ALASKA HEALTH CARE COMMISSION

DENA'INA CIVIC AND CONVENTION CENTER

600 WEST SEVENTH AVENUE

ANCHORAGE, ALASKA

WEDNESDAY, AUGUST 21, 2013

8:00 A.M.

VOLUME 1 OF 2

PAGES 1 THROUGH 312

P R O C E E D I N G S

8:00

(On record)

CHAIR HURLBURT: There are some others coming in, but I appreciate everybody being here. I think you'll find this to be a very interesting and exciting day, opening up a new concept, for some of you hearing more about the concept, for many of you here. For those of you who haven't had a lot of contact, the Alaska Health Care Commission was initially -- at least, the current iteration of it -- started by Governor Palin as Governor, what she could do for about a year, and looking at issues of cost and access and quality and affordability and so on. And then after that year, Governor Parnell and the Legislature established the Health Care Commission, looking at those issues for Alaska, and we've been meeting and looking at a number of issues.

One of the things we've talked about is evidence-based medicine, and in one of the very first meetings -- actually -- well, Jay Butler, who is on jury duty this morning, I guess. He's not here this morning, but while Jay was chairing it and before I came into the job, he had somebody up from North Carolina and estimated, basically, in the neighborhood a third of what we do in healthcare is not really supported by evidence, and much of it does no good. Some of it does harm. And the concept of evidence-based medicine helps us because,

often times, we might all agree on that's probably true, but we don't know which third is the third that either does no good or does harm, and the concepts of evidence-based medicine lead us to identify that to improve the quality, to improve the safety of what we do. And Mike and Sheri won't talk about it today because that's not a part of their charge, but from what I understand, we would not have a Health Care Commission if it had not been for issues of cost and consuming more and more of our national gross domestic product there. And truly, if we can identify the areas where, in healthcare, we really do good, that is what we want to do.

So I will come back to this a little bit later. Again, I'd like to welcome everybody here. This is an unusual structure for us. Normally, we have the members of the Health Care Commission around the table and the others in chairs in the back. There is always a public testimony time. They always are public meetings, but most of the communication is just among the Commissioners or the Commissioners and whoever has come to present to us, but we wanted to have this opened more broadly. There are a large number of physicians in the group, but not the majority. The minority of us are not physicians. The presentation will be geared -- there will be things -- and I think those of us who are physicians will tune in and pick up on some things, just based on spending much of our lives in the medical literature and so, but the

presentation really will be geared for all of us here who have an interest in healthcare and seeing that things are done right.

Now, I would like to start, and our time will be limited, and we have a busy program, but I want to start and just mention a couple of things. Usually, I start our Commission meetings with some comments.

For some of this -- some of you, particularly for the Commission members, this will be old hat, but I've developed a set of elevator slides that I've shared with guests, Representative Keller, and with some other (indiscernible - recording interference), but this is, why are we here? Why are we talking about this?

Well, I don't have them to present, but you can see this is a graph that goes off a lot. It starts way down here, ends up here. This is how much we spend, per capita, in the U.S. on healthcare, and it goes from $147 a year, and it doesn't quite go up to 2013, but now, we're over $9,000 a year. Well, I have gray hair. This was after I graduated from medical school. I graduated in 1959, less than $150 a year. Now, we're over $9,000. So that's an attention grabber.

So it's happening to everybody. Well, it's really not. This graph shows us compared to other countries in the world. We were the highest back here about 1980, but the United States has continued to outstrip other countries. That's

hurting our economic competitiveness.

Over the decade from 1999 to 2009, the average real earnings for a middle-class, middle income family in the U.S. went from $76,000 a year to $99,000 a year. All of that was consumed by increased healthcare costs. So you didn't move from a 15-year old (indiscernible - voice lowered) to a ten-year (indiscernible - voice lowered).

These are federal expenditures. We talk about entitlement. Social Security flat-lined. Those are a very flat line. Other federal programs, flat line, flat line. Healthcare is a red line going up.

A new chart, to me, was state expenditures. The same thing. The blue line shows -- and this is a projection of the next 50 years, as was the last slide -- all state expenditures, except healthcare, going down over the next 50 years, healthcare for state employees, Medicaid, and so on going up. So that's an issue.

We've seen a little bit of leveling off -- and Allen Hooper is not here yet; he'll be coming in and joining us. But we've seen a little leveling off. We're still increasing healthcare costs more than the Consumer Price Index, but this, again -- I know you can't see these -- shows the membership from 1988 and HMOs, PPOs, Point of Service Plans, and the new little bit up here is Consumer-Directed Health Plans. I mention that because we say, why is increased rate of cost

going up, more than GDP, more than CPI, but less than it has, and various reasons are suggested, but this could be it. The consumer has more skin in the game, and a part of that has to do with evidence-based medicine because the role of the provider now, the physician, the nurse practitioner or others, is being the expert, having the technical skills, but also being the educator. Whereas in the past, in my early career, if I had to said to a patient, I'm sorry, we have to amputate your arm, the answer might have been, oh, thank you very much, Doctor. No way. Not today. The physician is an educator, and you talk with the patient (indiscernible - background noise). You get a PSA test for prostates or you have surgery or you do whatever, and you make the decision jointly.

Back with the Consumer-Directed Health Plan, more of that is there and that's a little bit of the background that leads us to this, but quality and appropriateness are very much tied with cost.

Now, if we could just maybe start and just give your name and who you represent because there are a large number of people in the room, and we'll start with everybody at the tables and then go to the back of the room. Representative Keller, we'll start with you.

COMMISSIONER KELLER: Representative Keller, Alaska State House, representing the House.

COMMISSIONER ERICKSON: Ward, can you ask folks to use

the mics, so we can hear it?

COMMISSIONER KELLER: Oh, I'm Representative Wes Keller, and I'm from the House, supported by the House.

COMMISSIONER CAMPBELL: Keith Campbell. I'm the consumer rep on the Commission.

COMMISSIONER URATA: Bob Urata, family physician representing primary care.

SENATOR COGHILL: John Coghill, Senator for Alaska, for the Interior part of the Alaska.

COMMISSIONER PUCKETT: Jim Puckett. I'm on the Commission for the Office of the Governor.

COMMISSIONER MORGAN: Dave Morgan representing community health centers for Alaska.

COMMISSIONER HARRELL: Tom Harrell. I'm a cardiologist. I'm the DOD/VA representative.

MR. KNAPP: I'm Gunnar Knapp. I'm the Director of the Institute of Social and Economic Research at UAA.

MR. MONAGLE: Mike Monagle. I'm the Director of Division of Worker's Comp with the Department of Labor and Workforce Development.

MS. BRODIE: Margaret Brodie. I'm the Director of Healthcare Services.

MS. HARRIS: Jan Harris, Vice Provost for Health Programs for the University of Alaska.

MS. COTTER: Emily Cotter. I'm with the State of Alaska,

Department of Administration and Division of Retirement and Benefits.

DR. GRANHOLM: Marin Granholm, family doc here with the Alaska Academy of Family Physicians.

MR. HERMAN: Bill Herman with the Mental Health Trust Authority.

MS. WRIGHT: Peggysue Wright with the Section of Public Health Nursing.

MS. BERNER: Barb Berner, Director of the School of Nursing at UAA.

MS. RIPLEY: Elizabeth Ripley, Executive Director of the MatSu Health Foundation.

MS. LOVE: Jenny Love. I'm the Medical Director for API.

DEAN HOGAN: I'm Bill Hogan. I'm currently the Dean of College of Health at UAA.

MR. CAPISTRANT: Todd Capistrant, Medical Director, Tanana Valley Clinic, Fairbanks.

MS. FENAUGHTY: Andrea Fenaughty, Deputy Section Chief, Chronic Disease Prevention/Health Promotion.

MR. HOPE: Chad Hope with Division of Health Care Services.

DR. BROWNSBERGER: Keith Brownsberger representing internal medicine.

DR. KIESSLING: Bruce Kiessling, family physician, Primary Care Associates.

MS. HUDSON: Laura Hudson, Executive Director of Alaska Physicians and Surgeons.

MS. COX: Sherri Cox, Clinical Manager at Cross Road Medical Center, a community health center.

MS. SMALL: Alanna Small, Alaska Native Medical Center, Inpatient Medical Director.

MS. TSIGONIS: Jean Tsigonis, family practice, Tanana Valley Clinic, Fairbanks.

MS. PERDUE: Karen Perdue, President and CEO of ASHNHA, the hospital association.

MADAM COURT REPORTER: You all need to push the button; otherwise, we can’t hear you.

MS. PERDUE: Karen Perdue, President and CEO of ASHNHA, the hospital association.

MADAM COURT REPORTER: Thank you.

MS. EDMISTON: Gena Edmiston, Chief Nursing Officer at Fairbanks Memorial Hospital.

CHAIR HURLBURT: Could we go to the folks in back?

MADAM COURT REPORTER: If you could go to a microphone, that would better.

UNIDENTIFIED FEMALE: (Indiscernible - away from mic)

MR. MCCLUNG: Peter McClung, Health Economist with McDowell Group.

MS. ORELL: I’m Laurie Orell with McDowell Group.

MS. RYAN: Debbie Ryan, Alaska Chiropractic Society.

MS. OGAN: Janet Ogan with Representative Keller's office.

CHAIR HURLBURT: Thank you, and again, welcome. We'll have some others coming in. We have two more of the Commissioners who have come in. If you could just introduce yourselves? Val?

COMMISSIONER DAVIDSON: Good morning, (indiscernible - speaking Native tongue), Valerie Davidson, Alaska Native Tribal Health Consortium. I represent tribal health, and God bless all the parents who managed to get their kids out on the first day of school.

COMMISSIONER ENNIS: Good morning, Emily Ennis. I represent the Alaska Mental Health Trust Authority on the Commission. Thank you.

CHAIR HURLBURT: I'd like to just maybe make one comment. Dean Hogan, who also served as Chair of the Health Care Commission for, I guess, one meeting, Bill, probably -- was it?

DEAN HOGAN: Just one.

MR. LEE: It was memorable though.

CHAIR HURLBURT: It was during the time that Dave was here, and I came, and I'd like to welcome you here. I had a message this morning. I guess Alaska Airlines had a mechanical problem. Commissioner Becky Hultberg was going to be joining us, but since she wasn't going to get out of Juneau

until late in the morning, she will not be here with us.

Now, I just want to say a little bit, as we move into our meeting. Part of my own background, one of the times we left Alaska, I went down and joined Group Health Cooperative, Group Health Permanente in Seattle, in partnership with a non-physician administrator and led the non-primary care parts of the program there, and one of my duties was to co-chair with Dr. Mike Stuart, who is here with Sherri Strite this morning, our Technology Evaluation Committee, and that was a time of great learning for me and that was back in 1993 when I went there, but when my eyes really first started to be opened to evidence-based medicine.

Evidence-based medicine is a fairly new concept. Some of the pioneers -- there was -- and it does sound like it's an English-speaking world, but there have been others elsewhere. It's probably just easier for us to read the English literature. There was an Archie Cochrane in 1972 who published *Effectiveness and Efficiency*, and he is the one that the *Cochrane Collaborative* is named after and the *Cochrane Collection*, which is a recognized international source of analyses there.

There are a couple of Canadians, David Sackett and Gordon Guyett of McMaster University, David Eddy, one of the early names in the U.S. who was at Duke, later became Vice President at Kaiser for that large eight million member system on the

evidence-based issues and quality of care. He now heads his own consulting firm. Anna Gordon lives in Australia, who founded the British Medical Journal's Clinical Evidence systems that they do. And I have to mention John Wennberg, a young physician who wrote the book *Tracking Medicine* that showed the differences in care, but really, initially, assuming where there was less care and may have been less resources, less docs, less whatever, and then found that much care was not really evidence-based -- and those were some of the early pioneers.

There is a slide that I use that I read, which you're not supposed to read with slides, but I think it's important, and this was from David Eddy, and this was from an article in Health Affairs in 2005. And this was the quote:

> "Up until about 40 years ago, medical decisions were doing very well on their own or so people thought. The complacency was based on a fundamental assumption that, through the rigors of medical education, followed by continuing education, journals, individual experiences and exposure to colleagues, each physician always thought the right thoughts and did the right things. The idea was that, when a physician faced a patient by some fundamentally human process called the Art of Medicine or clinical judgment, the physician would synthesize all of the important information about the patient, relevant

research and experiences with previous patients to determine the best course of action."

In some ways, I describe knowledge as being like a lake, and clearly, in our lifetimes, all of ours here, we've had an explosion of knowledge in many, many fields, and very notably, in the medical field, but the unknown are like the shores of the lake, and the bigger the lake gets the larger the shores are, the more shores you have and the more we realize that there are still lots of unknowns there. But the knowns, because they have grown so much, it becomes harder and harder for any human being to synthesize and so we do need to look at that. We do need to help that.

But I mentioned that I got to know Mike and Sheri when I was with Group Health there. I think Mike appreciated my co-chairing the Technology Evaluation Committee with him because I had some administrative clout there, but Mike was my guru, and Mike was the one that I really learned the concepts from, and we had others there. We had an epidemiologist, who helped do the analysis for us. We had other members of the medical staff participating.

And in my bias, Group Health in Seattle, our neighbor to the south, has really, probably has done the best job in the country, particularly in primary care areas, in looking at what does the evidence really support, and one of the things they will talk about in their presentation, for example, is

Vioxx. It was a drug that came out and was felt to be a new, wonderful drug for some autoimmune conditions, like an NSAID, that has less gastrointestinal complications, but was pulled from the market because of the increased incidents of cardiovascular disease when you took that and ended up costing Merck hundreds of millions of dollars in settlements. That information and that data was there right from the beginning. It was there, available, and the FDA approved the drug, and because of the technology evaluation process, Group Health never put Vioxx on their formulary.

Now, it doesn’t mean that it was a totally bad drug, and with more selective usage for some local patients, it might have been appropriate, but after it was FDA-approved and got on the market, when you turned the television, you would have thought (indiscernible - voice lowered) created mankind with a vast efficiency and that it would solve many of our concerns. So that’s one of the kinds of things that we’ll do.

I’m not going to take a lot of time to introduce Mike and Sheri because they want to introduce themselves. They are experts in a new field, a word that’s new to me -- you might remember it, if you’re Scrabble players -- evidology, but it’s how do you look at the evidence, how do you look an article, how do you look at the charts, why should you look at it. And so for those who are in the field, physicians, the physicians here, I think that will be very useful to you either as

administrators or as active clinicians or both (indiscernible - voice lowered) you do for non-clinicians. It will give you the concept, but just because it's black and white doesn't mean that it's ordained and that's what it is.

The *New England Journal of Medicine* just not too long ago, I remember, had an article of the percentage of the articles in the medical literature that are supported by pharmaceutical industry, and it's not an evil industry, and they're not evil people there, but they are in the business of making money and pushing the products, and it was the majority. It was slightly more than half where there was that involvement there. So you need to look at conflicts of interest.

Without going further into what you've done specifically -- you can do that yourself -- I'd like to introduce Sheri Strite and Mike Stuart, who (indiscernible - background noise).

MR. STUART: Thanks, Ward. We are extremely happy to be here, and I'll say a little bit about us, and Sheri can say a little bit about us, but our goal today is to just have like a communication with you, a conversation, very informal, if we can make it that way, about how we deliver the best care for patients and that's the starting point for us, the patient, and I think it is for everybody.

In fact, the decisions that are made in healthcare are

made not only by patients, but frankly, they're made by doctors, by payers. They're made by healthcare professionals in all walks of healthcare life, and there seems to be, even though it's an overused word, a paradigm currently, and hopefully, we can change this a bit, but the paradigm now is what David Eddy said and Ward told you is, there is some sort of belief that there is a black box, that, if you are a doctor, then you will come up with the right answers for your patients, but if you ask them, how do they do that, and as David Eddy, it's very difficult to explain how that would work.

Now, the problem with that paradigm is that patients are not getting the information they need very frequently, maybe (indiscernible - voice lowered) to make decisions that are in their best interest, and why is that? Well, because it's a little more complex than going through some school and then your experience and then putting it all into your head and thinking hard, as David Eddy would say, and then coming up with the right thing to do.

The new paradigm is more like this. Patients are entitled to have the options -- and we're going to focus on treatment and therapies today because we don't have time to do more than that, but diagnostic modalities have very similar considerations in other parts of decision making, but we're focusing mostly on treatments.

So patients really should have all the options in front of them for whatever their condition is. They should know what their condition is. They should know the potential benefits and the potential risks of each option that they are offered, and they should understand these things before they are put in a position of making a choice.

We were in Oklahoma a few years ago. We went there every year, as a matter of fact, but the Medical Director's mom developed cancer, and this is just one story. The Medical Director is a very evidence-based guy and so Tim took his mom, went with her to the oncologist, and the oncologist says, well, this is the type of tumor you have, and this is how we usually treat it, and we'd like to get started next week. And that was about all she was told.

And so he knew, the oncologist knew that Tim was a physician and also that he did evidence-based medicine work -- and that's actually why we went there every year -- but he went home, and he quickly did a quick search and looked at all the evidence on his mother's tumor and found not a shred of evidence that what was being offered had benefits that outweigh harms. In fact, the harms were clearly -- both to his mom and to Tim -- outweighing the benefits. So they rejected the therapy and went for symptomatic care, based on what the evidence told them. But they were never given the opportunity to actually provide informed consent, so to speak.

And it’s because, I believe, there are so many little problems with our system. You have to actually start with what question you’re trying to answer. You have to get the evidence. Then you have to evaluate, or what we call, critically appraise that evidence. Then you have to summarize it or make recommendations from it or lay out the options in terms of benefits and risks so people can make choices. Without that information, none of us, as patients, are going to be able to make the choices that are best for us, and this has been shown repeatedly that, when -- and Cochrane, as Ward pointed out, is a very reputable group that -- these folks that work for Cochrane are mostly volunteers, really, and they really want to get the evidence in front of people so that people can make right decisions, but in fact, it’s not happening very well because we haven’t moved to this new paradigm where we have a formal way of posing a question, reaching out to the world of information, bringing it in, evaluating it, getting rid of the stuff that doesn’t help us and then using the rest of it to layout the options that include the benefit and risk for whatever the condition the patient is being treated.

So we’re going to be talking about different aspects of that process for that newer paradigm, and we’ll be contrasting it, through lots of stories and cases, with what’s currently going on and showing why this is so important. So we will be

talking about the whys, the how's, and really, this is not an extremely difficult set of competencies to learn. So we will get started.

My background is I practiced as a primary care doc at Group Health in Seattle for 30 years, and about half that time, I had a department that sort of evolved into this new paradigm. I had a lot of help, but started out as a Continuing Medical Education Office, and I was the CME Director, but I changed that, and I hired some epidemiologists, and we learned together how to critically appraise the evidence. That was the first thing we had to learn, and there weren't good books. I mean, we read the epidemiology books, but they're not sufficient.

