Advancements in Diabetes Seminar
Understanding & Using Clinical Data: critical appraisal of medical literature for understanding and applying results from clinical trials into your practice.

Ryan Schupbach, PharmD, BCPS, CACP, NCPS
Vice Chair, IHS National Pharmacy & Therapeutics Committee

Ryan Schupbach:

Thank you very much and welcome to everyone. As Jan mentioned, my name is Ryan Schupbach, and I currently serve as the Vice Chairman for the Indian Health Service's National P&T Committee. We're based out of Oklahoma City here, but those of you who may not be aware of us, we are a group of 14 physicians and pharmacists who meet routinely, and we are charged with managing and maintaining and updating the Indian Health Service's National Core Formulary. I've been in this role about three years now. And prior to this, I did spend most of my time as a practicing pharmacist at the Claremore Indian Hospital. And I'm using my time as a direct patient care provider really to guide me through these slides and use that kind of as a template. Before we begin, I do want to just extend some gratitude to Dr. Bullock and the entire Division of Diabetes Treatment and Prevention for extending again this invitation to me.

It's an honor to be here, and what I hope to do over the course for the next 45 slides or so, is cover obviously some of the objectives but more importantly provide you with some practical points when you pick up some medical literature or publication and how best to discern whether it's worth your time to even delve into that publication. My disclaimer here, let me put this out front, when I sort of was collating my thoughts, I thought how best to do this, and my understanding is all of you are pretty much practicing clinicians as well. And so, I thought about my time in the clinic seeing patients and there were limited amounts of times throughout the day where we could, my colleagues and I could share thoughts on publications. But someone would generally print a new drug publication or an FDA approval study off, and we might have time at the beginning of the day or over lunch to sort of go through this publication. So, again, as I crafted this, I tried to keep that in mind, we'll start at the beginning sort of through the abstract and then you work your way through the introduction, the methods, the results and discussion. And so, that being said, my disclaimer is simply that all the things that I'm going to talk about are really somewhat specific to the type of study that compares, A, drugs, and there are many studies out there that different therapies and so forth. But, being the pharmacist background, I suppose that what we'll be talking about are drugs versus

placebo or the standard of care. And most of this information is applicable when we're looking to show is one better than the other, and also known as the superiority study.

So, that being said, I do want to recognize one of the primary references or resources that I used. The National P&T committee because of the very nature of what we do, we're routinely assessing and evaluating the benefit of new medications versus ones that are available on the National Core Formulary. So, I took what, I sort of learned over the course of my 16 years as a pharmacist and my 3 years here, but also incorporated a lot of the information from this book as well. So, I just want to point this out. This is really one of the primary resources that I used and I would highly encourage. It's a pretty easy read, it's about a hundred pages, but its large print which is good for me, and I think that you'll get through it. It's written in a way that is very understandable. So, I would encourage you to -- if you're interested in this type of thing to take a look at that.

So, let's start with really the problem, what brings us here is that, we're expected as practitioners to stay current on all medical literature, specifically within our practice setting. But the problem lies in that the National Library of Medicine is stating that they're publishing on average about 13,000 new references each week. And a study I pulled from 2010 and that's eight years ago, quoted that there was roughly one new publication every 26 seconds and sort of compounding the problem there was that last year, and I suppose this is good but there were record number of new drug approvals through the FDA that just add to this. So, how are we expected to remain clinically up to date with this litany of publications that were just bombarded with?

Well, several decades ago, journals understood this and began putting forth these abstracts, right? There was a way that you could easily sort of peruse the abstract and figure out if it was applicable to you or if it had value and if not, you could stop reading right there. Well, the problem is and what we found is that the accuracy of the data in the abstracts has really been left to both the author and the journal's editor. And often times what we're seeing is that validation of those results are really not happening. In fact, this particular study that I've referenced here is up to 70% of abstracts and this is really in the top tier journals. So, JAMA, New England Journal of Medicine, Lancet, BMJ, so forth that information in the abstracts were actually not corroborated anywhere in the body of the study. So, that should make us question whether the abstract is the end all be all.

But additionally, and I don't know about everyone's discipline and their training, their school of training, but in pharmacy school we have essentially one semester where we cover the clinical aspects of drugs not only in the body but specifically with regard to study design and methodology. And it's a fairly intense semester but it's only one. I don't know if that is universal to all other disciplines but I would suspect that's probably not the case. And what we're finding out now from a variety of different resources is that most of our healthcare providers especially ones making decisions, don't really have the background or the training to be able to elicit whether these new medications truly have benefit or not.

