big data and being able to analyze lots of data and predict things that are going to happen, most of us know that because of our Google searches and then we get ads on our browsers all the time, but you actually use that kind of approach in understanding cancer today.

Ma	Right. While you mentioned a big data, it is definitely one of the hottest areas right now, and in cancer research, in statistical research, we have also moved into this big data era bioinformatics. So what we do is we use statistical and mathematical techniques trying to analyze those data trying to relay one big amount of data and really to identify what is the causal kind of mutations are for example for lung cancer.
Hochster	And so, how does that bioinformatics work, I mean we use that today when we are doing gene sequencing?
Ma	Right. So, what we do is well kind of step zero is to really get samples from patients, next step – step one is to do a profiling studying, like sequencing, you mentioned sequencing, so we will do a sequencing study to really see what kind of mutation profiles patients have and the next step is to conduct statistical and mathematical analysis using really complicated software to really identify what kind of mutations may cause a higher risk for cancer, may cause poor prognosis and may cause bad response to treatment.
Hochster	And so, people like doing these big data or bioinformatics analysis, how many calculations or numbers are they dealing with.
Ma	That is actually a huge number. So, if you look at number of genes, we have over 20,000. So, that is already a big number, but actually not too bad but if you look at sequencing data, if you look at mutation data, we are dealing with millions of mutations. So, basically we are trying to identify from those millions of mutations one or a small number of really bad mutations, which may contribute to cancer. So, while I still remember a long time ago, we said it is really hard to find a needle from a haystack, but right now, if you think about it, it is actually not too bad, I mean a needle actually looks different from haystack. So, what we are doing right now in those big data analysis of mutation data, we are trying to identify one or small number of very special hay from big haystack. So, that is actually much more thrilling.
Hochester	The needle is easier to find than a specific piece of hay. Okay. So, and this is really pretty much common now in every kind of molecular analysis. They look at proteins, RNA, DNA, whatever and kind of analyze millions of pieces of data to find out which things are more prevalent and which are less prevalent.
Ma	Right. So, what is really interesting about this research on cancer and also other types of complex diseases is if you look at those DNA level changes, amylase, proteins, you have a large number of measurements and what is more interesting and also more complicated is all those different levels of changes are actually interconnected. So, kind of the first step is to really understand for example how DNA level changes affects amylase and how amylase regulate proteins. So, to really understand biology another second step – the next step is to really understand how those molecular changes, how they contribute to cancer.

00:28:08 into MP3: https://ysm-websites-live-prod.azureedge.net/cancer/2017-YCA-0917-Podcast-Ma_315286_5_v1.mp3

Hochster   And is this all being done using bigger and faster computers. Is that what the basic thing is to be able to do this kind of research?

Ma   Actually, yes. I mean just imagine how large those data are and how complex the relationships are. Right now, our research definitely involves a lot of computers and programming and new hardware, now software. So, sometimes I feel I sit in front of computer longer than a programmer. Sometimes, I feel my job is more like a computer scientist.

Hochster   And where do you see the field going in the near future. What is going to be happening with this area of bioinformatics?

Ma   So, I think in the past decades we have collected enough data and finally we are at the point of being able to actually understand cancer from a more fundamental level. I am an optimistic person and I think in the couple of years, maybe 5 or maybe 10 years, finally we are going to have good models to really predict what is going to happen to healthy people and what is going to happen to cancer patients, and at some point, those findings those bioinformatics findings will guide us to really discover new and effective drugs to finally treat cancer.

Dr. Steven Ma is a Professor of Biostatistics at the Yale School of Public Health and Director of the Yale Cancer Center Biostatistics Shared Resource. If you have questions, the address is canceranswers@yale.edu and past editions of the program are available in audio and written form at YaleCancerCenter.org. I am Bruce Barber reminding you to tune in each week to learn more about the fight against cancer. You are on WNPR, Connecticut's public media source for news and ideas.