Epidemiology is the study of diseases and conditions in populations. Critical appraisal of the medical literature is quite different. You use many of the same concepts, but what you're looking at is you have now an article or you have a systematic review or you have a guideline or you have a monograph, and you have to take that apart, and the biggest problem right now is that people look at the bottom line because of their time, and we understand the time, but if you look at the results of a study or a review and you don't go back upstream and look at whether the information is reliable and clinically meaningful to a patient, if you don't do that step and you just look at the results of the study, you're

going to make some decisions that are not in the best interests of the patient, and ultimately, it's going to be the patients who are going to have to take some responsibility for participating this, and some do, even now. In fact, we've found, at Group Health, that I would go on a Saturday morning and all the Seniors would show up because the Seniors were very interested in anything that we would put on, and they could learn this. So we learned very quickly that Seniors can learn evidence-based medicine just as well as doctors can. So we really came to the point where we would instruct anybody. We actually instructed architects in Seattle to do evidence-based architecture. It was very, very fun and very, very different. We've also trained administrative law judges in California in this and that was also very, very interesting, but I'm saying anybody can learn this. It's not really that tough. It's just a matter of having a systematic approach.

So basically, I chaired the P&T Committee. I chaired the Technology Assessment Committee with Ward. Those are the two big committees in most organizations, but there are others. We did guidelines. We did pathways. We did evidence-based QI projects, quality improvement projects.

And Sheri was in the Research Department, and I finally recruited her out of that department, and she came to work for me, and she's been working with me, off and on, since '96. She went to the University of California San Diego where she

taught medical students and pharmacy students and others. And I'll let you tell a little bit about your background. Then we'll get going.

MS. STRITE: Well, actually before I do that, since I'm the least interesting member of the group here, let me just do a few housekeeping things and then I'll give you a few more pieces of my background.

But first of all, I want to mention that we will accept money from anybody who wants to pay us. The thing is, we do the same work for them that we do for anybody. So the last 12 months, we've actually had some wonderful experiences working with Amgen, Estelles (ph), and Genentech, and it's all been around evaluating evidence, which has been very interesting and fun.

And then we are members of the Editorial Board for DynaMed, through EBSCO Publishing, which is nonprofit -- I'm sorry, which is for profit. We are unpaid by that, but then we also are about to do a publishing relationship with them, so I'm excited about that. So that's a little housekeeping.

And then about us, this is the real deal here. This is my real background. Delfini stands for dolphins, and there are various reasons why we picked that as our symbol, but one of the things that we wanted to do was to have something represent the fact that we are fairly playful. And so even though today is about an extremely serious subject, we think

that we should all enjoy life and that learning is best when we do have some fun. We have some very interesting stories, again a lot that's very serious. We are very informal and so you can interrupt us and interact with us. Mike and I do not consider ourselves to be speakers, but we consider ourselves to be more trainers and facilitators. And so today is not a canned program, by any means. We constructed this specifically with you guys in mind and so this is your day. It also means that we really don't have a sense of timing because it depends on how engaged people will be with us. We have some exercises we might be able to get to. So we may work you, if you don't talk with us during the time that we're presenting. And they're fun, too. They're not torture. And we kind of make up things as we go. So occasionally, you'll see us kind of stop and consult with each other because we're trying to figure out how to do the best thing, just in time for you guys. So that's why we're here.

And what else do I want to mention about that? Here's, formally, what we have as our objectives to cover: why evidence-based medicine and critical appraisal are so important for patient care. And then we're going to take you on a little journey to give you some concepts about critical appraisal and what that means, and the big message that we really want to give people is that you see a lot of people using evidence or saying that something is evidence-based, but

what is missing from that is it's very frequent that people are not evaluating the quality of that information. They're not evaluating the quality of the evidence and that's what critical appraisal is about. So it's not simply using something that got published in the *New England Journal of Medicine*, but it's evaluating whether or not that science is likely to be helpful to us or not. And we'll tell you some facts about that along the way.

MR. STUART: Question?

MS. STRITE: Yes?

UNIDENTIFIED MALE: Can I just -- I hate to interrupt, but I.....

MS. STRITE: Sure.

UNIDENTIFIED MALE: It just would be very helpful if you would briefly address this. We send young people to medical school. They train for four years in how to be doctors and keep people healthy and fix them. Why -- can you briefly address why evidence-based medicine doesn't saturate the medical curriculum?

MS. STRITE: Absolutely. Yep. Yep. In fact.....

UNIDENTIFIED MALE: And you don't have to tell me now, but I'd be interested in understanding that.

MS. STRITE: No, because I think that's perfect, and I.....

MR. STUART: Good question.

MS. STRITE: .....think it's really important, and you know, Mike and I have spent a lot of time in this work, and I think that that's really the key is that evidence-based medicine -- and evidence-based medicine has gotten, sometimes, some negative press. We can just even say critical appraisal or (indiscernible - voice lowered) of medical science to kind of get the politics out of it, but science in medicine is actually fairly new, and I know that there are other healthcare professions, but I'm really going to kind of focus, especially during today, on physicians, partly because I worked in the medical school and have worked with a lot of doctors. But medicine generally is trained as an apprenticeship, and if the person who is your guide doesn't understand how to evaluate or the need to evaluate medical science, that's not going to get passed down. And medical schools are very political in terms of the real estate of the student, you know, that territory. And so unless you've got a structure in medical school that has embedded this and embedded it well and unless you have faculty that understand this and are utilizing it and know how to put it into practice, it is not getting down to the students is what's happening.

I was hired at UCSD to teach there, I think, in 2002. I am the first person that I'm aware of that was formally hired to teach evidence-based medicine in that medical school,

which, you know, has a great reputation. I got six hours with the fourth year medical students, and most of those students knew nothing about what I was teaching, and I was teaching.

So Mike and I travel all around the country, and we see this everywhere. We don't spend as much time with nurses, so I'm not going to focus on nurses so much, but we work a lot with clinical pharmacists, and it's the same thing. We've done a lot of precepting for, I think, about three schools in southern California that are pharmacy schools, and these people are coming to us with very, very little information about this, and thing is, I think that one of the big problems is not just that science is fairly new in medicine, how do you get it in there, how do you teach, how do you put it into practice, but I think another big issue there is not a good understanding that most of the information that gets published is actually not reliable information.

So Mike and I can actually find that we use about ten percent of the studies that we evaluate, and we're not even necessarily the toughest graders. One of the statistics I'll show -- so I'm going to do the numbers kind of from memory, but I think it was -- don't test me on this -- 60,327, I think, articles were reviewed by McMaster, and they found that only seven percent of those passed criteria for high quality and clinical relevance. So we've got a huge information problem.

We were talking about Cochrane, for example. Cochrane is considered to be one of the best places in the world for go-to for evidence-based information, and yet, there is huge variability there, and even one of the Cochrane groups, the Nordic Group, which is considered, I think, to be one of the better groups or maybe the best, evaluated the other Cochrane groups and the quality of their information and showed huge variability there. So there is a lack of awareness about this problem.

A lot of people -- you know, physicians will come up to us and say, well, you know what, here's what I'm going to do. Rather than learn critical appraisal techniques, I'm going to save myself time, and I'm only going to look at the *New England Journal of Medicine* and my other favorite specialty journal, and I'm going to avoid all industry studies. Well, you can't get around the problem that way. If I were to take the *New England Journal of Medicine*, whatever issue is out right now, my guess is that, probably, I could find that I could only rely on about ten percent of the information in there.

Do you want to answer that? I mean, this is -- what you're posing is, like, the heart of the problem.

MR. STUART: Well, just briefly, historically -- and I'm just speaking about medical schools because that's what you asked about and that's -- I know more about that, but also I

know some about pharmacy schools, but essentially in medical school, you're -- you have plenty to do learning anatomy, physiology, pathophysiology, and the various specialties compete for the student time, but none of them incorporate, as part of their piece of the pie, how to read the literature in that particular area. So if I'm now on respiratory medicine, I should be reading the current literature and critically appraising it for bias to make sure that the results are trustable because that's what, really, all critical appraisal is, but that's usually not part of the training in medical school.

So it doesn't get done. Well, there are exceptions, of course, and there are residencies that do and medical schools that do a fair job. I say fair job.

MS. STRITE: And it's rare.

MR. STUART: And it's not -- we don't see it very often. So essentially, it's not like many other things. We don't have complete agreement on much in critical appraisal. We have people that will agree on concepts, but in details, we will vary a little bit so that we don't have, you know -- we don't have a society. We don't have an American Academy of Evidology. You know, we don't have the structural things to make this happen in a big way yet, and it's just slowly -- the snowball is just slowly starting. So we have a ways to go. It's just new, and it's being put together because we have a

new type of evidence that comes out that says, if you don't blind in a study, you can distort the results of that study by up to 70%. So it could be wrong by 70% or even more, but it could be less. So we don't know exactly, in one study, how much the results will be effected by lack of blinding. We know it's going to be a significant amount. So that's part of the answer.

I think we should move on, but the idea that I want to leave you with is that it's not that the medical schools don't want good doctors. It's that the faculty and all of us are just not to the place yet where we can efficiently get this into medical students' competencies yet. So I'll have you move on.

MS. STRITE: So we're going to do some little snippets of critical appraisal. So we're going to give you some idea of some of the concepts and how they relate to each other. We're going to talk about the requirements for an evidence-based approach, what patients need for patient-centered decision making, and then what policymakers need for an evidence-based approach and that's going to be a point at which we're very much hoping that there is a lot of discussion with you all.

Mike's introduced us fairly nicely. I'm just going to mention, I guess, a couple more things and that is we consider ourselves to be evidence-based Clinical Improvement Experts. And so that really runs the gambit all the way from how

organizations set themselves up to do an evidence-based approach and clinical improvement activities and then all the way to physician-patient communications, but people know of us and this day is about critical appraisal because that's where we've concentrated our time because that's the area that we see as having this huge need, and it's very neglected. So that's where we spend most of our life's work.

Let's see. What else do I want to say about that? Mike is very humble. He actually did some incredible work at Group Health where he mentioned that he had been the head of the Medical Education Department, and I just want to, I guess, flatter him and compliment him a little bit because what he did in the early '90s was react to a frustration that he had where, as a physician, he would call a sub-specialist on duty and say, well, what should I do for this patient who has whatever, and he said, I'd get different answers frequently. And he said, so I didn't know if that meant they were all true or this person is right, this person is wrong, they're all wrong. You know, what's that all about?

And Ward had mentioned Dr. David Eddy. He wrote a series in JAMA that is just beautifully written, the *Clinical Decision Making Essays*, and that was put together in a book, and Mike read that book, and he said that's it. This is the answer. I understand now what the missing piece and that is that evidence needs to be evaluated to determine if it's

reliable or not. And so what he did is he turned his little tiny department into an evidence-based machine, and when I met you in ‘96 -- let’s see; I think I left, like, five years later -- Mike’s department, through different specialties in Group Health, had produced over 35 clinical practice guidelines, and the U.S. Bureau of Navy -- sorry, U.S. Bureau of Medicine of the Navy licensed the guidelines. The New Zealand Health Ministry had come to Mike to learn how to do this work. And so he really deserves, I think, recognition for a lot of great stuff he did, including putting Group Health on the map for quality.

Probably one of the most important things to know about me during this day is that I have no clinical training whatsoever. I actually have a Bachelor’s in English Literature. So when doctors come up to me and say, I can’t do critical appraisal, it’s too hard, I say, well, no, because you’re a doctor and all I did was read. So it was poetry that was done, you know, back even before English was codified. So the important thing to know about my lack of clinical training is that, sometimes, I’ll be kind of impressionistic with some of the clinical stuff. So everybody gets to give me a little bit of room, but the concepts behind that are what I want you to listen for because that’s stuff I actually do know.

One of the reasons, too, why we ended up with dolphins kind of as our symbol is that I was at UCSD and Mike called me

up, and he said, you know, I'm going to retire from Group Health, and how about if we do this work together? And then neither of us knew anything about starting a business, but I said, I think we need a name, so that would be a good thing. And I said, so give me an idea of kind of what you want us to do, and he said, well, I really want to help people navigate out of chaos. And so we loved the idea of the dolphins who helped the shipwrecked sailors and so that's another reason why we chose that name, but we think that there is a huge problem in healthcare here, and we are really excited to be here and glad that you are here because we see a lot of working with the evidence is a lot of the solution for improvement.

Mike and I came from sort of a public health orientation. Group Health Cooperative truly was a consumer-led cooperative. And so we are a little business. We're it. This is our company right here, but we consider ourselves a public service entrepreneurship and so we have spent our lives, for the last ten years, creating lots of tools and lots of materials that we give away for free on our website to help people. And I'm going to show you some things that you have in your binder, but I'm just going to mention there is a whole lot of other information up there that's useful and freely available, and I'm happy to spend -- it would only take maybe ten to 15 minutes to give people even a web tour on the phone, if there

is something you're interested in. And I'm going to show you just a couple things that we have here.

So you can always contact us at the envelope there, and our website name is on all of your materials. There is a selection here on our main menu that's resources. That's the place to go where we've got all kinds of stuff. And then I've got a little breakdown and a little submenu that's just below that, where we've got a lot of tools to help do not only critical appraisal but clinical improvement work.

We have a very important document that is in your binder that we're going to talk about, the Health Care Information Problem, which we've discussed some, thanks to your great question. We've got some tutorials. We write some commentary on the evidence. We've got some patient decision support examples. We're going to show you our Evidence Messaging scripts, which are ways to impart quantitative information to providers and patients. We have medical leader interviews, including one with Ward that is very wonderful.

We try to simplify the concepts. So we try to put things in natural language and so we've got a glossary for evidence-based clinical improvement work that a lot of people find helpful because it puts things in natural language. Recommended reading. If we mention any websites, we have them on our web links page, so you don't have to write those down. And then if people are interested, I'll be passing around a

sheet if you're interested in signing up for a short email from me about four to six times a year that just says what's new on our website, but we also have a button that says "What's New" so you can go on any time, but I'll be circulating that.

And then I'm going to do some definitional stuff here before we get started, really. We're going to do as Mike said. We're going to tell you some stories. We're going to do some cases. We're going to do a number of things along the way, but I want to make sure that we frame a few things because Mike and I start using terms -- like Mike used "bias" a few minutes ago -- and we want to define some of these things.

We also want to give you an orientation to your handout. This is differently set up than how we usually do this and so you have a couple of different places to go. Normally, we do not have slides as part of the binder. Instead, what we have is we have a lot of notes. We have -- up here, it says "Summaries," and those summaries are one-page documents, sometimes two, that cover different topics. And so the first one here -- and then the page numbers are kind of squished over here to the far, bottom-right with a little underscore because these are actually taken off of discreet documents that have their own page numbers, too. So the page numbers I'm going to be referring to will be these.

So the first one here on page three, for example, this is a framework that gives a lot of information about steps to do quality improvement activities. Page four, this one is extremely important, and I'm going to just read some highlights from this. Page four, this is really getting at kind of the heart of some of the problems with what we see in medical information. "As of December 2012, at least, 37 deaths were linked to fungal meningitis" -- sorry, I can't say this today.

MR. STUART: Meningitis.

MS. STRITE: "Meningitis, thought to be caused by contaminated epidural steroids." And here, we had a situation in which we didn't have good evidence that people would be helped by this treatment, and they end up in a risk situation where many people lost their lives, and in all likelihood, would not have been informed that, in fact, there is not good evidence to support this treatment. So again, our big point is, for patients to truly have informed consent, they need critically appraised information that is likely to actually benefit and help them in some way.

We're going to talk about Vioxx more, so I'm not going to cover that. Number five, "Leading experts estimate that 20% to 50% of all healthcare in the United States is inappropriate." We've talked about training in medical schools, number six. And we have a pre-test that we often use

for our training programs that has only three questions on it, and two of them are extremely basic, and I'll go over those later on, but roughly, 70% of clinicians fail our basic pre-test. So this is huge.

We have 13,000 articles that are published in the National Library of Medicine every fill-in-the-blank -- 13,000 every what? Go ahead. Week. That's a lot of information, and we're telling you that, of the clinical trials that we review of the randomized control trials, Mike and I find that we can only use about 10% because of trustability and clinical relevance.

So anyway, this information -- there is more information here, but this gives you a picture of some things about the problem, the magnitude of the problem and why the problem is. So that's a very important page.

And then we have some other summaries. One of the most important ones is here on page five, and what this is, this is a window to critical appraisal done in kind of a narrative form. So we put all of the key concepts to do critical appraisal for primary studies, which means original research, on this one piece of paper. And so when we do our critical appraisal sampler, this is where you have the notes. So you will have the slides in the back, but we've taken notes for you in these nice summaries for a lot of different topics.

One other page that I'm going to mention that is one of

our very most important pages is in, let's see, the tools section on page 25. Page 25 is extremely similar to page five. It's just a bit of a different flavor. Page 25 is kind of our core critical appraisal checklist. So page five is more instructional for people who are newer to this work, and page 25 is sort of the more professional version of that. There are a few more items on here. It's very condensed, and whenever I read a medical study, I always have this piece of paper with me, and the way that I use this paper is, first of all, I've learned the basic concepts of critical appraisal. So it's like learning the language, only it's much easier. We could fit it on one piece of paper. There are no verbs to conjugate. There are no articles.

So you get that in your head, because docs asked me, it's like, how do you do this fast? You just get those concepts in your head so it's kind of natural, and it's a bit of a rewiring. It's a bit of a paradigm shift. And then you make sure you understand the concepts on page 25, and online, we have an accompanying tool to this. It's about 14 pages. So it's a little, mini-EBM text book, and it explicates each of these items. And so if you don't remember or don't know, you can look it up in there. And again, this tool -- even though I'm experienced in doing this work, I use it still every time I read a study because you will forget what's not even mentioned, right? And so sometimes, that will be the thing

that's one of the biggest problems with the article, and no one knows what this means yet maybe, except a few of you, but they didn't tell you how they censored and so this reminds you what that is. What "censored" means, Mike is going to talk about that, but quickly, it's removing patients from an analysis. It can be a big problem.

So those are, I think, some of the key things to show you about the first part of this. And then after -- let's see. So the glossary starts on page 62. So you can look up things, if you need to, as you go. The glossary that we have online is much larger. So this one is just kind of constructive, sort of short for these programs. And then you have a beautiful piece of paper that's kind of marbled blue right after the glossary and that's where the slides start, and occasionally -- so this is what's going to be a little bit of a challenge, since you have both, is, occasionally, we're going to ask you to toggle back to the front of the book. Probably, you're going to spend more time here in the slides, but just be aware that that's how you'll be navigating.

We're also going to toss a few hypothetical cases for you to chat about. And now, I'm going to give you some definitions. There are a number of definitions of evidence-based medicine, but we came up with ours when I was teaching one of my fourth-year seminars, and I'll tell you, fourth-year seminar in June when you're first teaching somebody this is

really painful. Oh, my goodness. They want to kill you. It was kind of like -- wasn't it the Harvard -- one of the Harvard leaders who addressed the students as they were leaving medical school, as they were graduating and said 50% of everything we've taught you is wrong; we just don't know which half? So it relates to your 30% there.