And in addition to that, this idea of informed consent and bringing the patient in as a stakeholder and making therapeutic decisions with their voice in mind really, again, sort of compounds the problem. And I've got this I think it's almost a perfect example here. I've pulled up and I'm going to make reference to the SGLT2 inhibitors on several occasions. But the reason that I bring this up, A, it was we just reviewed this class in August. But, B, this alliance with what I'm talking about with regard to patient preferences, if we're able to glean meaningful information from the study, we need to translate that and give that to our patients in a meaningful way where they can make a decision as well.

And if we look at this particular study, the CANVAS study with again, canagliflozin, it's an SGLT2 inhibitor. I'm sure most of you are familiar with this but it reduces blood sugar modestly anywhere from about 0.5% to 0.8% on the A1C through excreting it in urinary glucose. The problem is, is that what we found from this particular study and this was a cardiovascular outcome study, but what we found is, it does -- we know that it reduces the A1C modestly but we also noted that there is nearly a doubling of the risk of lower limb amputations. And so, if we're able to give that information to the patients say, "Well, it may modestly lower your blood sugar but it doubles your risk of losing a lower limb." Chances are many of the patients may elect to look for different therapeutic options.

The good news about this and this term that we're all sort of, this global term is critical appraisal. And that's just comprehensively evaluating the study from start to finish. This term, it's really a fairly easy process. It doesn't involve a lot of required statistical knowledge. There's really no fundamentally no best way to do it but there are some characteristics that are similar in however we choose to appraise the study. The focus with critical appraisal is not trying to prove that the outcomes are accurate. It's actually the opposite. It's trying to find the problems in the study, and if you can identify problems which may introduce some uncertainty into the study, where you can put the study down and you don't need to go further with it. You can save your time and move on to particularly another publication of interest. And so, the Delfini Group they just say, "This process is inexact", and I can speak from experience to that this process like many others like practice or anything else just sort of build on your experiences from prior experiences.

I've tried to break this down this process into three fairly easy steps right here, the first being a review of external validity. And before we begin, I do want to just make clear some definitions. There are some terms that we'll be using and I just want to clarify these. But the term validity just represents the ability of the trial to actually find what it was designed to do. And so, we ask our self that does, if a study is valid then it gives us an accurate outcome that represents the truth. Whereas, reliability equally important is the ability of the study if reproduced would lead to the same outcomes, and this is really becoming a much bigger issue. These are being challenged in smaller settings and coming up with drastically different outcomes.

So, two other definitions I just want to make mention, external and internal validity. When we talk about external validity and the easiest way is just to ask yourself, do the conclusions of the study, are they applicable to my patients? Whereas internal validity, this is sort of the whole process that leads to an outcome and we have to ask ourselves, do I feel confident that the conclusions, the end point, that the measure of interest is accurate here.

So, if we want to start again with external validity, we need to -- you might ask yourself when you're reading through this, are these populations similar to my patients? I think, we all understand and recognize the fact that American Indian and Alaska Natives are grossly underrepresented in research in general. And that's for another time there's -- on both sides there's been historical mistrust and -- but ultimately what happens is that we see very, very, minute populations involved. The American Indian patients are generally in the less than 1% or other category when we look in this main stream publications and we need to appreciate that. There are genetic variations in all races, all ethnicities that play a role in how the body metabolizes and uses the drugs.

And so, the concern here is that if we look at a study which the patients are drastically different. We're making sort of a –we are taking a leap of faith to extrapolate and assume that what was seen, the outcomes that were seen in the study will be similar to what we see in our patients. And so, we need to at least appreciate if there is a start contrast in the patients then it may not be applicable to our patients, in our practice setting.

Some other things we want to look at, is the site is it international or is it very focused? I got a couple of examples, we'll talk about there. Is the study -- are there multiple sites where the study is conducted or is it really just one single site? In a good way, again, to look at your patients and all of this is to really look through the inclusion and exclusion criteria. And I get upset a lot of times because I hear practitioners talk about, these drugs in settings that they haven't been tested. And I think it's just ignorance to where it has been tested. If you think about most studies rarely are they ever studied in patients with a GFR of less than 30, certainly not pregnant patients or often times obese patients are excluded, especially morbidly obese. This is an issue that we run into all the time especially with our oral anticoagulants.

But, if you feel like that the populations are similar to your patient, then ask yourself, will results, will those results be again applicable. And the example I've got here, two examples. One study I read was about 400 Taiwanese diabetics in a single site hospital in Taipei, it's unlikely that we'll see a mirror outcome in our patients and really any study from the VA, and I love the VA, they're partners with us and that's -- they do an excellent job with the research. It's just that I think we all can appreciate that their study groups, their patient cohorts are greater than 65 years of age, white males predominantly. So, we need to be careful again when we make some leaps that the study, the outcomes in seen in these will translate to our patients.

Another question on usefulness on the study, again, you just pick this up and you're sort of gleaning -- trying to glean some useful information out of it to determine whether you want to keep going is, are the outcomes meaningful to my patient?

So, I'm going to start here sort of in the middle, and if you see surrogate outcomes also known as intermediate outcomes, you want to have something of a healthy dose of skepticism. What we're looking for is some degree of what we call the five primary outcomes. And these are universally agreed upon to have more meaning to the patient. Is there some degree of morbidity that's being tested or mortality, symptom relief, quality of life, or some degree of functioning whether mental, physical, or emotional? These should be your primary outcomes if it's really going to be important to your patients. If you start to see this, and we see some very common ones, and this is okay but we have to remember that not all surrogate outcomes, they're surrogates because they're intermediate, they may not all be validated in what we're looking for which are the hard outcomes, right, stroke, MI, death.

I also want to make mention of composite outcomes. We see this more and more, and really the reason that we see composite outcomes is when it all boils down to it is cost. What I mean by composite outcomes is just a culmination of two or more outcomes. So, I've got an example here, death and chest pain, so it's one or the other. And it's important to know with composite outcomes if one of these occurs, then the whole primary outcome is considered positive. And so, going back to why this is done. Well, if by combining multiple end points or outcomes, you can meet an event rate very quickly, right? And if you hit the event rate that you've pre-determined then you can actually -- you don't need as many patients and that will save you time and money.

So, that's really why you see a lot of composite outcomes and they're always -- some are mandated with diabetes drugs for cardiovascular risk, the FDA mandates the big three; all-cause mortality, non-fatal MI and non-fatal stroke. So, I want you to be careful with the composite outcomes because in my example, you need to look at each individually. So, my example might be, let's say, new drug A compared with placebo, and it shows much less primary outcomes. That's good, right? But when we look at this, we see that there's a lot less chest pain. However, death is actually increased with new drug A versus placebo. Clearly, we would want to step back and ask ourselves, is this really in the best interest of our patients to add this drug to our formulary or to give it to any patient for that matter. So, when you're looking at this just recognize that anyone of these in this multitude can serve as a primary result and need to look at those individually.

Alright! So, that's external validity. Now, we're going to focus on internal validity. And this is really looking at sort of the nuts and bolts of the entire process to which you derive ultimately at the end with an outcome. And so, the very basic, at the core, the studies that you're going to be reading about, either they're observational or experimental. And by experimental, simply we mean that there are two or more groups and their enrollment into one of those groups is randomly decided. In observational studies, you may see two or more groups but they are not

designed to go into this specific. They're not randomly allocated, they're not assigned, and so that's really the difference when one of the most striking differences anyway.

We want experimental studies, this is largely determined as the only method determined if there is really what we call cause and effect or a true difference. Observational studies have value. They tell us a lot about the prognosis, the natural history of these specifically. But because they're not randomized there's a chance for error in there and they're highly prone to bias. The results of observational studies should also-- should always be hypothesis generating. The design is not -- the observational study design isn't designed in a way to rule out the potential for error. So, ask yourself when you get in and start thinking about internal validity, is it an experimental study or is it an observational study? If it's observational, be very, very skeptical about what you can take from that because it's not designed to rule out the potential for bias.

Now I would be remiss if I didn't put this in. I'm sure we're all somewhat familiar with this, we've seen this. This is sort of just the hierarchy of strength methodologically, and we see really at top the randomized control trials. Because the randomized control trials have characteristics that are intentionally included, they are systematically designed to minimize or mitigate confounders, bias, and chance, and we're going to talk more about that. But if you have a new drug and you are submitting to the FDA, the FDA will mandate that only data from randomized control trials are included.

Unfortunately, I don't have time to talk much about meta-analysis or systematic reviews. But briefly, systematic review is just a review of a topic, generally several studies, but it's pre-determined what will be included in the review and what won't. It's somewhat similar and oftentimes combined with meta-analysis. Meta-analysis is just taking several studies and combining them. The idea is that you increase the number of patients and you can oftentimes determine more accurately if there's a true outcome or not. Imagine if you had five studies all studying basically the same thing. Two, were what we call negative studies, meaning there wasn't a difference. And three, we're positive. Well, if you meta-analyzed all five of these you might come up with a decided answer in the clinical question. So, for all intents and purposes for this presentation, we'll be focusing really mostly on randomized control trials.