And I had a student who came in late. She was very beautiful. She had gorgeous, flowing, long blonde hair, and she was tall, and here, I'm kind of squat. And there were these five men, and they were very interested in what we were talking about. We were having a great discussion, and she blows in, and she throws her books down on the table. So we're all silent at that moment. She sits down, throws her hair back, and she says, I just want you to know I have a real problem with evidence-based medicine. And I'm standing there, and now, all the guys are looking at me, like, okay; what is she going to do? And it's, like, they were my friends, and it's, like, now they're turning on me because it's, like, cat fight. Cool. Let's watch this.

So anyway, I looked at her, and I'm thinking as fast as I can -- and it seems so obvious now, but at the time, it was not -- and I just said quietly, tell me what problem you have with science and that just leveled here. The guys all started laughing. She said not another word. But I called Mike up after that, and I said, you know, all the times -- and again,

we go back with EBM a while -- that we've had people, you know, talk about evidence-based medicine, you know, cookbook medicine and all of this, I said, I found the answer; it's simply about science. And so we say that, to us, EBM is the use of scientific method and application of valid and useful science to inform healthcare provision, practice, evaluation, and decisions.

And then Mike talked about critical appraisal, so I'm just going to go into that a little bit more deeply. It's a scientific evaluation of evidence -- meaning the research data, so the publications from the research -- to appraise for validity. And when we talk about validity, we're talking about closeness to truth, and we don't necessarily always know that we can get at truth and so that's why there is kind of that hedge there is we're trying to get close to truth. And then usefulness, and usefulness is all about is evidence that's actually going to matter to a patient in some way, evidence that's going to actually help a patient or help avoid harm in some way. And that becomes something that frequently, you know, as Ward pointed out, may be very specific to an individual patient or it may be to a population as well. So there may be some different approaches in terms of how evidence is used because of that.

And to me, critical appraisal is a combination of what is clinical knowledge -- so here's my clinical knowledge right

here -- and understanding critical appraisal concepts, but it really is about critical thinking. And one of the things that I think is a beauty about doing this work is that, when you really understand what makes for good science, it's all very logical. And so we've actually seen people do -- within a workshop that's maybe a couple of days, even consumers come away with an ability to evaluate scientific information because there is commonsense behind it, and you get some information about shortcuts around that, like why blinding is important, but it really is just about logic and reasoning through, are these results likely to be true or does something else really explain the results? And so that's what that's all about.

If this is hard to see up here, on page 44, this -- I just want to show you what a critical appraisal looks like. Now, I have to give a little caveat in here. Well, I have to give several. One of the ways that we avoid me having to have any clinical knowledge is Mike and I just make up drugs, and sometimes, we make up conditions and then we can say anything we want about them. So -- but we wanted to show you kind of what the end game is and so page 44.

So what we have here is we made up an abstract. Now if you were going to look closely at our critical appraisal, you would find things mentioned in the critical appraisal that are not in the abstract and so that means that there is a phaco

study behind this that you don't have the parts and pieces to, but the reason why we do this is, for some groups -- does anyone here do critical appraisal? Okay. For some groups that do it, an easy way to report out what you've done is to just grab the abstract from PubMed and PubMed is just the entree, for those of you who may not know, into the National Library of Medicine to get access to medical research.

And so our critical appraisal then, what it does is it lists our findings. And one of the things that sometimes -- especially for people for whom this is newer, they come up to us, and they say, well, you're just focusing on the negative. That's right. It's not like evaluating an employee, and there are some strengths to them and some limitations, but you know, we're going to keep you because you're great and we love you.

With critical appraisal, it's like you're trying to make sure you've got a solid foundation and so you do focus on the negative. And so in this instance, in our made up study, we said study size was small. Okay. That doesn't necessarily mean that that's a big flaw, but when you have only a few people that you study, you have a greater opportunity for chance effects because you don't have a big pool of people with lots of characteristics.

Our primary endpoint, it was a composite endpoint. "End point" means the thing we're studying, like mortality. Composite endpoint might be something like all cause mortality

and hospitalization, and we combine those together and call those one thing, and Mike's going to do some exercises around that. So when we say a questionable composite, what we're saying that's shorthand for, there may be some things about putting these different endpoints together that may be misleading in some way.

Randomization, we're going to talk about why that's important, but in our study here, not truly randomized. Patients were assigned to groups by study consent date. Maybe this is a good moment to just chat about randomization for just a minute. Does anybody -- does someone want to talk about randomization and why, when I started talking about what Mike and I evaluate, that we focus on randomized control trials when we're dealing with therapies? What is randomization? What's it about it? Why do we care? What does it do? Yes?

COMMISSIONER HIPPLER: I would say randomization protects the study from the bias of the researcher.

MS. STRITE: In terms of how? I think you're right. I think that's a good answer, but there are some reasons why.

COMMISSIONER HIPPLER: Because the researcher naturally, instinctively selects people differently for different things.

MS. STRITE: Absolutely. Yeah (affirmative). So if I were somebody that was -- let's say there was a study looking at medical therapy versus surgical therapy, surgery of some

sort, and I don't look too good, I don't look like I'm going to be a good candidate for surgery, what does randomization help in that instance? As a clinician, you're going to want me to go to which group?

COMMISSIONER HIPPLER: Medical.

MS. STRITE: Medical therapy, right, because that's going to be the greatest likelihood that I'm going to have a better outcome. So if you can control that, then we've got a selection bias problem, and it could be now that we've got healthier patients in the surgery group, less healthy patients in the medical therapy group. Hey, surgery ended up looking better, not because it was better, but because we had healthier patients in that group. So what randomization does -- one of the things it does is, it keeps the investigator away from making choices, like that, that may be a selection bias. So we'll talk more about that, but randomization, when it comes to therapies, becomes very important.

Anyway, I'm not going to go through all of this, but I just wanted to give you kind of a sampling of what it is that a critical appraisal can look like, so when we talk about this, you see sort of a summary of where that's put together. And then what we do, as evaluators, after we have -- let's say we've got five critically appraised studies. We take a look at them, and we say, okay, do we think that this is information is information that we can trust? If it's

information that's likely to be trustable, will it actually benefit a patient? And then we keep together the studies that we agree meet those two criteria and then we take a look at kind of the summary of that evidence and how that may apply, and we'll talk more deeply about how that's done, but this is just a very kind of high elevator talk about that.

So, why do we want good science for medical decision making?

COMMISSIONER HARRELL: We want to have a firm foundation for this information.

MS. STRITE: You do.

COMMISSIONER HARRELL: You want it to be something other than the art.

MS. STRITE: The art, yes. And the art is important, but you want it combined with it. Exactly. And one of the reasons why we want that foundational information is, we're trying to get at cause and effect. So all throughout today when we talk about medical therapies, we're very focused on, if I take this pill, what is it likely to do to me? We're also very focused on probability because the fact of the matter is that it is maybe rare to never that a therapy is going to benefit every patient. And so part of what we want good science to help us understand is, what is that likelihood of benefit and that likelihood of harm?

So I'm going to do kind of another definitional thing

here. Number-needed-to-treat. I know there are people here who know what that is. Is there anyone who is willing to explain that to the group? What's number-needed-to-treat?

COMMISSIONER HARRELL: It's the number of individuals you need to treat with a particular therapy to see a difference in the outcome.

MS. STRITE: Great. Yes. So if I had two agents and -- so science is all about comparing -- again, we are always talking about therapies here, unless we say otherwise. So we're interested in comparing two treatments. And so number-needed to treat is, what is the differential benefit to somebody in a group, if a treatment has shown itself to be superior? And that is the likelihood equation. And so what would be the ideal number of people you would want to treat to benefit one person over the comparator agent? What would be the ideal number of patients to treat to benefit?

UNIDENTIFIED MALE: One.

MS. STRITE: One. Okay. To all the doctors in this room and any other clinicians, how many times do you see that happen in medicine? Zero. Right. Every group we've ever been to always says zero. So this becomes very important in terms of the science because we do want to understand that probability of benefit, and often times, the probability of benefit is very, very low.

So when we're doing critical appraisal, to break this

down in very, very simple terms, we're asking two very important questions. Is it true, and it is useful? And then there is a third one that Mike and I focus on in terms of clinical improvement, and we're going to talk about that later in the day and that is, is it useable? So you could have -- you know, Oregon Health and Science University has done some great work through the DERP program. Are people familiar with that, the Drug -- oh, help me; what's it stand for? Drug Utilize -- no. Drug Evaluation Research Program, I think it is. And they would produce these documents. I don't know if they're doing it anymore, but like, you know, 200 pages that Mike had to sift through one day for three hours to determine that it said all proton pump inhibitors are essentially alike, when that was produced. So the usability part is then, how do you then take this information and put it into informational tools that can help providers and then help patients make decisions?

So is it true, is it reliable.....

MR. KNAPP: Question.

MS. STRITE: Yes. Sorry.

MR. KNAPP: I'll just make one additional observation.

UNIDENTIFIED FEMALE: What's your name?

MS. STRITE: Say your name, please.

MR. KNAPP: One additional observation. So you want to use science to -- obviously, let's say you're talking about,

you know, this new drug, and the test is, it is reliable, is it clinically useful and so on. Is that going to make the patient better? Now, I happen to be an economist, and economists have this irksome way of saying, actually, there is another question you could throw in there, which is, is it worth the cost.

MS. STRITE: Right. And there is more to that. She wants your name though, too. Everyone needs to, I guess, say their name into the microphone.

MR. KNAPP: Oh, I'm sorry. My name is Gunnar Knapp. Yeah (affirmative). So that's just another point. There's whole -- aside from the whole effect -- the question of, does this thing make you medically better off, economists would argue that there is a question of, does it make you enough better off or enough people better off to be worth the cost?

MS. STRITE: And that's not an argument. There are a whole bunch of decisions that go into it. It's not even that. At Group Health Cooperative, we had a woman who had (indiscernible - recording interference) can help her. Well, gee, wouldn't we like to be splashed on the front page of the paper, denying this patient, and you know, three small children of care? So there are many, many decisions that go into what you're going to do for a population, what you're going to offer to a patient.

Here, one of our issues though is that people need to

start with the science and get that clear and straight because that's the universal thing. And then you've got all these localized things or things that may change over time because the economics of it can vary. And so we're always trying to encourage people look at that later because you don't want to bias yourself because, oh, now we've got an agent that's, you know, $100,000 a year for treatment. Well, get it straight first whether the science tells us if that's actually going to work because maybe tomorrow, suddenly, it's free. And we see some people leaning toward the cost, saying, oh, you know, they want to find everything wrong with it scientifically so that they're decision really ends up being a biased one, not looking at the science in a clean way. So we say start with the science, and this is actually what Ward and Mike did. The Technology Assessment Program that we had at Group Health, they said let's get away from all those other factors at the start and focus just, first, on the science, and then as Mike puts it very cutely, throw it over the wall into all of these other considerations of which there are many. And we will talk about that when we get to the policy making discussion, too, which doesn't mean people shouldn't engage with this, but we have more to say about that.

MR. STUART: I would just add that, when you're looking at the economic issues, they're so closely tied with the science issues many times that, if you don't do them both,

your cost inclusions are going to be wrong. So I saw a systematic review years ago that really stuck in my mind. It was on Cox 2's (ph), and the economic model was based -- they put 50% of the weight on a study that was totally flawed, that showed a lot of benefit, but in fact, because it was flawed, the benefit wasn't there. Therefore, the cost projections were wrong. So the cost is a big piece, but it is really a big mistake to go look at results of studies and then start doing economic modeling until you know that the studies that are talking about the efficacy are reliable. That's why we start with the evidence, and we'll go through this several times because that's not an uncommon problem is to get the cost wrong because you didn't get the evidence right.

MS. STRITE: So in terms of trying to assess a study for its reliability, the "is it true," we are looking for three things when we critically appraise. We're looking for bias. We're looking for confounding, and we're looking for chance. And what bias is -- those of us who live in lay people's land think of bias as, like, a bias researcher. Well, that can be part of it, but the formal definition of bias for when we do critical appraisal is this, something that systematically leads away from truth. And all the word "systematically" is doing there is saying it's something that led away from truth, not due to chance. So we can be led away from truth due to chance. As I mentioned, small study -- it might be that we

have too few people to show, you know, kind of what would happen if we had enough people there and so chance explains our outcomes.

And then confounding, which I'll define in a few minutes, is a special form of bias. So in effect, what we're doing is we're looking to rule out bias and chance, i.e. leading away from truth, and if we can do that, then we conclude it's likely to be true.

And then clinical usefulness is what I consider to be looking at meaningful clinical benefit, and to this gentleman's point, part of the meaningful clinical benefit has to do with not just that we show benefit, but we show benefit that's large enough, you know, for enough people that it really is meaningful and matters.

So just a few more points there that I think are important because some people say, well, you know, can't we just rely on the FDA? I mean, that's why we have them, right? Well again, a lot of people that work in healthcare decision making do not have this framework, and so in every group, there is variability in terms of how good people are at doing critical appraisal or not. Also to the FDA, there are certain things they just don't do. They don't do comparative assessments. So you know, that's an issue, too, and that's frequently a question.

As we've mentioned, there's a general lack of skills.

And then we've talked about biases. And it's important to know people's conflicts of interest. I agree with that, and Mike and I always pay attention to when a manufacturer has funded a study, and we've always been interested in looking to what the exact role is, but there is actually a lot of manufacturers' research that's done that's good. But the other thing that's important to know is that everybody engaged in research has a bias. So it's more than just the manufacturers. If I have an academician -- let's say I'm an assistant professor -- I am probably working in a department where there is pressure to publish, you know, the whole publisher Parish adage. I may have spent three years getting a grant, three years doing the grant, and now I get the results, and I didn't get a statistically significant outcome that I was hoping to get, and oh, by the way, I'm in the research because I'm rooting for the thing I'm studying. Well, I walk down the hall to my friendly neighborhood statistician and say, well, there's got to be something in here that's publishable. So it's very important to be aware that anybody involved in doing research is potentially biased and so what you want to do is understand the critical appraisal concepts and know that good study design and execution helps prevent those things from happening.

One of the things that we use is a framework just to help people understand the magnitude of this problem is we use the

Vietnam War, and just in terms of U.S. lives, there were roughly 60,000 people who died.

Here is an example of a study, however, that had to do with women with advanced breast cancer who got autologous bone marrow transplant and high dose chemotherapy, and this was through observational studies. Observational studies are not experiments. They're not clinical trials. And this became high on the uptake because doctors, quote, literally watched tumors melt before their eyes. And then when randomized control trials were actually done, it showed no benefit. In fact, one of the trials showed something like a 30% increase in mortality within 30 days for the women who got this treatment that was horrific. I mean, this treatment is so beyond painful that it's beyond description, and here, we had a huge uptake on it because we didn't have good science telling us that this was not going to be a benefit, and then to the cost question, $3.4 billion.

Some caveats. Mike said that we're going to focus on therapies. There are several kinds of comparative designs. So we're going to focus on what's called superiority studies, unless we say otherwise. Superiority studies just means that I'm looking at Agent A, I'm looking at Agent B, and I'm trying to figure out which is superior.

There is also what's called non-inferiority and equivalence trials, which have more to do with equivalence

kinds of issues.

Some groups focus on making decisions in terms of looking at drug classes. We understand reasons for people doing that, but we think that that's sometimes not going to give patients the best information, rather than looking at individual agents. There are exceptions to everything we say. There is a lot of judgment that's required in this work, and our best answer to everything which, at the end of the day, we want everyone repeating with us is, it depends. But because of that, we have to do a lot of pattern learning kinds of things and so we generalize things.

A few more terminology things. When we talk about primary studies, that's the original research that's published from a study. Secondary studies, that would be studies of studies, groups of studies that are studied together, like a med analysis. Mike mentioned systematic review. That's an overarching term for a certain type of secondary studies that's done in an evidence-based protocol kind of way, and med analysis is a subset of that. And then secondary sources, we use that to refer to any medical information that is citing a primary or a secondary study.

Other terminology thing: PICOTS. Has anyone heard of PICOTS? How about PICO?

MR. STUART: She has.

MS. STRITE: Where people -- PICO.....

UNIDENTIFIED FEMALE: I'm a nurse. Yeah (affirmative).

MR. STUART: Okay.

MS. STRITE: A lot of people have taught PICO and then they added the "TS" on there, and what this is, is people who were helping people create clinical questions came up with this outline and so it's very nice because it can be used to summarize a number of things about research. So PICOTS -- so when Mike will talk about PICOTS, he'll say, "The PICOTS of it," and what he's talking about is the patient, the intervention, comparators, outcomes, timings and settings. So these are contextual elements of different settings, and they help us look at, for example, heterogeneity, which is important because some studies might not be combinable because the populations were so radically different.

And then for searching, we actually focus not on that, but we use CI, Condition and Intervention, because you get more synonyms if you're doing a PubMed search. So that's just a quick tip for those of you that are doing searching.

What have I not covered here? Confounder. So let me show you a picture. A confounder, it's a factor other than what you're studying that actually might be responsible for or affect your study results. So an example would be, here's an observational study in which people were interested in looking at whether vitamins reduced risk of coronary heart disease. Wow, they found out, yes, it did. Well, it is really the case

that the vitamins caused the reduced CHD or was it something else? And I mentioned that confounding is a special form of bias.

Confounding happens when we've got a couple of variables that are related to each other. In this instance, if I take vitamins, it's very probably the case I'm interested in doing other lifestyle things that are going to benefit me, like exercise, for example, and pay attention to how I eat. And so in this example, the real cause was the healthier lifestyle, but it appeared that vitamins reduced, and in fact, they did not. So this would be an example of a confounder that confuses us as to the true causal agent.

And grading. After we review a study -- so for example, there, on page 44, we show you we would take a look at the threats to validity and then we would kind of do a Gestalt and say -- again, a lot of judgment in this -- you know, do we feel comfortable with these results or not or do we think the results could be so distorted by bias that we cannot trust them? We would then assign a grade. And I will tell our grade system, just because we use it in language a lot because it's kind of fast to do -- but I will mention there are over 120 grading systems out there, and so whenever you're looking at a grade of a study, you need to look at the criteria for getting that grade, and you'll find some of the criteria can be extremely loose and make low quality studies appear to get

a grade “A” because they’ve got bad criteria. And if you see level one evidence in one system and someone else is using that term, the criteria may not be the same. So you need to always kind of look at the legend and see what they’re doing.