I mentioned there are some things that can be introduced within a study that really confuse us that lead to inaccurate results. And these are generally called the four reasons that explain the study or the drug itself and the outcome.

The first one I'm to talk about is cause and effect. Now this is just truth. There may be a situation where new drug A is truly better than placebo and that is just cause and effect or truth. There are also conditions where we may have an inaccurate outcome. Bias may lead to that, it may be due to confounding, or simply may be due just to chance. It's a statistical anomaly. And so, we're going to talk about all three of these and how to look for this and what tools to use to rule this out. Because if we can find any of this in the study then we can throw

the study out, but if we can't, then we have to be led to think that there might actually be some benefit to the drug in question.

Lastly, the study, if you have gone through and you can't find any bias or confounding and it doesn't appear to be that the outcome is not due to chance, then the study is said to have internal validity.

So, let's move forward and talk about bias. What is bias? It has a myriad of different definitions. It's essentially anything that leads us away from the truth. Another definition might be anything other than the study drug which is intentionally separate and different from placebo. Anything else in that process is a bias. And it always tends to favor the intervention. We see this time and time again, that if there's a problem within the study and it leads to inaccurate results, they always tend to favor the intervention drug.

I think we can appreciate that certainly if the researchers are part of or have vested interests in seeing the drug be approved, then there clearly could be some potential for bias, obviously, a conflict of interest. But we need to keep in mind that even the researchers, whether it be consciously or subconsciously, probably contribute a little bit to some degree of bias. If you think about academicians, their clinical currency, if you will, is to publish as many manuscripts and publications as possible. That's oftentimes how they're graded for tenure. And so, if the result is negative they're much less inclined to produce a lot of publications. But if it is granted approval, then there would probably be a lot more publications for them.

There are number of different biases, and I don't want to take time. These are just a few, just a snippet. Attrition bias deals with dropping, patients dropping out. Classification is how they're grouped together. Publication bias is whether the study has a negative outcome. Is it still published? What we see is generally, well, significantly more studies that have positive results that favor the new drug are studied whereas ones that are negative don't seem to make that as well. Selection and recall bias, any component of the clinical trial, the process in the clinical trial, if it leads us away from the truth, there is likely some degree of bias and it's been named.

We mentioned, why do the FDA and why does everyone rely so heavily on randomized control trials? Well, they seem to have the tools that we're looking for, the characteristics of the study that really look to minimalize or mitigate bias. And so, when you're looking through these studies, you should look, is the study randomized? And I think we're familiar with most of these for the most part, but you may see it referenced as allocation or allocation concealment. That's just means if the patient comes in and let's say they're enrolled in the study in the ER, that concealment that they're unaware how they're grouped is hidden. So, it may be through like sealed envelopes, they're handed a sealed envelope and no one knows. Sequencing is how they're put into that. Sometimes the patient may be in the ER and they're given a phone number and they call and the third party-vendor has a random sequence generator, it's a

program that gives them, "Are you in group one or group two?" So, we want to make sure that allocation and randomization is occurring in our study.

Blinding, again, another fairly common term. This is just preventing one or both sides of the clinical study, the research team and the patients, from knowing if they're receiving an active drug or if it's a placebo. We see single, double, and triple-blinding. Triple-blinding just implies that if you have a situation where you need a radiologist to review an imaging study to determine the outcome, well, the trial may enroll some third-party vendor of a group of radiologists and those radiologists also need to be blinded so that they don't introduce any particular type of bias. So, blinding is incredibly important. Double-blinding, it's like sort of double-dummy. I don't know if you're familiar with double-dummy. It's not used as often, but double-dummy just means that let's say there's an instance where you're comparing two drugs but one is an oral tablet and one is a subcutaneous injection. Well, in order to be double-dummy, obviously both sides would have to receive both an oral tablet and both an injectable but one would be active and one would be placebo. So, you may see double-blind or double-dummy and that's what that ultimately means.

The last component that I want to touch on is just following up of patients who are dropping out. They call this the attrition or potentially attrition bias. Why did half of the patients drop out? That's obviously very important to us. Did they deviate the protocol? Did they move out of town? Was there something so egregious that it was significant, diarrhea or who knows? Whatever the case may be, we need to evaluate that, because that plays a role in how applicable it maybe to our patient population.

There are a couple of different designs that look to avoid this attrition bias. And you're probably familiar most commonly with the intention-to-treat design. And there are really two, intention-to-treat or per-protocol. Really, the difference is in intention-to-treat all patients who start the study are included. Whether they finish or not, they are included in the final analysis. So, you really get a better sense of why they dropped out, and that's a clinically important question. In per-protocol, this differs a little bit because it only includes patients who actually finished the treatment series. So, it may be a 12-week study or something. And all those who drop out, they are not included in this per-protocol.

I've got an illustration here that I think will kind of make a little bit more sense. So, here's the intention-to-treat. And it includes in yellow here, we've got all of our patients finishing, only those that finished the study, and then here's our 20% that dropped out. Well, when you analyze this based on an intention-to-treat design, you're getting all the patients. Whereas as per-protocol only looks at those who finished. So, you'll never know how many dropped out. The good thing about per-protocol, and it's pretty limited on the advantages of per-protocol; one, the most significant advantage of per-protocol is you know exactly what to expect from the drug in patients who can tolerate it. So, it's not sort of watered down because it does not include those discontinuing that.

So, I'm going to go back to our study. And if you want to know, we are always looking for intention-to-treat. That's the one that is highly preferred. But the easy way to tell, they should tell you in the study, but in the event they don't, what you want to do is you can easily look at the number who were randomized and enrolled and then that should be exactly the same as the number who end up finishing the study. If that is the case, you know that it is an intention-to-treat design.

The second thing that can sort of lead away from the truth is confounding. Confounding, it's very similar to bias, it's technically not the same. But oftentimes you'll either see the term "confounding bias" which is a little bit misleading. But bias is in the end when you look at the outcome and it's inaccurate and that is because of it is biased, that outcome is biased. The confounding or the confounder is the actual variable that's in the study that actually leads to this biased outcome. I hope that sort of makes sense. The confounder is the actual element within the study that led to the biased outcome. That's probably the best way I can explain that.

I've got an example here that may help. It's actually based on some degree of truth. A new antidepressant is believed to decrease the risk of suicide. So, if that's the case, many prescribers will put their highest-risk patients on the new antidepressant and they're going to leave older patients, or more stable patients, rather, on older antidepressants. Well, if this is done long enough, someone may go in and review a database and they start seeing, they're going to note higher rates of suicide associated with the new antidepressant. Well that isn't entirely true. Yes, it was done. The prescriber was justified in doing that. But it is obviously misleading.

And this term called "confounding by indication", that's when clinicians tailor the drugs to meet the needs of the patient. Which is completely fine unless you're in an experimental situation where you're trying to discern if one is better than the other. And this one is severity illness, oftentimes you get a selection bias from that. The good news is, with confounding, you can rule it out if your study has effective randomization blinding. If you see those two trial characteristics, you don't have to worry about confounding. For the most part, it's been ruled out by that particular study design.

Now, Chance is simply just that, it's just purely by accident. Drug A was found to be better than placebo. Now there are a couple of things that actually increase our risk of chance findings. Really, the most specific and significant for that matter is small sample size. When you start getting patients less than 100, you really ratchet up the potential for just random error. It could be just the numbers worked out. And so you want to be really skeptical on these smaller studies.

Additionally, other areas that may increase your risk, when you start looking at outcomes that weren't designed before the study started. You see this a lot with interim analyses. What happens oftentimes is the sponsor is putting up $500 million for this and they'll stop and they'll

say, "All right, at some point maybe halfway through we want to stop and re-run the numbers and see if it's worthwhile." It could be for safety, we would like to think that they're doing for safety. But chances are they're doing it to hedge their stake in this. If it's not showing benefit, then they may pull the plug on this. We see this all the time.

But these interim analyses, a perfect example was, it happened about two years ago, 25% into a study, the CEO leaks out that they're showing significant gains, I think it was in HDL increases and LDL decreases, and there was a big lawsuit that came from it. The problem is because you've not designed the study to look at it at that point, there maybe all kinds of problems and risk is multiplied at that point, or at least the potential for risk. The good news here is we have a couple of tools that we can use to reduce the chance of chance. And I apologize for the way that it came out, but I think you understand what I'm saying.

The P-value and confidence interval can really help us determine if the outcome is due to Chance. P-values, probably if you recall anything from at least from pharmacy school, I remember that P-values was the end-all be-all for statistical significance. If you knew that P equals less than 0.05, you can pretty much get a B in the course. And so, this is commonly referred to as a cutoff for significance. And really, at some point eons ago, some statistician decided that 5% is okay. And the P-value says that if it's less than 5%, we're okay, we think that it's okay, it's all right, we'll assume that it's not based on chance. It's really not as helpful as we think. There are some assumptions that have to be made. That studies have to be randomized, you can't really use P-values in observational studies, it doesn't make any sense. So the p-value, while it has some merit, it's really sort of we're moving away from its over-emphasis in statistical significance.