Ours. Grade A is like it sounds. We have graded one study a Grade A, and it was a manufacturer study. Grade B we give about three percent of the time. And then we didn’t want to have something that sounds passing. We wanted to do kind of a teaching moment, so we have what we call Grade U, saying we’re still uncertain and so we can’t rely on these results. And then we’ve got kind of a middle of the road there that’s a BU, so we’re saying maybe highly likely that it’s true, but we still have enough uncertainty that we want to signal that to patients and providers, and about seven percent of the studies that we evaluate get a BU, and we’re really happy when we see a BU. We use A,B, BU for efficacy. We will use Grade U for safety, and Mike will explain later why safety is so hard to get at. So it’s not that we are saying that safety gets sort of a free pass. We still describe the evidence as being of low quality, but because safety is so hard to get at and because it matters so much, we’re more open to reporting the Grade U evidence for that.

The PMID number. If we’ve got anything referenced, you’ll see about a ten or so digit number. That number, you can put into the PubMed search window, and it will call up the

citation and the abstract. So we use that a lot. We like it.

Now, I'm going to teach people how to read a Forest Plot. I don't know where this is taken from, but isn't this beautiful? Mike and I have been enjoying the mountains out here, which are gorgeous.

How many people have ever read a Forest Plot before, that they know of? Okay. For those of you who have not, and actually for those of you who have, there are certainly the slides you can follow, but I'm just going to mention that, on page 11, you've got a one-pager there that puts all of these slides together in one place, and we use this for some of our case studies. So that's why we're going to show you this now.

What a Forest Plot does is it takes a bunch of studies and lines them up and gives you some graphical information about the results. And so here, we've got -- I'm never very good with this. So this isn't meant to add up to this number. It's just a sampler because we wanted to show you the various elements of a Forest Plot.

So here, we've got, you know, our sample studies that we picked out here, and this -- let's see; where do I want to start? This square the calculated outcome for that study. So whatever it was that they were studying -- in this instance, I'm going to guess it was mortality -- that is the calculated difference between the groups.

What the lines are, the lines are the confidence

intervals, and a confidence interval is a statistical calculation that says equally, statistically plausible results are along this line. And the nice thing about confidence intervals is that it gives you a sense of how uncertain you may be in a valid study -- we're going to assume everything here is a valid study -- because, if the line is really tight, it may mean that you had a big difference between the groups and/or you had a big sample size. So the confidence interval is helpful, but what research has shown is the true answer to a valid study is usually that. So again, this gives you a range.

So look at this one here. See how wide that is? This is why. There are only 50 people in the study. Everything on this side of the line on this graph is favoring the intervention. Everything on this side is favoring the placebo. When they added up all the studies and the missing ones we took out, they represent the summary in a diamond form.

MR. STUART: So each horizontal line is a study, and what it's doing is adding up all the little studies to get the answer as if, instead of studies, they were people. So we're picking lots of studies and putting them together. The idea for meta-analysis -- the rule is when many small studies existed and there was not a good likelihood that a very, very large study would be done -- so they put the studies together,

and as you can see in this chart when you add up all the studies, it shows that the intervention favors -- or that the results favor the intervention. And so at the bottom, there is a diamond that's clearly not touching that vertical line. So that tells you, if you can trust this, if the studies are valid, that whatever this intervention is, if it happened to be for systolic hypertension in elderly patients, that the mortality rate was lower, if they treated the blood pressure.

But does everybody understand the Forest Plot, the idea of just putting a number of studies together to get what's the equivalent of a big study? That's what a systematic review of meta-analysis -- that's the goal of those secondary type of studies.

MS. STRITE: So this line here that divides favoring intervention and favoring placebos is just The Line of No Difference, and there are some fancier terms for that. For the lay folks, you don't need to pay attention to this part, but for anyone working clinically, sometimes, you hear infinity or unity. It just means that centerline of no difference.

It is statistically impossible that you can both favor the intervention -- not just statistically. What I am saying? It is not possible.....

MR. STUART: No. In reality, it's not possible.

MS. STRITE: In reality, it is not -- or statistically.

But you can favor the intervention and the placebo. So anything that touches that line is not statistically significant, i.e. the P-value, you know, is not reaching significance there, which has to do with whether or not the results are likely to be due to chance or not.

So if I were to ask people what here is statistically significant, what would you tell me? Study C.

UNIDENTIFIED FEMALE: Ask that question again.

MS. STRITE: What's statistically significant on that graph? C is. Uh-huh (affirmative). Mike had told you before, there is one other thing. And the diamond. And the summary diamond. That is not touching that line. So when they added up all the people, we ended up with a statistically significant outcome.

Now for the lay folks, you probably will never experience this again so you don't have to worry about this, but for the clinical folks, there is a little trick here. Can someone tell me why, if The Line of No Difference is no difference, why is there a one at the bottom of it? Doesn't that seem odd that one would be no difference? What would you expect it to be? Zero. So why one?

Well, the clue is up here. This is what's called the Measure of Outcome. There are a bunch of Measures of Outcome. Like Numbered Needed to Treat is a Measure of Outcomes. Odds ratio is a Measure of Outcomes. And when you have a ratio, No

Difference is a one-to-one correspondence. So if you see one, that's what's going on there. So relative risk, that would be one. If you have Measures of Outcomes that are expressed as percentages, like Absolute Risk Reduction and Relative Risk Reduction, No Difference is zero. And so if this were ARR up there, Absolute Risk Reduction, The Line of No Difference would equal zero. So that's just a little trick.

And on the Forest Plot, you don't really need to know that so much because you can see, graphically, what's going on, but when you get the confidence intervals just presented as numbers, knowing that trick becomes very important so you can see if it's statistically significant or not.

So I think we've covered sort of all the sort of background stuff that I wanted to cover and so.....

MR. STUART: Okay. So let's -- we have here a healthcare provider and a patient, as represented by a mom and her child. What is going on in this meeting? What's happening here?

UNIDENTIFIED MALE: Examination.

MR. STUART: That's right. There is an examination going on. Other things that are going on, in people's minds, for example, in all these three people's heads?

UNIDENTIFIED MALE: He's using a stethoscope without taking the shirt off.

MR. STUART: Yes. That has had a little evolution. In the old days, we were not taught that, but we do see it as the

norm now.

UNIDENTIFIED MALE: It’s not necessarily right though.

MR. STUART: No. We don’t have good evidence on that, do we?

UNIDENTIFIED MALE: Yeah (affirmative). Well, there’s pretty good evidence that you’re going to miss a few murmurs if you listen through all the clothes.

MR. STUART: Oh, I haven’t seen that evidence, but that may be -- it sounds right. So what else could be going on here? These people have hopes. They have fears. They have thoughts. What kind of things do you think are going through these heads? This is a medical encounter.

COMMISSIONER HARRELL: The actual patient, at the moment, is uncertain about what’s going on.

MR. STUART: That’s for sure. Okay. So what I want to do is sort of get into a different little module here called outcomes. What outcomes do people want from healthcare visits and interventions? What is it that we all want? Well, we want to avoid avoidable mortality; that’s number one. Number two, we don’t want any unnecessary morbidity. So we want to avoid morbidity that we don’t have to experience. We don’t want to have a lower quality of life, health-related quality of life than we need to. We go to doctors. We go to others. We go to healthcare professionals because we don’t want any of those things. We want high healthcare quality of life. And

number four, we want our symptoms to be relieved. There is nothing worse than severe pain. I mean, it's just unbelievable. So we have to have something for symptoms, whatever the symptoms are. At least, these are our hopes as patients. And then the last one is we want to function. We all want to function as humans in these ways. We want to emotionally, physically, and mentally be at our best, and we ask the healthcare system for what they can do to help us with this. And we will come back to these five outcome areas repeatedly because these are the outcomes that matter to patients. Yes? There is a question back there.

DR. GRANHOLM: Is there.....

MS. STRITE: And say your name first.

MR. STUART: Name first. Light on.

DR. GRANHOLM: Marin Granholm, family physician. I wanted to ask, is there a community function of healthcare as well, not just individual?

MR. STUART: I.....

MS. STRITE: Is there a community function of healthcare?

MR. STUART: Oh, of course. Absolutely. And I think the various groups could -- we always start with the patient, and as we're working in different ways, there is the physician or other healthcare provider. There is the cultural unit. There is the organization where you get your care. There are the payers who are paying for it. There are the decision makers.

There are the commissioners. There are all these different groups, and they all need the same skills at looking at how to deliver these five outcome areas. It's the responsibility of all of those groups, starting from the patient and all the other ones. Yes?

MR. KNAPP: I'd just like to make an observation -- this is Gunnar Knapp -- that, in America, a not insignificant number of the patients in that picture would also be wanting -- that I hope this visit isn't going to cost me $800 or whatever the doctor tells me. So there is an economic side to this.

MR. STUART: Sure.

MR. KNAPP: And a lot of people don't think about that, but a significant number do.

MR. STUART: Thank you for bringing that up. We're going to come back to that, and it is definitely a quality of life issue and maybe even a function issue, if you don't have enough money to go downtown because you don't have any gas or a car or anything else because you've spent it all on healthcare. So it's an incredibly important piece, and we're going to come back to that when we get to what we call triangulation areas. But we want to start with the evidence first and then add these other considerations, cost and many others, too. Yes?

DR. GRANHOLM: Marin Granholm again. I don't want to

belabor the point, but I guess my point, thinking about your picture, was the evidence-based part of that visit for that kid was probably his immunization, which will benefit the public more than it will benefit him.

MR. STUART: Yes. We're -- the point that I'm trying to make is that we want the most benefits we can get for the patient by involving everybody that has any contact with the information or the patient that gets to the patient in some way. Yeah (affirmative).

So here again are those five outcome areas. There is morbidity, mortality, symptom relief, functioning, and health-related quality of life. And we're going to repeat those many, many times because we want people to remember these because, if it isn't one of those things, it's called an intermediate outcome measure or a surrogate outcome, and people don't experience those directly. That would be a lab test. That might be the result of an x-ray. Although the shot might -- you might experience the shot from a certain of these things, but the idea is that those five areas are what people really experience, and when we have studies that don't have those as outcomes that are being studied, they're going to be using a stand-in for one of those things and that might be a hemoglobin A1 level, hemoglobin A1c level. It might be uric acid blood test. It might be any kind of lab or diagnostic technology result, but those things, unless they

can be directly related to one of those five areas, are considered to be -- you have to be very careful about looking and accepting the results of studies that don't have those five outcome areas in them. And then I think we should a little bit about safety. We're going to say more about this later, but if we're -- let's say we have a study, and because we don't have a large budget, we can't do as large a study as we want. What we're likely to do is use what are called composite outcomes. So we might say, if they have a death, a stroke, a heart attack, we're going to put those all together and that's called a composite. And you're going to see lots of composites because it's more efficient to do a study with a composite, but what you want to be aware of is that the components of the composite endpoint should all make sense and fit together.

And in this case, the first one, we have all-cause mortality, MI, stroke, and rash. Well, why did they have rash in there? This is an obvious sort of silly example that we put in there, but in fact, if the rash is the most common, that will drive the outcome of the study because it's the most common event that people experience.

The second on here is a little better, but it's got a different problem. All-cause mortality, myocardial infarction, stroke, admission to the hospital for cardiovascular problem. Now, what is the issue with that last

item, admission to the hospital for a cardiovascular problem? There is a potential problem with that one because it's under the control of the doctor. It is not something that is experienced in the same way that these other things are because there is choice involved, and we'll get to -- we have understanding studies by whether there is choice or whether there isn't choice, and if there is choice involved anywhere along in a study, you have to, at least, consider whether that might bias that study.

The last one is hemoglobin A1c, a blood test for diabetes, advanced retinal diseases determined by an opthamologist, end-stage renal disease, cardiovascular outcomes, and all-cause mortality. The point being there may be some disagreements on some of these about how good that combination is. So there is an example, right away, of judgment that comes into this, and some people will have different judgments about composites. And I'm just using composites as one example of where judgment has to be applied to see if they're reasonable composites. Uh-huh (affirmative)? Light on and name, please.

MS. EDMISTON: This is Gena Edmiston from Fairbanks. One of the endpoints that we see widely used in more, say, quality improvement types of studies is reduced length of stay and that's one that has a lot of baggage that goes with it.

MR. STUART: Would you repeat that?

MS. EDMISTON: One of the common outcomes that we see used in quality improvement studies is reduced length of stay.....

MR. STUART: Length of stay, yes.

MS. EDMISTON: .....which is a pretty baggage laden endpoint, but we see it used really frequently.

MR. STUART: Uh-huh (affirmative). Yeah (affirmative). We'll get to performance measures a little bit later, but thank you. That's a very good point. It's quite different than a myocardial infarction. It's a different type of animal, and.....

MS. STRITE: Well again, you've got the potential for someone being in control of that and so there is a possibility that they can drive the outcomes of that research by choices that they make. So you've got a validity problem, potentially.

MR. STUART: So let's take a look at this statement and have somebody comment on it. Did you just go ahead one slide? I don't know where we are anymore. You went back one? Now, you went forward one? So that's fine. I can keep going.

To demonstrate the efficacy of coronary artery bypass or angioplasty when you compare it to medical treatment, a reasonable study would be one in which people undergoing bypass are compared to those managed with medical therapy. We'll call it optimal medical therapy. Just, what sort of

thoughts come into your mind about that study, right off the bat? Yes?

COMMISSIONER HIPPLER: Are the people undergoing bypass medically similar to the people who are not?

MR. STUART: Right. That's a good question that you would ask, and how would you -- I mean, how would you know? How would you answer that?

COMMISSIONER HIPPLER: Were they randomly selected?

MR. STUART: Right, because there is nothing on this slide that says, what type of study is this? It doesn't say it's a randomized control trial. It doesn't say it's an observational trial, and you would want to know that and that's kind of the punch line for this slide. And it brings up the idea that, in healthcare -- and when you're looking at a study, the first thing you want to look at is, what is the study type? We don't know what the study type here is. It may be a database study. In fact, this one was. It was from a Canadian, Ottawa database, and what they did was they compared the outcomes in CABG, coronary bypass grafting, angioplasty, and medical therapy, and they found that medical therapy came out at the bottom. Although later, a randomized control trial said that, in stable disease, there is really not a great deal of difference in outcomes. Well, that's because there were two types of studies that were done. This one was what's called an observational study. Anytime a

doctor or doctor and patient choose a treatment, that's an observational study.

Now for this group, you don't need to know all the epi that -- the epidemiologist -- you don't need to know cohort study versus case control versus prospective cohort versus retrospective cohort. That isn't important.

What's important for you to know is there are two buckets. Bucket one, observational studies, meaning all I do is look at what people chose to do and report it in my study. So I will look at all the people who had medical therapy. I will look at all those that had CABG and all those that had angioplasty. I will statistically try to adjust to make them similar, but I'm really reporting on what people chose to do. Now that's very different than if I take all that choice away and I randomize people, as he said, to two groups, and why do I randomize? And he asked the absolutely first important question, the correct question; was it randomized? What does randomization do that you don't get in observational studies where there is choice involved? You want to continue on with that, since you started?

COMMISSIONER HIPPLER: It makes the populations the same.

MR. STUART: It makes the populations similar, and how does it do that? When you randomize, you're putting people into groups by chance, and because, as human beings, we have an infinite number of variables -- our diets, our cultures,

our habits, where we live, our genetics -- it goes on and on and on; it’s infinite.

So what we want to do is distribute those variables, and those are called prognostic variables. The way we distribute them is we randomize people to two groups. Each study should have, at least, two groups that we compare, and everything should be exactly the same in the two groups, except for what we’re studying, but in this case, it’s very likely we did not start out with equal (indiscernible - background noise). So we are getting a high chance of bias from that type of study, but that’s just one question, the design study.

How about this one? An appropriate study method would be to compare patients receiving surgical treatment with patients treated medically in a different healthcare facility. Thoughts about that one? What would be the problem? The way this is presented, you know, in the worst case scenario, what could happen -- what could go wrong with this type of study? Well, is there anybody else that wants to take.....

COMMISSIONER HARRELL: (Indiscernible - away from mic) methodology - different things.

MR. STUART: Great. Could you use your light and your name and repeat? That was good.

COMMISSIONER HARRELL: It’s Colonel Harrell. Methodologies would be different, different facilities.

MR. STUART: Yeah (affirmative). There would be

different processes. There would be different cultures in the different hospitals. They would have many, many different experiences. The contextual elements are going to vary greatly, for sure, between the two hospitals. So you are not starting out with two equal groups. You are not going to have equal experiences and so you have a type of selection bias because the patients may be different, and you're going to have what's called a performance bias because they get things -- they experience different care in the different groups. Thank you.

So that gives you an idea of what we do. We start out with the type of study and then we start getting more and more granular, which we're going to slowly do as we go along.

Here's Bill Clinton. Let's do another one. Bill is admitted to Hospital A for CABG surgery, and a newspaper reports that Bill has been admitted to the hospital with the highest rates of mortality resulting from CABG in his metropolitan area. His hospital comes in at three percent compared to two percent and one percent for other primary hospitals in his area, the Hospitals B and C, and the difference between these hospitals is statistically significant, and Bill just gave me a million dollars to advise him, should he change hospitals? Do you think he should go to one of those other hospitals or should he stay?

COMMISSIONER HIPPLER: It depends.

MR. STUART: It depends.

MS. STRITE: Yay! Early adopter. Yay!

MR. STUART: Okay. Now, go ahead and restate your name with your light on and tell us more.

COMMISSIONER HIPPLER: Allen Hippler. It depends because the populations could be different. For example, maybe three percent mortality means that all the hardest cases go there, and it's actual evidence that it's the best hospital.

MR. STUART: Uh-huh (affirmative). Right. So you could adjust these things many different ways, but in fact, this -- we don't really know for sure, even though it's reported, that it's the worst hospital. We don't really know that from this data, and this -- we're not going to sort of dig really deep into this issue right now about performance measures, but this is something that comes up all the time is that, can you rely on that information? We're just giving you examples of where you may or may not be able to rely on information and how that's going to come at you. Yes?

DR. BROWNSBERGER: Well, an example we have in this town is -- I don't know if any of you have seen the new *Consumer's Report*, but Providence Hospital gets the lowest grade of any hospital in Alaska on their surgical complications.

MS. STRITE: And we need your name, also.

DR. BROWNSBERGER: Keith Brownsberger.

MS. STRITE: Thank you.

DR. BROWNSBERGER: And I suspect it's because most of the really tough cases go there. If you polled physicians in Anchorage and asked them which hospital they would like to go for surgery, I think the majority of physicians in Anchorage would probably select Providence Hospital, as they would go. So I think it's an example of this same kind of thing.....

MR. STUART: Yes.

DR. BROWNSBERGER: .....that's really close to home.

MR. STUART: Yes. And we're exploring it very superficially at this point. We're going to get deeper as we go, but we're just sort of raising some issues here.

The goal of clinical trials, or one of the main goals, is to determine causality. Is the explanation for the results due to the intervention that we're studying or -- and that would be cause and effect or causality -- is it due to bias or chance? And that's the goal of critical appraisal is to find out which of those is true. Is it -- are the outcomes due to bias or chance or are those outcomes true? And it's really not that difficult to get at this, if you learn some basic principles, concepts, and maybe have a tool or two around.

Now -- so once we have critically appraised a trial for reliability, then we look at the benefits and the harms, but we don't look at those benefits and the harms as a first step because, if, upstream, the design and the methods of the study are flawed, we cannot trust those results. So in evidence-

based medicine, critical appraisal specifically, we always want to critically appraise the study design and methodology before we seriously look at the results. Granted, we look at the abstracts and the results, too, just to get an idea of what's going on, but we don't trust them or accept them until we've gone back and done the critical appraisal. We also -- when we're looking at safety, we might do some other things. We might look at more than randomized control trials, and we might look at, actually, registry data, observational data, and we do it in a different way, but safety is a unique area, and we'll come back to that towards the end. But for efficacy, meaning does it actually works, we're going to be really tough on these studies, and even though we're tough, we have, like David Eddy you mentioned -- I called him up and said, how many -- what percentage of studies do you think really are valid? You know what his answer was? One percent. So one out of 100. Now that's not really very good. This was a couple years ago. But he doesn't know that. It's just that, after living in this world for so long, you get used to nice, beautiful-looking studies and glossy covers of journals, and yet within them, you have flawed studies.

So the outcome -- the goal of clinical trials is to find out whether the study is reliable and to make sure that the benefits outweigh the harms so that patients can make decisions. And the way we do that is we just look at how many

people in each group experience the outcomes that are being studied and we compare them.

The patients should really know that, when we're talking to them about safety, almost invariably, we have limited data because studies are too short. They're not formally looking at safety. Many of the investigators participating or the clinicians don't pay as much attention to the safety issues as they do the "does it work" efficacy issues. So safety -- patients need to know, many times, we're not going to be able to give you as accurate information about safety as we would like.

So in summary, we want, at least, two groups. We want to study an intervention, a therapy, and we want one group to get that and the other group to get the placebo or the usual care or whatever the comparator is and that's the only difference. We want those two groups to be the same at the beginning, so we randomize. We want to make sure that the contextual elements that they experience are the same.

So the first one is called selection bias. The second is called performance bias and that goes for everything from, do they have the same co-interventions with drugs, is the frequency of visits the same in both groups, all these things, and then is the data management done the same in both groups, and are they assessed the same, and it's really that simple, except there are about 20 things you have to actually look at

and then you've critically appraised the study and you can make a decision about whether you want to use those results.

Another way of saying this is you want to isolate the intervention. You want everything to be the same, except the intervention. Randomization is the way to get the first step done right, and blinding is the key to making sure the experience and the co-interventions can't be different between the two, and there are some other things there, but we want to make sure we have randomized control trials for therapy, that they're well done, that the randomization process is done well, that the blinding is done well, and that the data assessment is done well.

We want to remove choice from everybody in the study; that's the key word. If you allow choice, you will allow variation to creep in so the groups become unbalanced, and there are many ways for groups to become unbalanced, and we want to make sure that they stay balanced as much as possible.

Then same with measurement. We want measurement to be done the same way. We really have to think about, if there is no difference between the groups, was the study too small, was it underpowered is the way that's addressed, and we can talk about confidence intervals later as a way of dealing with power.

The final thing in the result department is, are the -- even if the study is valid, meaning likely to be true, and

there is a difference between the groups and it shows benefit, the next question is, is there enough benefit, and as we will agree, is it worth the money? So how big is the benefit? What is the benefit? Is it meaningful? And is it worth the money? In a way, those are the key principles on which we're going to build. I think we don't need to do these again. We've discussed them. It's now time for a ten-minute break.

COMMISSIONER HARRELL: (Indiscernible - away from mic) before we break?

MR. STUART: Yes. Of course.

COMMISSIONER HARRELL: So this is Colonel Harrell again. As you made the comment about randomization and blinding, you mentioned early on you're going to get into censoring.

MR. STUART: Censoring.

COMMISSIONER HARRELL: I'm very curious as to how that interplay between blinding and patient informed consent for the randomized blind trial ends up impacting that because you can end up excluding patients in the informed consent process and it affects the blinding portion of the study, so I'd like to talk about it, at some point, offline or in the group.

MR. STUART: Maybe the best way for you and me to do this at the break right now and then we can bring it up to the group as -- maybe in a really short summarized way, but yes; there are a lot of things in what you just raised. There are things that have to do with included criteria, excluded

criteria, the randomization process, including concealment of allocation, generation of the sequence, baseline characteristics table, are they the same, and then in the blinding, if there are side effects, can the people be unblinded by the taste or the side effects. So all these things have to be looked at, but in terms of censoring, we're really talking about time to event trial. So there's a -- that brings in another whole dimension. It's a whole unique type of study design that's.....

MS. STRITE: Analysis.

MR. STUART: Analysis that's being seen more and more all the time, a time to event analysis. So let's talk about that because you brought up, actually, a very complex set of issues, and we'll -- I want to dissect to get to the main one that you're interested in and then we can talk about this after the break.

DR. BROWNSBERGER: Before the break, I would like to come out in defense of medical schools. In 1956 at Stanford University, our pharmacology professor started out by telling us that most of the drugs that were used now would not be used in the future, that we would need to evaluate each drug, and he gave us a statistics course in order to evaluate new drug testing, and he made us all extremely suspicious of every new drug that came out.

MS. STRITE: And we do need your name, and we want to

(indiscernible - simultaneous speaking).

DR. BROWNSBERGER: Keith Brownsberger.

MS. STRITE: And some medical schools do it well, but most of the physicians that we work with, most of the medical students, most of the residents are not getting the kind of training that you did, unfortunately, or it's not done in such a way. You know, for a lot of medical students, they get statistics, and it's, like, this is a course to just get over and get through. You know, where is the patient? I want to be focusing on the patient.

DR. BROWNSBERGER: And as a result of that, our medical school class started an award and that award was to the professor who gave us the least amount of wrong information.

MS. STRITE: Oh, I love it. I love it.

MR. STUART: Very good. Okay. Ten-minute break.

9:55:39

(Off record)

(On record)

10:07:38

MS. STRITE: Slide 57. Not page 57, but slide 57.

MR. STUART: Right. Okay. As soon as I can use this, let me know, would you, so I could put up a slide? Yeah (affirmative).

The slide 57 says "Watershed Moment 1991 Study Design." This is a story about what's called the CAST trial. When I

was.....

MS. STRITE: How many people are familiar with that?

MR. STUART: Tom was, but he just had to step out. The cardiologists all know the story, but I think we only have one cardiologist. Are there other cardiologists here? Okay.

This study is to illustrate a key point. What happened here is that it was observed that people who had had heart attacks and had what are called premature ventricular contractions, PVCs, or extra beats or extra systoles -- they're called different things, but the heart goes beat, beat, beat. You get that quick, extra beat. That's a premature beat. So it should just be beat, beat, beat, but if your heart goes beat, beat, beat, beat, beat, beat, that extra beat in there, if you've had a heart attack, is a signal that you're at twice the risk of cardiac death than if you didn't have that premature beat. So this well-known.

It was also discovered that, if you treat these premature beats with Encainide or Flecainide -- two drugs that had come out at this time -- you could decrease the number of these premature beats. Well, the way medicine works is doctors think -- as David Eddy would say, they think real hard, and they say, this makes perfect sense to treat these people.

I want to emphasize we had no randomized control trials, which is how we determined whether something works. What we did have was a bunch of observational data saying that these

people were at higher risk. So you can't blame physicians, including myself, who didn't know anything at that time, for using these drugs, and we said -- we actually believed we were helping our patients. They would continue to come back, and we did not see that we were hurting patients, but in fact, we were hurting patients, and here's the way that worked.

Well, the -- a randomized control trial finally got done, and it showed that, basically, the mortality rate was higher in the group that got the drugs than the group that did not get the drugs. We don't know what it was about these drugs that, even though they got rid of ventricular beats, they got rid of patients. So what we had here is, out of 100 people, three died with the placebo group, but eight died with the new drugs. So that's a difference of five percent, right? So five out of 100 died, but wouldn't have died had they not gotten the drugs.

Now if we wanted to convert that into number-needed-to-treat, since we talked about that a little bit, we said number-needed-to-treat is the number of people you need to treat to benefit one. So this, in this case, we had five out of 100, so that would be one in 20. You just divide five into 100 to get number-needed-to-treat there. You take the absolute difference between the groups, divide it into 100. That gives you the number needed to treat. So for every 20 patients that I treated, one died. I never knew that, and I

stopped doing this when this study came out and so did everybody else. But here's an example of where I, personally, and all my colleagues were treating people and hurting them. Well, we didn't hurt all of them, but we hurt enough of them, and we stopped using this. Does everybody see how the observational data led us the wrong way and the randomized control trial brought us back to the right way? So you have to have the right study, the right study design. It has to be done well. We haven't really in detail about how you do the study well, but that's what you've got to have or you're going to have a situation like this, and this has happened repeatedly over the history of healthcare.

So here's that Vietnam number again, 60,000 American lives lost, and we probably lost about 63,000 people due to the use of these drugs. Now most of the time, this type of information doesn't get presented in this way. So you can get some sort of a sense of it. You just hear, well, I guess there was a problem with that drug and people don't use it anymore, and you don't really get anymore information. But underneath that story, we're talking about personal lives that are gone and their families that are suffering because they lost loved ones that they didn't need to lose and that's what happened to this.

Many of you, I'm sure, already know this story. This is the hormone replacement story, and again, I was practicing

medicine doing this, too. Years ago, it was known that, if a woman has a heart attack, she is twice as likely to have another one, if she doesn't take hormone replacement therapy, but these were observational studies. There were about 20 of them, and they all showed that women who took hormone replacement therapy after a heart attack had half the rate of heart attacks that the women who didn't take that hormone replacement therapy. Anybody have an idea of why that was when, in fact, it wasn't true? I mean, the drug was not preventing this. The drug was not preventing it, but the studies all told us that there was a 50% decrease. So what do you think was going on? Any ideas?

MS. STRITE: And say your name first.

MR. STUART: And I'm also going to give you a hint, choice. Okay. So anybody want to take a shot at that?

DR. BROWNSBERGER: Well, they were not selected -- Keith Brownsberger. They were not different populations that you were looking at.

MR. STUART: Correct. Correct. In other words.....

DR. BROWNSBERGER: And some of those women were probably more -- that took the drug were the ones that were more careful about their health.

MR. STUART: Correct. This has got a name in epidemiology. It's called the Healthy User Effect, which means that the women who take care of themselves, just as he

said, were likely to do all the things they were told to do, which were watch your blood pressure, watch your diet, watch your cholesterol, watch your weight, and take this hormone, but the hormone was actually not benefitting them. It was the lifestyle, just as you said. They were different groups. Had they randomized people, they would have found different results.

So here is, basically, what we have. We have a cabbage-eating, jogging lady who chooses to take HRT, and we have a bon-bon eating, sedentary, smoking, non-HRT user on this side. Now these should not be compared, unless they're randomized, and if you randomize them, you'll distribute these different variables. Does everybody now get the importance of randomization in healthcare studies? And for the doctors here, I apologize because this is, like, 101, but there are a lot of lay people here that haven't had what you have. So I'm going very slow with these things, so the concepts get in there because we're going to speed up later on and get involved in talking about what policymakers can do about this, but that's at the end of the day.

And I think Sheri will give you an even more dramatic example of this with the Salpeter study.

MS. STRITE: Although, I do want to make a comment, and it's on the hormone replacement therapy story. And Mike and I were doing a training program a number of years ago, and a

young woman came up to us who was in training, I think, to get a Master's in epidemiology or a Master's in public health. And how many people here are with the VA? Just the Colonel. Anyway, she was with the VA system, and she came up to me -- and we had just talked about this -- and she said, I am so excited because, she said, I'm going to be working with a database that is so huge, and now, I can do all kinds of research with that database, and I said, okay. And I said, databases can be helpful for answering certain kinds of questions, and they can be very important for hypothesis generating, and they can lead you astray.

And one of the things that Mike and I are very concerned about, right now, in this country is that we see CER, comparative effectiveness research, as the latest buzz word that really is about what we're talking about here, just with a different flavor in terms of its new name and a lot of government funding and a lot of it, which is going to go to doing a lot of database research, trying to prove cause and effect therapies. Well, that's a big problem.

So she came up to me though, and she said, but I'm going to have millions of patients; therefore, because of the size of this, I'm going to be able to get at truth in this way. And I said, well, actually, you could add, you know, 50 billion women in that database because it was true that those women, in fact, were avoiding the second heart attack.

But the key point here is -- we talked about cause and effect -- it is not true that the hormone replacement therapy caused that outcome. So again at the end of the day when we're looking at therapies, a big question is always cause, and effect and not all associations are causal, as we know, but it's easy to get fooled, especially when numbers are big and questions are important.

So this is actually an interesting case because this a very serious change in medical practice. Mike talked, in various ways, about hormone replacement therapy. Well, this study came along later, I believe, in 2004, and it's *Salpeter, et al*. And you actually do not have these slides, just FYI. So this is just kind of a sit back and enjoy the show-and-tell.

Mike and I became aware of this study because there was a commentary published in the *American Family Physician*, which is, quote, the peer review journal of the AAFP, and the journal is supposed to be preferring original articles from experienced clinicians who are writing evidence-based authoritative clinical reviews. Well, in fact, we got a commentary on something that was published in *The New England Journal of Medicine*, I think. Oh, shoot. Sorry. No. The *Journal of General Internal Medicine*, again 2004. This was a meta-analysis, and the person writing for the AFP, Dr. Welbury, her commentary is, "Does beginning hormone therapy

earlier decrease mortality?" So this was an effort to try to see if there was a select group of patients that would, in fact, benefit from HRT after we saw problems from the Women's Health Initiative.

And so here, we have the citation and then here's just a little bit of a background here on this before I get to kind of what our analysis did. The analysis, again, was to discover how the age at which women began hormone therapy affects mortality. And when you do a systematic review, you're supposed to have a question that you pose to the literature to try to capture all potential studies that may be useful in answering that question. So this is a systematic review. And so in doing that, they have to have criteria for the studies that they're going to evaluate to include in the review, and they only chose randomized control trials, to which we would say that's a good thing because our clinical question is over a therapy, and randomized control trials are going to be the most important studies, potentially, we can look at, the most likely to be valid when we're dealing with an efficacy question. They wanted trials that lasted, at least, six months or more. Does anyone have any comments about that?

It might be short to show mortality, so that might be an issue. And then they had selected -- part of their criteria was there had to be, at least, one death reported. Comments

about that?

DR. BROWNSBERGER: Be more common in the older age group than in the younger age group, death.

MS. STRITE: That could happen. Also, too, we're neglecting non-mortality, right? So it's like we're choosing the outcome that we're trying to study, so there were some real problems with that. The outcome measures that they were interested in were total deaths, deaths resulting from specific cause, and all-cause mortality. Any comments about that as a composite? Oh, I guess it isn't a composite. I'm sorry. Those are the various outcome measures.

The age cutoff was 60 years, to determine whether you were considered older or younger. And then Welbury writes, "The authors conclude there may still be a role for hormone therapy use in younger women beginning shortly after menopause." Now, I realize that she's just repeating what the authors are concluding, but if I am a family practitioner and I am reading the AFP, I'm going to look at that and say, oh, this is advising me that this may be important information for me to attend to, and I may be changing my prescribing practice, and in fact, we know a lot of people who changed their practice as a result of this.

MR. STUART: Well, let me give you just one example. I used to teach a course at the University of Washington for mid-career healthcare leaders, and it was really a quality

course, a course in quality improvement, but I did the evidence-based medicine part.

And one year, I had, by chance, several gynecologists who are the heads of the OB/GYN sections in their systems, and I asked them about this study, and their response was yes; this study affected our group, and we are now treating younger women with hormone replacement therapy.

Now, we haven't gotten to the punch line from this study, but you will see that you might have some questions after Sheri goes through this study. And I quickly put this study on my computer and showed it to everybody in that room, just so these two gynecologists could see this. I don't know what happened after that.

MS. STRITE: We are so technologically challenged.

MR. STUART: No. It's actually cutting out.

MS. STRITE: Next time, we'll have Mike wear the dress. That will take care of the tie problem.

So how many people have heard of DynaMed? Dyna is a great source; yeah (affirmative). Brian Alper is the Editor-in-Chief, and he, like, started it in his basement, and he tries to systematically roundup all the information on primary care, and he got (indiscernible - voice lowered) at Besco (ph), which is where I mentioned that we're participating on the editorial board and does an amazing job, again not perfect. There are caveats there, but he's got a whole crew

now of people that do critical appraisal.

And anyway, Brian wrote us, and he said, you know, I'm really concerned about this study. He said, I'm looking at this thing, and he said, I think this is going to get past some folks, and it's going to end up potentially changing practice, and I've found some flaws, and so how would you like to write a letter with me?

So here's what Brian found. His first thing that he noted was the analysis was not based on a woman's age. It was based on the average age of the trials, and because Women's Health Initiative was so huge, you had 5,000 women that were misclassified, you know, younger versus older. So that was a problem that he found.

The other thing that was very curious, in addition to the choice of one death in a trial as being one of the criterion, they actually weighted by deaths. And if you're going to do weighting in a systematic review, usually, what you see happening is they're weighting by the study population. Now, there are some issues with that, but to weight by death is very, very unusual, and obviously, what that's going to do is give prominence to some very unusual populations, one of which was an ovarian cancer patient study and so that got an overemphasis in terms of the study outcomes.

But then here's what we found. Mike found the tables in the study, and he said, wow. He said, just check this out.

So now, I'm going to show you.

So we had -- okay, Forest Plots. Everyone remember our Forest Plots before? And in this particular one, everything on this side favors treatment, and everything on this side favors control, and this is a meta-analysis of the women, the trials that were identified of the women that were greater than 60 years of age.

Now, what I want people to do is I want you to just concentrate on the black dots. Don't worry about the confidence intervals, but just look at the black dots. Those are called the point estimates. Those are the calculated results of the study. What do we see here in terms of favors treatment versus favors controlled?

COMMISSIONER HIPPLER: Favors control.

MS. STRITE: Favors control. Okay. So in other words, that means that, for these women, hormone therapy was not providing benefit in terms of preventing mortality.

Now, we've got the second group. We've got the younger women, and what do we see in terms of the black dots? Favors treatment. What do we know about this? If it favors treatment, what else do we know about that?

UNIDENTIFIED MALE: It doesn't touch the line.

MS. STRITE: It doesn't touch The Line of No Difference. Therefore, this is telling is what? That it's statistically significant. Right. So on the basis of this, wa-la. They

concluded it has a place in younger women. Well, let me show you something.

We noticed that a lot of the studies in the younger women group were unblinded. None of the studies in the older women group were unblinded. What I did is I put red dots on all of the unblinded studies. Where are the unblinded studies? Favors treatment. And this is going to be a really, really important point that we're going to start raising every now and then and that is that, when you've got bias in a study, it tends to favor the intervention of interest. It doesn't always, but we're going to show you -- and we've seen study after study after study after study where you get a picture like this. Okay. Now when I take the red out, what do the black dots look like? It favors control.