I put this little guy here. When I first saw him, I thought he was sort of bowing down to the P-value, emphasizing our over reliance. But I think he's sort of mourning the death of the p-value. At least I hope that's the case.

Interestingly, a few years ago a statistician out of the United Kingdom published this study talking about false discovery rate and misinterpretation of P-value. He's a statistician and he has got it all backed up, it's not a very interesting read but I've sort of clipped the important pieces here, but he states if you accept 0.05 as your benchmark for saying that something is significant, you're going to make an idiot out of yourself about 30% of the time, I think it was 29.2. But accepting P as equal to 0.05 means you're going to be wrong and it's going to be due to chance about 1/3 of the time. And he went on to talk about he ran the numbers on there, and it's pretty impressive. But he said we really need to change our ways and consider a P of less than 0.001 as a demonstration of significance. And this equates to like 1.9% chance that's it's due to chance." The Delfini Group, they're not very impressed with the P-values either and they've talked about either using confirmatory studies or patterns to really determine if the outcome is true and not due to chance alone.

Confidence intervals, again, another fairly common concept for reviewing this, is actually more helpful than P-values. And they're really this range or interval where the true value can lie 95% of the time. That's what it essentially tells us. And this is sort of written backwards. It says that a 95% confidence interval tells us that 5% of the time the value lies outside the range. Well, I would look at that and say 95% of the time the true value lies within this interval. And we want narrow confidence intervals rather than wide confidence intervals.

You can see here, here is our example. This is our confidence interval, 95%. This is what we call the point estimate. It's sort of the average or the mean, if you will, from the point estimate. But really, we can say that because the outcome could really be anywhere in here, that we must be cleanly on one side of the line of no difference. This vertical line right here and you can an example with both C and D, if it crosses the line of no difference, also known as the "line of unity" but I think that's a little silly, if this confidence interval passes the line of no difference, then we have to say there's really no difference between these treatments because the outcome could be right there on the other side. We don't know that, only in situations where these intervals cleanly are on one side versus the other.

So, again, if we would look at the study and we see the characteristics such as randomization, blinding, intention-to-treat, we look at our P-value and our confidence interval, we see all these things and there are no problems there, we can say that our study is internally valid. So now we can go on to the last section. And that's appreciating the outcome and trying to discern whether it's truly meaningful or not.

So in our situation, let's say new drug A is statistically significant. It has benefit over placebo. But an equally important question might be, "How much better is it?" And this is known as the effect size or magnitude of difference. And this is really the basis behind statistical significance versus clinical significance. Drug A can be better than placebo. But if it's just marginally better and it costs $13,000 a month, are we really going to spend all this money on this? This just doesn't make any sense. So, we have to be able to appreciate how large of an effect the study drug may have. And there are two ways to look at that, these measures of effect size, and the broad categories include measures of probability and odds.

What you're probably beginning to say is, "All right, this is where I usually get lost." This was a problem for a lot of us, especially in pharmacy school, sort of understanding. And I would question that it's really not important to know that these are probability and these are odds. Yes, they have some value, but interpreting the specifics are more important than actually knowing how to calculate or understanding these specifically. You're probably familiar with absolute risk, absolute risk reduction. I've got some very easy examples we'll go through and we'll talk about odds ratio and hazard ratio and then we'll sort of wrap up here.

The first thing I want to talk about is absolute risk, and this is very basic. We're talking about probability here. What's the probability that an event is going to occur in one group? That's absolute risk. Here's my example, and it's going to be the same example we use throughout.

It is extremely basic. Two groups of patients are in a study and they both have an outcome, a bad outcome, at different rates. In our placebo group, 15 patients out of 100 die. We'll just say death is the bad outcome. The study group, this is new drug A, only 10% out of a 100 die. So our absolute risk is 15% for placebo and 10% for our study. Absolute risk reduction is simply the difference between these two, simple subtraction. Absolute risk reduction in this case is the absolute risk of the placebo minus the absolute risk of the study group, we get 5%.

Said another way, what does this mean? This means 5% more patients who took new drug A, did not die -- and that's terrible grammar and I apologize -- but if you took the study drug, you died less 5%. That's an equally bad way to explain that as well. But more importantly is the next slide. Because who cares about absolute risk reduction, right? Well, it helps us to form a calculation that we can use to again measure the difference or the benefit of the drug. And that's number needed to treat. You probably see this, you may not have a firm understanding of what it is, but it's ultimately the number of patients who need to receive the study drug in order for one person to avoid a bad outcome or to have a good outcome. Why we use absolute risk reduction is because in order to generate number needed to treat it's just the reciprocal of the absolute risk reduction. In order to get this, we just take one and we divide it by the absolute risk reduction.