And so what we would be doing, if we did this meta-analysis, is that we would be critically appraising these studies, looking at more than just blinding. We'd be looking for other biases, too, and we would only be including in our meta-analysis the studies that passed a critical appraisal screen. So only on the basis of blinding do we see a big change here.

Now just for fun, I'm going to circle the two studies in which the outcomes are remaining favors treatment. Now, I'm not going to say you can't do a valid small study, but it is more prone to chance effects, and the likelihood that there is

more bias in there is probably pretty high. One of them has a total of 26 patients, 13 patients in each arm. So my guess is we probably would have tossed that study and that leaves just one study remaining, 164 patients, and maybe those outcomes are due to chance, too. You know, who knows?

So anyway, even though we did not do a full critical appraisal on this, we think this is fairly compelling and supports that there were significant problems with this study in terms of their not doing critical appraisal work, and again, seeing that bias tends to potentially favor the intervention.

COMMISSIONER URATA: Question.

MS. STRITE: Yes.

COMMISSIONER URATA: Bob Urata. When you looked at this, did you look up all individual studies?

MS. STRITE: No. We didn't.

MR. STUART: Well, actually, afterwards, I did.

MS. STRITE: You did?

MR. STUART: Yes.

COMMISSIONER URATA: How do you process this to find the information that they were unblinded and stuff? Is that just stated in the.....

MS. STRITE: Oh, they had a table.

COMMISSIONER URATA: They state that in.....

MR. STUART: They state that.

COMMISSIONER URATA: .....their article?

MR. STUART: Yes.

MS. STRITE: Yeah (affirmative), and not even in that. Just in the Salpeter study, they have two tables, and the tables had a column for whether or not it was unblinded or not.

MR. STUART: With some other study characteristics, but I paid attention to the blinding as one, as an audit really. And then after we did this project, I went back and looked at the two remaining good studies -- they were not good studies -- the two remaining studies showing benefit, terrible studies. They were flawed, fatally flawed studies. So we, in my opinion, have no evidence from this study that hormone replacement therapy benefits younger women.

MS. STRITE: So let's all just stand up for just a minute and stretch. It's my way of keeping us all engaged physically. Okay. And then you can sit back down whenever you want. So I've talked about -- you know, we've talked about the training issue, and one of the points I want to make is that we've got a lot of people who are working in healthcare decision making who are not trained in critical appraisal, and those are the people that go on to be researchers, medical editors, peer reviewers, et cetera.

And to just kind of add another little story here, Mike and I want to try to get to everyone we can to try to help

improve the situation for patients. Someone asked me, they said, well, you should be talking to the scientists, and it's like, yeah (affirmative), we want to talk to them, everybody, but we also wanted to talk to the medical editors. So I did some digging around on Google, and I found the World Association of Medical Editors, and we looked at their criteria, and they did have criteria, but it was things like clinical relevance, must be clearly written, concise, blah-blah, and not a single word about the quality of the evidence.

And so we contacted them and said, we want to help you. We will help you for free. You know, we've got tools. Take advantage of us. It took six months to even finally get the right person to respond back to us. Do you know what the answer we got was? It's true; we have huge variation in our medical editors. It's true; we don't have these criteria. We should, and you're right; there is a big variation. And that was it. And that was it. So we've got a big problem in terms of the pool of people that are involved, in some way, in our medical information problem. Okay.

Now, what I want to do is, because the text on this is going to be small, let's go to slide 65, and we've got an abstract here, and this is going to be an abstract that was based on the Vigor study. How many people -- and I'm kind of sight challenged, so I need hands up high. How many people have any familiarity with the VIGOR trials? Okay. What's

notable about it? Anybody want to comment on that? What were they studying? They were studying Vioxx. It was compared to what drug? And we allow cheating here, so you can read ahead. Anything.....

UNIDENTIFIED MALE: Naproxen.

MS. STRITE: Naproxen. Very good. That's right. Compared to Naproxen. Now, one of the things that's important about the VIGOR trial is that was, quote, a pivotal trial. To be a pivotal trial means that this is the study that the FDA used to base its approval on.

So what I want everyone to do -- and if this is word salad to you and confusing numbers, don't worry about it. So don't feel stressed about it, but I want everybody to have just a chance to look at the abstract, and my question, as I start, is going to be, we've told you what clinical outcomes are, so what are the five clinical outcomes?

UNIDENTIFIED FEMALE: Mortality.

MS. STRITE: Mortality.

UNIDENTIFIED MALE: Morbidity.

MS. STRITE: Morbidity.

UNIDENTIFIED MALE: Symptom relief.

MS. STRITE: Symptom relief.

UNIDENTIFIED MALE: Quality of life.

MS. STRITE: Health-related quality of life.

UNIDENTIFIED MALE: Functioning.

MS. STRITE: Functioning. Yay! Okay. So what you to do is I want you to look at that abstract and then I'm going to have you tell me what you identify as the clinical outcomes that are mentioned in this abstract and then what I'm going to want to do is have a discussion about the most significant clinical outcomes here and talk a little bit about results. And again if you're not trained clinically, you know, just hang in there for the ride. You will have, I think though, the ability to, at least, identify the clinical outcomes and then the rest of us can engage in a discussion about what are the results telling us and talk about some interesting things that happened as a result.

(Pause - review of abstract)

MS. STRITE: So again, don't worry about the math part too much here. Does someone want to tell me, what's the first clinical outcome that you see here?

UNIDENTIFIED MALE: (Indiscernible - away from mic)

MS. STRITE: Sorry?

UNIDENTIFIED MALE: Gastrointestinal bleeding.

MR. STUART: Well, events. Yes. There were three. There was bleeding, obstruction, and perforation, but yes, bleeding.

MS. STRITE: What's one in the first line? They don't state it quite in the way that we did, but.....

UNIDENTIFIED FEMALE: Rheumatoid arthritis.

MS. STRITE: Right. Their -- similar efficacy against rheumatoid arthritis. Okay. We’ve got GI bleeds. What else do you see that’s a clinical outcome?

UNIDENTIFIED FEMALE: MIs.

MS. STRITE: Myocardial infarction. Okay. So.....

UNIDENTIFIED MALE: The mortality was (indiscernible - away from mic).

MS. STRITE: And mortality was what?

UNIDENTIFIED MALE: Equal.

MS. STRITE: Yes. Mortality was equal between the groups. Great. Okay. Now, does someone want to talk about any of the results they see in any of these clinical outcomes? Let’s focus on the complicated GI effects. So complicated gastrointestinal effects, we’ve got perforation, obstruction, severe GI bleeding.

UNIDENTIFIED MALE: (Indiscernible - away from mic)

MS. STRITE: Okay. Let’s drop down a little bit more though to the complicated events. We see a relative risk 0.4 and then the 95 -- we talked a little bit about confidence intervals. The 95 CI just means there is a five percent (indiscernible - voice lowered) chance outside of that range. And we see that as statistically significant. So what do you think about the relative risk of 0.4? It converts to relative risk reduction of 60%. Say what?

COMMISSIONER HIPPLER: That seems impressive.

MR. STUART: It seems impressive.

MS. STRITE: It seems impressive, doesn't it?

MR. STUART: And that's what everybody said.

MS. STRITE: Because relative risk -- so relative risk of 0.4, you just subtract that from one, make it a percentage. That equals 60% relative risk reduction. So it's just a flip side. And 60%, I mean, if I have a newspaper to sell, 60% is a number that's going to drive that. If I have a drug to sell, 60%, that sounds good, doesn't it? So what do people think about that?

MR. STUART: It sounds good.

MS. STRITE: It sounds good, right? Okay. Take a look at the MI. In fact, someone actually just read that aloud.

UNIDENTIFIED FEMALE: "The incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1 percent vs. 0.4 percent)."

MS. STRITE: And that sounds good, too, the way that's stated, doesn't it?

MR. STUART: It does?

MS. STRITE: Sure. The incidence was lower, and you know what the company said? Well, yeah (affirmative); of course, the incidence is lower. It's just that Naproxen has a protective effect, just like some of the other non-steroidal anti-inflammatory agents. So notice how that's framed fairly

positively? Well, okay. So that study comes out in November, I believe, of what year I cannot find.

UNIDENTIFIED FEMALE: But it's the reverse, right?

MR. STUART: It's reversed; correct.

MS. STRITE: Gee, how about that?

MR. STUART: It is odd that they flipped the presentation mode; that's right.

MS. STRITE: Unless you're a critical thinker. Okay. So here's what happened. After the study came out in the hallways of Group Health Cooperative, where he worked, Mike kept running into GI specialists who were very excited. "A 60% reduction in ulcer complications, they enthused. Many others all over the world were also clearly impressed" -- and this is why I'm reading this because I can never believe the numbers; they always shock me -- "because worldwide sales for Vioxx in 2001" which I believe is when this was published -- sorry, but I can't this for the moment. I'm pretty sure this was, like, November 2001. "Worldwide sales were $2.6 billion." Billion. And the reason why I read this is because I always reduce it, in my mind, to millions because I can't even cope with a number like that. "Up 18%" -- okay. Oh, November of 2000. Sorry. "Up 18% from 2000, and the medicine claimed half of all new prescriptions in its class of medications. For the first quarter of 2002, sales were $650 million compared with $485 million in the first quarter of

2001." So uptake, I mean, was huge.

And I'm ignorant to this, and in talking with Mike about this, he said, you know, we really didn't have anything that was going to help the serious GI problems that people had in taking NSAIDS. And so it wasn't just the number that was big. There was excitement because there wasn't anything else out there available, but what's the punch line on this 60%? Does anyone know? Let's go to slide 67.

Can someone interpret, for me, why I've got these red arrows here and why I have these things circled in red? What's this telling us?

COMMISSIONER URATA: You have to treat 125 people to avoid an upper GI event.

MS. STRITE: You have to treat 125 people to avoid one, taking Vioxx as compared to Naproxen. Well, where is the 60%? So what just happened there?

UNIDENTIFIED FEMALE: Needed to know how often it was happening.

MS. STRITE: Well, that's telling you.

DR. THOMAS: It's the difference in the relative risk and the absolute risk.

MS. STRITE: Teach us about that. What is that all about? You're right.

DR. THOMAS: The absolute risk.....

MR. STUART: Light on and name. Thank you.

DR. THOMAS: I don't know that I can teach you. I'm Kim Thomas. The.....

MS. STRITE: I'll help you.

DR. THOMAS: So the relative risk isn't taking into effect the baseline rate in the population, whereas the absolute risk is thinking about that, I think. Am I saying that right? Absolute risk is thinking about that baseline.

MS. STRITE: Yes. And so let me bring it to you in terms of being a consumer because we can all understand this as consumers because we use the concept day-in and day-out, and a lot of people working in healthcare are not taught this.

If I have a little store and everything is 90% off in my store, are you going to hand me over your credit card or do you have a question for me?

UNIDENTIFIED MALE: Ninety percent off of what?

MS. STRITE: Ninety percent off of what is the question; that's absolutely right. And what happened was we begged the question in that abstract because the answer was it was 60% of less than one percent. Sixty percent of less than one percent of people were going to avoid a serious GI outcome taking Vioxx as compared to Naproxen. That's why this number here is over 100 because it was 60% of less than one percent. So that's actually the real size of benefit.

Now, someone discuss, for me over here, all thrombotic events; what's going on there? So say more. I guess, name

and.....

UNIDENTIFIED MALE: (Indiscernible - away from mic)

MS. STRITE: The risk is higher than the benefit, right, because, for every 103 people that I'm treating, one of them is going to have a thrombotic event because of the Vioxx as compared to the Naproxen, but again, the company spun it, but as good critical appraisers, our question is, well, wait a minute; is it actually true that Naproxen confers some kind of benefit over Vioxx, and even if it does, we've got a bit of a problem here, don't we? Yes?

DR. THOMAS: Can I just add one other thing that a lot of people who do critical appraisal add in, too, is the source of funding? It would be interesting -- of the PP icons that John and I were talking about, it would be interesting to know, were any of those people on the boards of Pfizer or whoever?

MS. STRITE: Merck, actually, in this instance.

DR. THOMAS: Merck.

MS. STRITE: Yes. And that is an important question, but again, I say Mike and I pay attention to that. So we definitely make note of not only the funding, but the involvement because, sometimes, it's a matter of funding, but it's, you know, through a research grant that's unrestricted. But again, everybody in research has got a bias, too, so we try to let the evidence direct us. And so.....

COMMISSIONER URATA: Question.

MS. STRITE: Yes?

COMMISSIONER URATA: Bob Urata. So why did the FDA not see this? Is it because they don't.....

MR. STUART: The FDA did see it. This is one of the examples that is a mystery to us. They actually were on this faster than anybody else, but they just didn't publish this in a way that got out. We went into the FDA site to get that information Sheri just showed you. It was there.

MS. STRITE: But yet, they approved it. And one of the problems with this is.....

MR. STUART: But I want to just say one more thing. Drug companies see their customer as the FDA, not payers. Payers are secondary customers. We've spent the last two years talking to a lot of drug companies about, you need to see who your real customer is; it's not the FDA. The FDA, yes. It's a hurdle. Your customer is there and the patient. So some of them are getting it, but in fact, anything you can get by the FDA, as a drug company, is the way the game is played. So they will present data to the FDA with composites for efficacy and then single outcomes for safety because there is no difference because they're small studies, and the FDA, sometimes, just let's it go through. I don't know why. The FDA is like every other group. They have good groups and groups that are not as sophisticated. So there is a lot of variation, even within the FDA evaluations.

The FDA is not a 100% reliable entity, and they will approve Avastin for advanced breast cancer without a shred of evidence for benefit and say you're required to do a couple more trials, and they knew the harms -- ulceration of the skin, all kinds of reactions -- but they approved it. And then years later, it was withdrawn because the other studies didn't confirm it because they were poor studies, too, and we never did have any good evidence for Avastin helping, but we have good evidence for Avastin hurting people with advanced breast cancer. Yet, the FDA was very, very passive, really, from the perspective of many payers we've talked to.

So the real customer is us, and drug companies are not really getting that. And so the FDA, why they didn't make this information available, we don't know, but we found it in the FDA site, which is not an easy site to use, I will tell you.

MS. STRITE: I was listening intently, so I missed what I was going to say. So good; take it over.

COMMISSIONER CAMPBELL: Keith Campbell. Is it true that the FDA is basically funded by the drug companies through the fees that they're assessed?

MR. STUART: Well, the FDA does -- yes. It is true that the FDA is financed to a large -- I don't know the percent though, but that is true.

MS. STRITE: For approvals. So what's happened is safety

has actually suffered some because the money from the drug companies, it's pooled, but that's giving more resources toward efficacy and so safety has become more neglected because of that.

The other thing I'll mention about the FDA, too, is that, with pivotal trials, sometimes, the FDA isn't as in control of the outcomes and certain choices. And so sometimes, some weird outcomes we see, it's because that's been FDA-driven, and I think that part of that harkens to what Mike says that, even within the FDA, there is variability in these skills like everywhere else.

MR. STUART: Well, there is a lot of statistician skill and sometimes not as much critical appraisal skill, which are different entities. If you are a statistician, you could be a miserable critical appraiser. There are two big things you've got to look at. Is the study threatened by bias, which is not -- does not require lots of statistics, or is it not? That's the first question.

Then when you're looking at the results, that's where the statistics come in. Are they using odds ratios, confidence intervals, P-values? Are they -- you know, that's all statistical stuff, but unless you know the study is valid by whether the randomization was done well, the performance of the study was done well and balanced and everything was equal, that's not really statistics. That's just basically bias

evaluation or critical appraisal for the quality of the study, almost without statistics. Then you move to the statistics when you look at the results, but you shouldn't even be looking at those results, unless you know that the study is valid, which does not require a statistician.

So in the FDA, there are some groups that are heavily statistician laden, but some of those folks don't have the critical appraisal skills of looking for selection, performance, attrition, and assessment biases. Yeah?

CHAIR HURLBURT: Just a comment on FDA. Having spent a lot of years in the public health service, I have a lot of colleagues there. They're good folks, and.....

MR. STUART: Yeah (affirmative).

CHAIR HURLBURT: .....they're good scientists, and they want to do a good job, but they are subject to mixed pressures. And the example, as cited, you know, over and over and over and over again, that we were protected from phocomelia from Thalidomide in this country back in, whenever it was, the '70s because it didn't -- it wasn't approved and that's the way -- and it was in Europe -- the process should work, but the FDA, yes, they are heavily funded by pharma, but they're also subject to the pressures. Probably very notably, *Wall Street Journal* often will say that they will have a, quote, expert -- and I know Mike and Sheri will talk about experts sometime during the day -- come in and say, why can't

I get this new hotdog cancer drug and save my patient's life? And I could, except for the FDA.

And I think the takeaway is to recognize FDA approval does not say this has been subject to a rigorous evidence-based objective process. They try to do the best they can, but FDA approval or Medicare approval, CMS approval does not mean that. So those who are deciding benefits, coverage, and clinical decisions need to have this kind of rigor in our thought processes that we're talking about today.

MR. STUART: Well said, Ward. That's exactly right.

MS. STRITE: And another problem with this is that the drug company did not have evidence to really state that it was true that Naproxen was protective, and if we had had a placebo arm in this trial, which really is the ideal -- the ideal is to have, you know, whatever your therapy of interest is, placebos or usual care or something like that. The best of class is a comparator. And in this instance, if placebo had been included in this, it would have been very clear that it wasn't the Naproxen that was protective at all; Vioxx was harming patients. But as critical appraisers, we get to do critical thinking and look at that statement and say, wait a minute; it's very possible that that reverse is true, that Vioxx is actually harmful.

And so again, here, we have our comparison. We had about half as many people that had heart attacks and sudden cardiac

death that otherwise would not have had, within a four year period, because of taking this agent.

Oh, I'm sorry. There is one other thing I wanted to mention. So one of the things that's kind of fascinating to me is that the VIGOR trial, we had actually spent, as Ward pointed out, a lot of time and energy helping various groups, including where we worked, keep Vioxx off the formulary, based on this study. And it was only a few years ago that I went back because I was looking at this study that I'm going to tell you about that is looking at just abstracts, and I said, I wonder if that information was in the abstract or was it hidden in the body somewhere, and by gosh, we just saw that, in fact, it was in the abstract. Well, this is fairly compelling because, as Mike and I would do our pre-tests, one of the things that we will ask physicians is, when you read an article, what do you read? And about 75% of them only look at the abstracts.

Now, the abstracts are very useful because you may be able to determine clinical relevance and whether you want to spend time with it. Occasionally, you can find problems like with the VIGOR trial. Sometimes, you can determine a study is not valid, but you cannot determine that a study is valid by only looking at the abstract. And then to make it even scarier, somebody named Pitkin took a look at the abstracts and compared them in some of our favorite journals here,

right, JAMA, *New England Journal of Medicine*, *British Medical Journal*, *Annals of Internal Medicine*, *Lancet*, and the *Canadian Medical Journal* because they happened to be Canadian, and what he found was that 18% to 68% of the time the information in the abstract could not be verified in the body of the text. There will be errors. Mike and I pay attention to both text and tables. You'll see numbers that don't mesh or match. So big caution here: If you are just looking at the abstracts, you really need to go beyond, if it's a study that you think is something that's going to change your practice.

MR. STUART: Okay. I want to read a Letter to the Editor of *The New England Journal* written by Steven E. Nissen, who is a cardiologist at the Cleveland Clinic, and I'm sure, Thomas, you probably have heard of him. He has thrown wrenches into many, many drug situations; let me put it that way.

Here's the letter: "In the original article, the APPROVE investigators reported event rates using an unusual censoring rule." The unusual rule was this: "Events" -- and these are blood clots -- "were excluded if they occurred more than 14 days after the study drug was stopped." So the investigators, in planning the study, made up a rule -- and these are called censoring rules -- that, if somebody has a heart attack or a blood clot 14 days after they have stopped taking Vioxx, it couldn't have been due to Vioxx. Now, why they arrived at that we have never been able to figure out or get information

on. “All data in the new report” -- the new report being the one I, Steve Nissen, am submitting with this letter; this is just part of the letter -- was “assessed by a conventional intention-to-treat analysis,” which I know you don’t understand now, but that just means that he put the data back in.

In other words, if you randomize people to a group, you must analyze them in that group. It’s kind of one of the agreed upon rules of many, many editors and evidence-based specialists and scientists throughout the world is that intention-to-treat analysis is the preferred initial analysis. At any rate, he put the patients back in.

Now, I’m going to show you what happened. Now, this is from the original study and so this if from the *Journal*. This is from *The New England Journal*. And you can see two curves, and I’ll ask people to comment on anything else they want to comment on as they look at this, but the one that I want to point out first -- you can find other, hidden features in this slide, if you can, but essentially, the main one is that, at this point, 18 months, we see the cumulative incidence of confirmed thrombotic events being higher in the Vioxx group than in the placebo group. Can everybody see how the curves diverge? Okay.

Now, this was important because people, at this point, were already suing and so anybody with a blood clot up to here

was very likely contacted by a lawyer and invited, well, to join in. Okay.

So here is Nissen's re-analysis where he put the data back in. Where do the curves separate? Three months. So you see how dramatic the difference in the outcomes are, depending upon how you actually do your study? This is just one little piece. This is an analysis problem that we see in Kaplan-Meier curves, which we're not going to have time to go into in detail today. But at any rate, does everybody see what happened here is the authors took people out of the study and Steve Nissen put them back in, and the results were different, and it costs millions of dollars. Was it Merck, I guess?

MS. STRITE: Yes.

MR. STUART: Yeah (affirmative). That had to pay out. Okay. So.....

MS. STRITE: As compared to the billions that they made.

MR. STUART: Right. It was a small slap, but still, in my mind.....

MS. STRITE: And I have a question for the group, too. So on that example that Mike just showed you, what if they hadn't had that censoring rule? What's -- oh, how do I want to phrase this? If they hadn't had the censoring rule, what should they face in terms of their concern that the deaths couldn't be attributable to the drug, if it had been over two weeks? I didn't ask that really well, so I'm hoping someone

reads my mind. Will you clean up my question?

MR. STUART: Well, I don't even know what it was.

MS. STRITE: Okay. So I'll just make the point. I'll just make the point. So the bottom line is, to have a censoring rule like that -- and I -- Mike and I actually did a little scenario. We sort of played out. It's, like, well, why did they do that? And we could actually come up with some innocent logic for them creating the rule like that. It's, like, maybe they had test data with animals that said they didn't see, you know, any half-life after 14 days, blah-blah-blah, but that doesn't matter because, with the study, you're comparing differences between the groups, and if they had just left those people in there and not had that censoring rule, if they were right, that would have come out in the course of the study.

MR. STUART: I see. Yes. Yes. Okay. So now, I'm going to switch to the evidence about the evidence we've presented to you. In other words, I'm going to show you why we're saying what we're saying and that is critical appraisal -- and we haven't given you anything but a few samples. We haven't given you an organized approach. If we were teaching you to be critical appraisers, we would -- as we went along, we would summarize and put it into a way for you to do it tomorrow, but we don't have to do that, since you don't have to do this tomorrow.

The low quality studies are likely to overestimate or amplify or distort results by, sometimes, up to 70%. We usually say 30% to 50%, and a reasonable number to remember, to be conservative, would be up to maybe around 50%, just to be safe. But really, that's for each, individual bias. So if you've got a whole bunch of biases, you've got a real high chance of having really non-trustable results.

So we have a bunch of studies that are telling us things are more effective than they are or that are telling us they're effective when they are not effective. And so what I'm going to show you now is some of the science that we used to figure this out. Other people have, too, but this started many, many years ago.

Here's one of the early 1983 studies done by Chalmers, who is a pioneer in evidence-based medicine. And what he did was a very unusual study, and he looked at mortality from myocardial infarction, acute MI, and he divided all the studies into these three categories: studies that were not randomized; studies that were randomized but did not have concealed allocation, which means that they did create random numbers, but they didn't really keep everybody from knowing which group they were directed into. So people could direct traffic into the group, possibly. And then the third group is they're randomized and they have concealed allocation. I want to emphasize that these are just a couple of the things we

look at, but this is a starting place because it's very easy to grasp.

The non-randomized studies show definite benefit for the cardiac inventions. You see almost a 10% difference here, so that we basically get lots of benefit if we do bad studies. Okay. So if we do poor studies, we could show benefit very easily. If we do a little better study, we could allocate -- it's -- we randomize. All we do is randomize. We drop that ten percent to around five percent. And then if we do even a better study in this particular arena, we have no benefit. There is no difference between the intervention and the placebo. So we see how the distortion of results can occur with the quality of the study.

Now, I'm going to show you some more slides that are going to be a little more complex, and as long as you get the bottom line, that's all we care about.

This is from Cochrane. Up here on the left, we have what's called Risk of Bias. Cochrane divides all studies into three categories. They're either -- well, each bias, I should say. The biases could be -- there could be a high risk of bias for that entity, like randomization, generation of the sequence, for example, or it could be blinding. But let's just say that those with a high risk of bias are low quality studies or low quality parts of a study.

On the other hand, there are low Risk of Bias studies

which are higher quality studies. Does everybody follow that so far? Okay.

So now if we look at the high risk of bias studies, we look at this review of 61 studies and find that their relative risk reduction is about 50%, 0.52. So that's a 52% relative -- well, if it's relative risk, it's a 48% relative risk reduction, but let's just say it's about a 50% risk reduction. So that's favoring the treatment. But now, we look at the higher quality studies, and we only get half that benefit. We get about a 20 -- about half of 50 is 25. So we get about a 25% benefit, meaning -- this is just more evidence that the higher the quality the evidence the less benefit is likely to be shown, and the lower the quality of the evidence, when we look at the biases in a study, it's more likely to show a benefit of whatever we're studying.

MS. STRITE: I want to do just a terminology thing here, really quickly. When Mike says "benefits shown," what he is really saying is "benefit reported."

MR. STUART: Reported. Thank you.

MS. STRITE: And there are a lot of reasons why something may be reported as a benefit. That doesn't mean that it was a cause of the intervention under study.

MR. STUART: Yeah (affirmative). Thank you. Now here is -- there is a lot on this slide, and for the people who are actually interested in critical appraisal and having a system

to do it, I'm going to just make this real quick, but that wasn't the goal of today. But here's the way we do it.

We divide a study into Stage I, Stage II, Stage III, and Stage IV. The first stage is getting the groups set up, and you do that for therapies by randomizing. And the first consideration is generation of the sequence, and if it's done wrong, results can be distorted up to 75%.

The second thing we look at here is concealment of allocation of the randomization sequence, and we get, again, up to 73% distortion.

There is a third item we look at, which is the Table of Baseline Characteristics, Table I in most studies, and we look at the groups to see if they're comparable, age, sex, clinical, co-morbidities, things like that.

Now we're down to Stage II, the Intervention and the Context, and we call -- the first stage is selection bias. It's called performance bias. Again if there is not double blinding of the people who are interacting with the patients, the results can be off by 72%. And if we drop down to the third stage, Loss of Data, if you lose patients out of your study -- this is one of the toughest areas because it hasn't been studied very well, but we get about -- we've been able to find up to a 35% distortion. It might be more than that, but we just -- you know, the studies just haven't been done.

And then the Assessment Bias, which is how do they handle

the assessors, are they blinded, do they do the right type of analysis, and again, we get this distortion.

All I want to do is make the point that each bias that we find -- and this really -- this is about all we do to start with. We look for other items, but I mean, this is really what critical appraisal is all about is looking at these stages and these bullets to see how well they did, and if they did really well, then they get lots of points. And if they do really bad, they'll get a Grade U, which is we can't use it.

Now, the CAPPP Trial.

MR. KNAPP: Quick question.

MR. STUART: Yeah (affirmative)? Hold on.

MR. KNAPP: Can you briefly talk about, or if you did already talk about it and I missed it, the question about, what about whether studies are actually reported? In other words, is there some bias about a study was done or sort of done and then it didn't look like it was going to turn out right so.....

MR. STUART: Yes. Yes. There is data on that and that's one of what Cochrane calls -- Cochrane has that as one of their main criteria and that's a complete reporting. So in the protocol when you're doing a study, you write down a protocol and say we are going to do this, and we're going to do it this way, and these are our outcomes. But then when you look at the study that's published compared to that protocol,

frequently, the things they report are not what they say they're going to report and that's a very important point you just made and that is -- you have to ask the question, why didn't they report on what they said they were going to report? Well, the answer might be because they didn't get the results they wanted. So, it's true.

MS. STRITE: And I think you're talking also about studies just not being published, correct?

MR. STUART: Yeah (affirmative). That's what I thought he was starting. A lot of studies that turn out the wrong way for whoever is sponsoring it just never get published. They don't see daylight. They just get put away. And everybody is trying to make sure that comes to a halt. The studies have to be registered, but still, there are ways to get around that, and it is possible to hide studies that you've done. Hopefully, it's going to get harder and harder to do that, but yes; that's also another problem. That's a good point.

MS. STRITE: Although, I think that's very complex though, too. There was a survey -- well, first of all, a lot of investigators, if they get, quote, negative findings -- negative findings doesn't mean, like, a bad outcome, except if you're the investigator, I guess, it means that you got non-statistically significant findings. And there is a belief among researchers that, if they got non-statistically significant findings, now they don't have a publishable paper.

And so there was actually a survey that was done of editors asking them, well, is it really true or are they wrong, and they said, oh, no. In the survey, they say, oh, no; we're definitely willing to publish non-significant findings. But in fact, I think what the reality is, is that doesn't happen because what sells newspapers, right? The positive findings. The dramatic results.

And then the other thing, I think, there are some groups that do a very extensive job of trying to find what's called the gray literature, the stuff that isn't published, but sitting in file cabinets and all of this. And frankly, I think that, if we actually had an opportunity to look at that, I think a lot of it would not be valid.

So I think that's an area where, even though there are studies looking at the impact of unpublished studies, I think that there are some flaws in that, too. So I think that's a complicated area, and I think that, in terms of clinical trials, you know, the hope, by registering studies, is a hope, and I think that -- I don't know how much that's going to help either, so -- but you're right; that's an issue. Yes? State your name and turn on that light.

MS. PENWELL: Hi, I'm Vicki Penwell, and I'm a Patient Advocate. I'm not a medical professional. And it has occurred to me this morning that there is a difference, perhaps, between bias and motivation. And so if you're a

pharmaceutical company and the FDA is your audience, are you more likely to do a low risk -- a low quality study, thinking that, if you make Vioxx, for example, you're going to make billions of dollars before somebody gets around to doing a quality study that would question your original results?

MS. STRITE: The FDA's in control though of a lot of things having to do with the pivotal trials. So the drug companies don't have as complete control over them as I think that it would appear by looking at it. So that's one issue.

Do you want to answer that more?

MR. STUART: Yeah (affirmative). Sure. She's raised many issues, and they're good ones.

MS. STRITE: Yeah (affirmative).

MR. STUART: The first thing, I think, to say is that we don't have any proof that drug companies intentionally design deceiving studies. We don't know that. We do know that, basically, if you aggregate all the studies that are sponsored by drug companies, they tend to show better results than those that are sponsored by others. So that's indirect evidence that there is something going on.

Quite frankly, it may not be the drug companies' fault many times because the FDA requires them to do the things that, sometimes, appear to us that they're not very useful to patients at all. They're to satisfy statisticians or some other reason that we have been unable to figure out, but I

don’t think that we can say that we know of any case where they’ve thought they could get their money pulled in before they got caught. I just don’t think we know of that.

DR. BROWNSBERGER: Keith Brownsberger. A number of years ago, a new drug came out called Zetia, which lowered the cholesterol, and there were some good studies showing it lowers the cholesterol. And statins were already out, a number of them, and statins were tested both lowering cholesterol and some good studies showing that they decreased the rate of heart attacks.

MR. STUART: Right.

DR. BROWNSBERGER: So when they brought this drug out, they attached the Zetia to a statin drug and that’s the drug that they were selling. And all their studies were done with statin plus Zetia. So I asked a drug company guy, I said, when is the study coming out that shows that Zetia actually cuts down heart attacks, and he said, why would we do that? It’s selling just fine.

MR. STUART: That’s true. He picked an arena to work in where he could do that, but that isn’t always the case, whether you can. So -- but yes. In that case, in Zetia’s case, that didn’t really work so well for a lot of people who did critical appraisal because they said the same thing you did. Well, we (indiscernible - voice lowered) trial, and until we get that, we’re not going to cover your drug. But

you're right.

DR. BROWNSBERGER: But they made billions of dollars off Zetia.

MR. STUART: They did. They did.

DR. BROWNSBERGER: So it was good publicity, but it was not good science.

MR. STUART: That's right. Exactly. Thank you. That's a good illustration, actually. Okay. I want to show you one.....

MS. STRITE: And you don't have these slides for this particular example, and I think that's the last of the slides they don't have.

MR. STUART: Yeah (affirmative). I just threw this in because I want to show you how I got burned. I totally missed this, and a Brit picked this up.

This is the CAPPP trial about Captopril. Some of you may remember we didn't use to have ACE inhibitors, but when Captopril came out, this CAPPP trial was going to compare Captopril versus diuretics, together with beta-blockers to see if Captopril was more effective than this combo of diuretics and beta-blockers. And so it's a large study, over 5,000 people, and all we need to look at is the primary outcome, which is heart attacks, fatal and non-fatal, stroke, and other cardiovascular deaths. So that's the study. Does everybody see what the study is aimed to do?

Here's my critical appraisal in short form. I say, in general, the goal was that it's a randomized trial to compare the effects of ACE inhibition and conventional therapy on cardiovascular morbidity and mortality in patients with hypertension. And so here's the population. It's an open label RCT, which is a problem.

MS. STRITE: And open label, for those who may not know, means it's unblinded.

MR. STUART: Unblinded, except the evaluators were blinded. So it's not as bad as it could be, but it still got some threats right there.

In randomization, there was a generation of the sequence by a computer. That's good. They had sealed a number of envelopes. That, we say, is not good. That's called concealment of allocation through putting them in envelopes.

Now even if the envelopes are opaque and you can't see through them, so I can send this patient who I want to go into this group there, people can steam them open. They can -- and it's been shown that people actually do this. We had a hard time believing this at first, too, but then we had investigators tell us that they did things like this, and they were usually fellows who were participating in -- it was the University of California San Diego where we got the best insight when this guy said, well, you know, we have -- a lot of these patients are enrolled in study at night, and if they

go into the group with the intervention, they have to have this big work up, blood tests, x-rays. So we don't want to be doing that at night. So we just take the envelope, hold it up to the light. If it's for the intervention group, it goes to the bottom of the pile, and we keep going until we get somebody that's in the control group. And so at night, we don't have to do these work ups. And so what's going to happen? I mean, people that come in at night are probably going to be quite different than the people who come in to the day -- during the day and so this distorted their study. Well, this is a similar thing. We don't know how this happened, but look what happened here.

First of all, I'm just giving you my little critical appraisal of the study to start so you can see we look at the selection, the intervention, the follow-up, and the assessment of the outcome, those four phases of a study, and the -- I won't go through any more of this, but they didn't lose many patients, less than a percent, which is almost unheard of, which is really good. And they did an ITT analysis, which is good, and they did some things that weren't so good, but it wasn't a fatally flawed study. But if you look at the baseline characteristics -- well, first of all, let's look at the endpoints. The baseline characteristics in those who -- I'm sorry. These are the results.

Let's take a quick look at the results, and what we see

is a higher rate of stroke in the group that gets Captopril, and it's statistically significant. This 1.25 means it's -- the risk is 125% greater for a stroke if you take Captopril. Well, this almost destroyed the drug because the -- why would a cardiovascular drug cause stroke compared to the other drug? It makes no sense. We don't want a drug that's going to cause stroke. So Captopril almost went down the drain.

But when we look at the Table of Baseline Characteristics, which I did -- and I thought they all looked well-balanced, the two groups looked well-balanced. But the truth is that there was a slight difference in blood pressure where these red arrows are, just a little bit, but the blood pressure was a little bit higher in the Captopril group at the beginning of the study when they put the -- when they randomized them to the two groups. Why do you think there would be a group with a higher blood pressure than the other group? Why wasn't it the same blood pressure? Everything else was pretty similar, but -- and this is not a big difference, but it's a difference. What could explain that? Why would people with higher blood pressure be in one group than another?

MS. STRITE: Could be chance.

MR. STUART: It could be chance.

MS. STRITE: It could be chance, point being that randomization doesn't always end up in equal groups.

MR. STUART: Uh-huh (affirmative). So.....

MS. STRITE: You could have some over-balance due to a chance effect.

MR. STUART: So it could be chance. Have we got another idea?

COMMISSIONER PUCKETT: It wasn't blinded, so did they know they were taking a new treatment?

MR. STUART: Say that again, loud into the microphone.

COMMISSIONER PUCKETT: Jim Puckett. It wasn't blinded, so did the patients know they were getting a new treatment?

MR. STUART: Well, if the side effects tipped them off, that would be one possibility. It's possible; yes. That's always a possibility. Yes. It could have been unblinded, but that's not what this researcher discovered when he wrote his letter. Here's the letter from Sir Robert Peto.

MS. STRITE: Richard.

MR. STUART: Richard Peto. He's been knighted for many of the things he's done in medicine. So he's a knight.

And he says, "The small, but highly significant difference between the two treatment groups in the pre-randomization height, weight, systolic and diastolic blood pressures with respective P-values of ten to the minus four, ten to the minus three, ten to the minus eight, and ten to the minus 18 show that the process of randomization by sealed, numbered envelopes was frequently violated."