The same example here, two groups, bad outcome, placebo group 15%, study group 10%. We already calculated this on our last slide, 5%. Now, in order to get the number needed to treat, we're just going to take one, divide that by the absolute risk reduction or 0.05, and we get 20. So, our number needed to treat is 20. What does this mean? For every 20 patients who took the study drug, one additional patient would not die versus those in the control group.

So now, the number needed to treat is really helpful when you are comparing drugs within the same class. It's clearly pretty easy to calculate, it helps us for a comparison. It's easy to do, we can generally pull the numbers out of the study. But it does have some limitations. Oftentimes it averages sort of the patients, if you will. If a patient is high risk, or low risk, you can't really take that into consideration. It just sort of averages everyone's risk and gives you sort of what you call the most probable value.

Another limitation is that the number needed to treat, the actual number is subject to interpretation. So in this example, this is a different example, the number needed to treat was a hundred over five years to avoid one event. Well, is that good or bad? Some people may say, "Oh, that's great," some people may say, "I don't think that's – that's a lot of patients taking a drug for a long period of time, it may not be very beneficial." Well, you have to know what we are avoiding, for starters.

The last limitation I want to talk about briefly is that the timeframe at which the drug was studied is important. Because in this example, the number needed to treat is a hundred, but we have to remember that the study was over five years. So, it's not just a hundred people take it, they have to take it for five years. You've got to have that information in there because

benefit of treatment is sometimes not linear. You may see significant value in the first three months, and then that clinical benefit seems to be mitigated as the drug stays in the system. We see that often times.

Here's my other example. This is a live example. This is one that we calculated in our August meeting for the National P&T Committee. But this deals with, again, one of the other SGLT2 inhibitors, empagliflozin, and again, this was in the EMPA-REG study. Excuse me, let me back up and say that the primary outcome was a composite outcome of all-cause mortality, it was nonfatal MI and nonfatal stroke. So, it was three different outcomes, if you had any one of those, it registered as a hit.

So, the absolute risk was 10.5%. And again, these are all straight from the study. The placebo was 12.1%. And so, if we do the absolute risk reduction, we subtract the difference, this is our rate. The number needed to treat is one over the absolute risk reduction. So, we get a number of 63. So, 63 patients need to be treated with empagliflozin for three years to avoid one of the primary outcomes. My question to you is, is this good? What if the outcome where ER admissions or urinary infections or something? And so, that's where the subjectivity of the number needed to treat comes into play.

All right, relative risk, I want to keep going here. This is just the risk of an event relative or compared to the other group. So, the same example, 15% in the placebo group, 10% in the study group, the same risk. The relative risk is just this is the study group relative to the placebo. And when you divide these, you get a 0.67 or 67%. This 0.67, this is our point estimate and when you're dealing with relative risk, because it's relative, less than one represents a lower risk for the drug, whereas a number of one or higher means there's higher risk for the study drug. In this case, patients in the study group have a reduced risk of 67%.

Pretty simple. The way that epidemiology text books put this together, it makes it a lot more confusing. There's a two-by-two contingency table that you're supposed to use and there's a very fancy formula that you're supposed to use. But really when it blows down to it, this is it at its core.

So, relative risk reduction, as you can imagine, is just the relative risk, which we just calculated. One subtracted from the relative risk. So, the same group, same everything here, same example, we know our relative risk was 0.67 or 67%. So, one minus the relative risk, you get a relative risk reduction of 33%, right? Patients in the study had a relative 33% reduction in risk.

Now, it's important to understand how these numbers can be exploited. And so, I brought this up. This is probably one of the more egregious exploitations of relative risk reduction using it that way. This is actually a study from 1996, from Lancet. This looked at using alendronate which is a bisphosphonate, Fosamax used in osteoporotic patients to reduce fractures. Back 20 years ago this was very expensive, there was concern about esophagitis and all kinds of

problems. But in this particular study, it was alendronate versus placebo for three months. And the primary outcome was looking at new vertebral fractures. We're just looking at fractures in the vertebrate. These patients already had them so they were at high risk. But what came out of this, and what we were hearing in industry representatives were coming in to prescribers' offices and saying, "Well, Fosamax reduces hip fractures by 50%."

Well, and that's not necessarily what it was designed to do. But as a secondary outcome, and this is actually page four of the study, this little secondary end point I've blown up for us here, if you look, here's any clinical features, these are some of the things they were looking at, and here's placebo versus alendronate. What we see if we look at hip and we move over here to relative risk, there was a relative risk reduction of 50%. And I've done the numbers down here, this is all accurate. What I think is more clinically meaningful is looking at the absolute risks. So placebo patients, patients on no drug, 2% of them had a hip fracture, whereas one 1% on the drug now had hip fractures. So, there was a reduction of 50%. But if we don't stop and appreciate absolute risks and we take relative risks, relative risk reduction into consideration, I think we can be statistically misled.

All right, odds ratio. Now we're switching. We've been talking about probability, now we're going to talk about odds. And it's simply just the odds of the event occurring versus not occurring. Again, I don't know if this is terribly important. I can count on the number of times on one hand where I've calculated an odds ratio. This information is always given to you. I think it's important to know some details of it, but will you be calculating this? Probably not. If you wanted to, you could search online, there are online calculators and it's very easy to do. But the difference here is that odds ratio tends to over-estimate risk sometimes. As your incidence gets higher, you can see over estimates are a little exaggerated. Relative risk really doesn't do this for the most part. Relative risk is only used in prospective studies.

So, our example, it has change just a little bit here. Our control group, 20 out of 100 died, whereas our study group, only 10 out of 100 died. And the difference here is really how we calculate this. Instead of a denominator of the total 100, now we have a combined total of 100. And so, that gives us a percentage. Then we put this, because it's a ratio of odds, we divide these and we get this number which means the odds of dying in the control group are greater than two times that of the study group. And that makes sense, right? The odds of dying are 25% if you took placebo and 11% if you did that. So that's really odds ratio. I don't want to spend too much time on it because, honestly, I have four minutes and I've got a couple more slides here.

Hazard ratio, this is becoming more popular as an outcome, as a measure of effect. These are used in what we call "time-to-event" studies. And so, you'll see this. And time-to-event, it used to be hazard ratios were generally relegated just to cancer patients where you had a survival curve, and unfortunately, time to event, we knew they had cancer, it was just unfortunately a matter of time until they died, and so we sort of created this curve, this Kaplan-Meier Curve, and we could apply that. Now we're seeing hazard ratio as the primary measure of effect in a

lot of these studies, especially in the studies we just talked about where these were high-risk cardiovascular patients with diabetics. The empagliflozin had a hazard ratio because it was a matter of time until they had one of those outcomes. I think how, it's calculated, it's similar to OR, I don't think it's really important. But it will suffice to say that if a hazard ratio was two, and again, it's somewhat similar to how you interpret odds ratio, that a patient who has not experienced the event is twice as likely to have it at the next time point. And that's kind of what's different with hazard ratios. It's always from one point to another. They're on this curve or slope, there are these time periods.

We're getting close to the end here. I want to wrap up a little bit with just considering, so we read through the abstract we read through the introduction, we've made our way through the methods and the results. And now we're getting to the discussion, the last thing. I really want to encourage you to consider possibly not even reading the discussion. And I say this, it's really hard for me to do as well, but generally the discussion are the author's opinions and they are just filled with speculation and conjecture, their opinions. Imagine if the sponsor spent $750 million on the drug and you are the primary lead author and you don't know why but it didn't show any difference. Well, if you want to be employed by this sponsor again, you're going to start coming up with some ideas about why that may be the case. And so, there's always some degree of the potential for bias involved in that. The Delfini Group talks about don't even read that, you've got what you need from the intro, the methods, and the results. There's really no reason to read the discussion.

All right, I'm going to summarize here. When you open up this study, ask yourself. Is it applicable to your patients? Are the outcomes measured, are they meaningful? Do they have the right patients? Do you think you can extrapolate these results and they would be useful to your patients? Is the study an observation or an experiment? What's the design there? If it's an observation, you really just need to leave it be. I wouldn't even read the results because it's probably going to mislead you. When you look through the study design, the methods, can you identify if there are problems there? Is it prospective? We hope so. We want it to be randomized. We want it to have two arms where one is the control and one is the active group. Is it blinded? Do they take into consideration drop-outs? When you look at the results, do you see that the confidence interval clearly is on one side or the other? Do the numbers needed to treat, does that indicate that there's clinical or statistical significance?

And lastly, do we have other evidence to support this? Are there other post-marketing reports or real-world evidence to support this? I had a slide on real-world evidence but I was afraid that I wouldn't make my time, and I'm right at the top of the hour now. But I'll leave you with this. I just pulled this, this is Dr. Glasziou, he is sort of a stern-looking Brit from the University of Oxford, evidence-based medicine guy. He had a nice editorial and had some comments here. He said, "The search engine is now as essential as the stethoscope," and he went on to say, "A 21st century clinician who can't critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system." That's probably going to

an extreme, but I thought maybe I could send home the message there. I've gone over by one minute I want to go ahead and stop right there.