What he’s saying there, is this didn’t happen by chance. Even though -- this is really a low likelihood of chance. Somebody did something. And he says, “I think these envelopes were fiddled with.” And that’s what it looks like turned out later on is that one group -- it was a multi-center trial -- did do that and wrecked the study. So it falsely told us that we were causing stroke with Captopril when, in fact, we had no evidence from the good studies that there was an increased risk of stroke. This finally got ironed out, and Captopril didn’t get removed from the market, but it could have been and that would not have been a good thing.

So he continues on, “Sometimes, unsealed envelopes allowed people to be directed into one group or another.” And of course, Captopril was the new drug, and new is better. So let’s get these really high risk patients into the new drug group, so people would undo the envelope and say let’s get them into the right group. So they directed traffic into the Captopril group. Does everybody see what happened there? It biased the study and gave us the wrong results.

So here are the key points. Concealment of allocation is required to prevent selection bias; that’s one. Envelopes can be opened or steamed; that’s another point. And even though you statistically adjust things, you can still have distorted results, like we got in this study, and if they had done the randomization process in a different way, this wouldn’t have

happened. If they had gotten the sequence for the generation of the randomization sequence and put those into locked boxes and distributed by interactive voice response systems so patients could dial in and see which group they went to, which is now much more commonly done -- although, we still use envelopes -- this would not have happened.

MS. STRITE: So let’s have people stand up one more time and stretch. I do that, if I notice that people get a little bit low because I know this is fascinating stuff, but occasionally when you’re in presentations like this in groups, suddenly, the sleep fairy shows up and is a little overwhelming. So if I notice that, I may actually go up to people and tap them; that usually startles them. So if we just all stretch a little bit because this is interesting stuff and that sleep thing is so uncomfortable when you’re not in your own cozy, little bed. Okay.

MR. STUART: So we look at -- so back to the same statement, critical appraisal is really important. We look at the design. We look at methods. We look at the execution. We look at the outcomes and whether the performance of the study is going to be done in a way where we can trust the results, and we look at how it’s reported.

This I put up just because we went to Boise, Idaho a few years at the invitation of a Medical Director because they had the highest rate of spine surgery in the United States, or

actually as he put it, in the universe, and *Newsweek* did a front-page story about Boise's neurosurgeons, and the Boiseans didn't like it very well. *Newsweek* really put them on the map in a way that they didn't like. And so he said, well, why don't you come down and show them that they're doing a lot of unnecessary surgery? And we go, well, that's not going to work. If we show up down there and tell them their job, they're just going to kick us out, and how are you going to get them to even come to the meeting? Well, he said, well, we talked it over. I'll put them in the paper as having gone to a very important meeting and so they'll get their names in the paper for doing evidence-based medicine; that's good. And then we said yes. And the way we'll do this is we'll have the neurosurgeons pick out their best literature, and we'll go through it with them and see the quality of the literature, and we'll do it together. We'll teach them how to do it. Then they'll do it and then they'll tell us the quality of their evidence.

Well, we got down there, and we did that, and we had a lot of fun with them. They all wore cowboy hats, and they were.....

MS. STRITE: (Indiscernible - simultaneous speaking) cattle.

MR. STUART: .....very nice people, I have to say. And what happened was they couldn't find any article that was

valid, not one, for back surgery that they were doing, and they said, well, just because there isn't good evidence that it works doesn't mean that it doesn't work, and we said we agree. We would like to have evidence that worked, but patients should know that the evidence is not good for all these back surgeries you're doing, and for younger people, many times, they should be given the option to wait and see if it goes away without surgery, and they said, well, we don't necessarily disagree to that, but if we do that, they'll go down the street, and they'll go to another neurosurgeon, and he'll operate on them.

So we had all these conversations about, you know, how they were going to solve this problem. Some of them -- one guy had a breakdown. He literally said, I just can't believe this. I learned this stuff, and I can't believe I've been talking to my patients the way I have been for, you know, a couple decades. And when he learned this, he just kind of fell apart. He was actually talking to Sheri. I mean, you could explain what happened to him. He had some sort of a.....

MS. STRITE: Crisis.

MR. STUART: .....crisis. And the.....

MS. STRITE: He's gone to every one of our programs since.

MR. STUART: Yeah (affirmative), but he was exposed to

things like this. You see this 1982 Volvo award here for clinical science. Well when we looked at this study, it was one of the worst studies we saw. It was just so poorly done that there wasn't one neurosurgeon in the group that wanted to defend this study, but yet, it got the Volvo award.

So the only point I wanted to make is that, when you see things in print, just because they end up in print doesn't mean they're reliable, and it just happens with all -- with medical studies in all journals, too.

This is one that -- we will start talking a little bit about costs here and then we'll talk a little bit about it more as we get into the afternoon, too.

MS. STRITE: Well, first, I want to talk about Group Health Cooperative a little bit. Mike and I collectively worked there, like, 55 years, which sort of a scary number, but Group Health Cooperative was started in 1947, and it was started by consumers who were unhappy with their healthcare and healthcare choices. So it really was a consumer cooperative, and people actually mortgaged their homes to rent a building. Their first doc was a pediatrician, and they said, you know, we're going to do this monthly fee, this capitation thing.

And so I was actually very proud because, when I first started there in '77, I actually got to meet some of the founding mothers and fathers, and they were very involved in

the governance. They did not want to be guinea pigs for research. So there was a very strong consumer perspective and component that Mike and I had come from in terms of our tradition.

And when Mike started doing the evidence-based medicine work at Group Health, one of the things that, I think, was really important was that we were not involved in doing that kind of work for cost-saving purposes. Now, HMOs that came around in the ‘80s and later, a lot of them would not be what we would call Health Maintenance Organizations, which we called ourselves, but really cost maintenance or cost (indiscernible - background noise) organizations. And so our focus was always on patient quality and then we had kind of an adage that we used to say that we had no evidence for, but we had hope around and that was, if we focus on patient quality, our hope is that cost will follow.

And we actually had a number of wonderful successes. Mike might have to leap in here for some of the details, but one of the first experiences I had with one of the clinical guidelines that Mike had promulgated and implemented at Group Health had to do with acute dysuria for women who are at low risk of.....

MR. STUART: That’s bladder infection, for those that don’t know what dysuria is.

MS. STRITE: Yeah (affirmative). And so there was no

evidence that women really benefitted by having to come into the doctor's office and give a UA sample. And then some of the medication that were used were going to give you some new side effects, so you would get a new problem.

MR. STUART: Well, and they were used for many, many days, too.

MS. STRITE: Many, many days. And so that guideline ended up such that -- well, personal confession here. I happened to have a UTI, I think, the week before I started working for Mike, and I was very unhappy because I tend to have very painful UTIs. And so I was getting ready to go into the medical office and just thinking, you know, I know what I have and now I've got to go, you know, suffer for hours. I called up, and they asked me a few questions, determined I was low risk, and they said, oh, your prescription is waiting for you. I'm like what? This is fantastic. And by the way, it's only, like, three days and not ten. So I'm going to have a happy experience. And then I went to work for Mike, and I found out why this happened. And so for that guideline, we ended up creating a lot of happiness. We actually did surveys, and there were many happy women, like me.

MR. STUART: And we saved money.

MS. STRITE: And we saved -- let's see. I should put this in perspective. We were responsible for about half-a-million lives, and we saved about half-a-million dollars a

year on that one guideline. So we had high satisfaction. We had improved quality as far as I'm concerned, and we saved money.

Mike and I go to the state of Washington every year, and we train people up in doing critical appraisal work. And one of our friends and colleagues, Siri Childs, who is a Pharm D, who is recently retired from there and whom we like very much, she created some data that she presented to AHRQ, the Agency for Healthcare Research and Quality, on Washington's prescription drug program.

And one of the things that I want to say that I feel, personally, is that, if you've got a situation of waste, that is a patient harm. So that's part of the benefit-harm equation, and I know, at least, one person here is going to agree with me that that's true.

So we just wanted to show you, by focusing on quality, some of the outcomes that Washington State had in terms of some of their financial outcomes.

Now, here were the components of what they did. They did engage the legislature to pass an evidence-based Preferred Drug List and P&T Committee. So that was a mandate. And then they worked with the Oregon Health and Science University at its practice center, and then again, the DERP program to get research that would help support their own research reviews and then we did continuing education with them on an annual

basis. In fact, we're going there in a couple of months.

And what Siri actually told me is that this is actually under-reported, but in terms of cost avoidance, it's estimated in the year 2008 of $46 million. I'm not going to go kind of through this whole slide, but I think that that's fairly compelling. So we think that doing an evidence-based approach is right for patients and will help eliminate some things, potentially, that maybe we shouldn't be doing.

Again, I want to stress though that there are all kinds of decisions that people may make. They don't necessarily need to be evidence-based, but again, our feeling is that you look at the evidence first. You look at that cleanly. Then you take a look at the other considerations. And then you become very transparent, and you label why it is that you're making the decision that you're making so that you don't mix or confuse a non-evidence-based decision with something that sounds like it, in fact, is. Any comments or questions so far? Yes?

COMMISSIONER PUCKETT: You say a cost avoidance was $46 million, but what percentage of that (indiscernible - voice lowered)?

MS. STRITE: I don't know. All I have is this slide from Siri, but I'm sure that she'd be happy -- if you passed that question along to me, I'm sure I could get an answer for you for that.

So let us turn to page -- so not slide 45, but page 45 in the earlier part of your book, and it's sideways. So this is just a fragment taken out of a bunch of mini-cases that we have, so you're not missing any pages. It's just that there is one case that I wanted to have us review and that is the one that starts number 53, and it goes over onto the next page.

So what I want folks to do is to just read that abstract, and then based on the conversation that we've been having so far, is there something of note there that you want to comment on?

UNIDENTIFIED FEMALE: What page again, is the abstract?

MS. STRITE: It's page 45.

MR. STUART: Well, here's page 45.

MS. STRITE: Yes. I know.

MR. STUART: Okay.

MS. STRITE: I've got it. We're going to help her get it. So 45, it's in the first part of the book where the pages numbers are kind of over to the far right with the little underscore. Yeah (affirmative). You've got it.

MR. STUART: And which question did you want them to look at?

MS. STRITE: Fifty-three.

MR. STUART: Fifty-three.

MS. STRITE: Fifty-three on 45.

(Pause - group reviews abstract)

MS. STRITE: So my hunch is that people are actually taking too long with this abstract. Why would I say that? There is a punch line here. See, I can save you so much time by learning critical appraisal.

MR. STUART: I'm sorry; you were saying something?

MS. EDMISTON: (Indiscernible - away from mic)

MS. STRITE: Go up to the.....

MS. EDMISTON: The perspective of the third-party payer.

MS. STRITE: Well, is that necessarily a bad thing?

MS. EDMISTON: It's a bias.

MS. STRITE: Well, it's a perspective, isn't it? So you know, it's a good thing that they identify that. When you're doing cost-effectiveness studies, one of the things you want to know is, what is the perspective they used? And then that gives you an opportunity to see whether that fits your circumstances, for example. So the fact that they reported it that way is good. It may not apply to you, so it depends.

MR. KNAPP: Is it the data sources?

MR. STUART: Yes. He's the first one in the room who got it. You are the first. Good work.

MS. STRITE: Tell us about that. Read to us all what you see. Yep, tell us what you saw there.

MR. KNAPP: Well, it's that their data sources were the systematic review of MEDLINE and published abstracts. Well, I

don’t even what MEDLINE is, but you’ve just been telling us all morning that abstracts don’t tell you enough to know what the study even was.....

MR. STUART: Right.

MR. KNAPP: .....and so all those abstracts could have -- so whatever you’re assume -- whatever these authors -- who probably think like economists, really -- they didn’t necessarily do a check on what the effectiveness actually was.

MS. STRITE: Oh, perfect. You said that perfectly.

MR. STUART: That was well done.

MS. STRITE: And that’s been kind of a big thing that I’ve been trying to help people understand is that cost-effectiveness implies what? He said it. That it’s effective.

MR. STUART: Effectiveness.

MS. STRITE: Yeah (affirmative). And you can only tell if something is effective if you have done a critical appraisal of review of the science. And so what we see happen -- and I’m so sorry for your colleagues. I’m like, you never do this and don’t report me on this, but I have yet to see a cost-effectiveness study where I’ve seen them say, “And we did a critical appraisal, and we found these studies to be valid and now we can start doing our modeling,” which has new problems as we all know, potentially.

So when you see cost-effectiveness studies, you know, just buyer beware. We’re going to depress you about some

other secondary sources, too, as we get further into this, but I'm so happy that the economist did that pick up. Yay! Our world is now going to change for the better. Yeah (affirmative).

Now, Critical Appraisal Matters. Yes; it does. So I, when I lived in San Diego, got to learn a teeny-tiny bit of Spanish, and I said, so now, my made up drug is going to be Nuevo-Magico, and it can do anything I want. So -- and we're back on this slide. So this is slide 85 in the second part of your book. Can I help you with anything? Mike's hunting for something.

So we are going to look at this case study and discuss it. We've got 240 patients presenting with a number of symptoms of mystery Condition X treated with my new drug, Nuevo-Magico. People may be ill for weeks. The disease is highly contagious and can lead to significant complications, and I would add, especially in the elderly. Side effects of Nuevo-Magico are documented in numerous well-done studies to be very rare. Of the 240 patients treated, 232 patients are asymptomatic within three to five days of coming into the doctor's office.

Now, I'm not going to put anybody on the spot here. So I'm going to have you put all your colleagues who are physicians and not in this room on the spot, and I'm going to ask you, of doctors that have answered this little case, what

percentage of them have told me that they're going to prescribe this drug on the basis of this information?

COMMISSIONER HIPPLER: Ninety percent.

MS. STRITE: No, 70-75.

MR. STUART: But that's pretty good.

MS. STRITE: But you're right; it's a big number. Should they? On the basis of this information alone, should they or should they not? And if they shouldn't, why not? Is there something that you might want to know more? Are there some clues here if something is missing?

DR. BROWNSBERGER: You need to know the natural history of the disease.

MR. STUART: Yeah (affirmative).

MS. STRITE: What about the natural history of the disease? And how would you tell that?

DR. BROWNSBERGER: Untreated.

MS. STRITE: You've got a situation here, don't you, where we've looked at the treated patients. We've not compared them to any group. We've not compared them to untreated. And so what if I were to tell you then, of the untreated patients, 232 patients out of 240 were also asymptomatic after coming to doctor's, three to five days?

DR. BROWNSBERGER: What you've described is probably a common cold.

MR. STUART: Correct.

MS. STRITE: Yes.

MR. STUART: And what kind of study was that? Was it an observational study or an experiment?

DR. BROWNSBERGER: Well, I don't know. An awful lot of antibiotics have been used for the common cold, and none of them work.

MR. STUART: You don't see any words like "in this group or that group." You don't see randomization. You don't see two groups, and like we said, there are two major buckets. They're either observations or they're experiments. The best experiment is a randomized control trial, and you're going to run into lots of little observational studies that say things like this, and I'm going to show you some data on why you have to be really worried when you don't have the other group to compare it to.

MS. STRITE: And so I mentioned earlier that Mike and I -- in our early years, we used to do a pre-test, and we stopped doing it because we were getting the same answers all the time, and the charts looked like this. And because I promised the docs that I'm just going to pick on them, we see the second line there, that we have, essentially, over 75% of physicians answering that question that yes, they would prescribe it on the basis of that information alone. And then we do our critical appraisal training, and they're, like, oh, no. Oops.

And so one of the delights of my life was actually teaching second-year pharmacy students at UCSD, and the Drug Information Director there, a real nice guy, a fairly playful personality, kind of a trickster, he came up to me as I was preparing this class for them, and he said, can I invite my wife? She's a pediatrician. And I said, oh, you know, anybody who wants to come. You know, this is great because I want everyone to understand these skills.

Well anyway, so here are the second-year pharmacy students, and most of them are pretty young, and this is, you know -- everything is a fairly new experience for them, but we've got a doctor in the group. And so I decided, at one point, that, if she was willing -- and since her husband was kind of a trickster, I thought, she's got to have a good sense of humor to be living with this guy -- we're going to see if we can do a little play, "How is Medicine Practiced."

So I went up to her, and I said, do you want to play along with me? And she said, sure; I'm game. I'll do it. So I looked at the students, and I said, do you want to see how medicine is really practiced? And all of a sudden, all of the students are just, you know, head forward. They're so interested.

And I said to her, okay, so here's what happens. I'm your patient, and I've got some symptoms. And someone in the industry has given you some samples, and you look at the

samples and say, oh, well, it looks like this indication and so you give me the pills. And I call you up three days later, and I say, hey, doc, I'm better. I said, what do you say? And she says, it works. And I said, great. Okay. You know, I'm another patient. I come in. I've got similar symptoms to Sheri, and you're remembering what? She said, that worked. And I said, right. And so what happens? You get those pills. I'm like, right. I call you up; I'm better. What do you say? It works. I said, great. Now, I'm a third patient. So we go through this whole routine again, and I say, now what do you say? It really works. And I said, not only that, you have a publishable paper.

So that's the definition of a case series and that's what we just saw there. And at this point, she is laughing so hard that she's actually wiping her eyes because she's crying, and the students are just like enrapt. And I said to her, I said, now, let's get's serious here for a minute. Tell me the truth; is that how medicine is practiced? And got very somber looking, and she nodded her head, and she said, yes; that's exactly right. And that's why we see this, but you'll never do that, right?

Here is another pre-test question that we have. It's probably hard to read up there, so I'll read it here. "A well done study reports a statistically significant relative risk reduction of 60% for patients in the intervention group. Is

this a result that may persuade you to use this intervention with your patients?"

Now, we already talked about this, but I just want to show you, here again, are the results. The light blue is showing us the failure rate to, basically, say, what's the correct answer for this one? We told you our favorite answer earlier. It depends. That's right. It depends because what we don't know -- even if this is a valid study, we don't know if that's 60% of what and so is that going to be meaningful clinical benefit? What's the benefit-harm ratio and all of that?

So we have a huge problem in people not understanding some basic concepts, and what I'm going to show you later on, probably after lunch, is that the math on doing this, to answer that question for yourself, is actually very, very simple.

MR. STUART: The headline for this slide is, "Informed Consent is Not Possible Without Critically Appraised Information." If you think about that, what that means is informed consent doesn't mean that there is good evidence that something works. It just means that you've done the evidence and laid it in front of the patient, so the patient gets a chance to see the reliability of the evidence along with the potential amount of benefit, if it's a reliable study.

I'll say it another way. The reason that we do critical

appraisal is to determine the quality of a study. Now let me ask you this, why is it so important to have high quality studies? I know the answer to many of you is obvious, but we found that it isn't obvious to everybody sometimes.

So let's go through this. Why does critical appraisal to determine the quality of a study -- why, for treatment, is that so important? Because we're saying it's really important, and yet, we're also saying it's really not being done as much as it should be, but it's important, right? I think we've probably convinced you that it's important.

So let's just summarize. Why is it so important to do this? I started the day with what I thought the answer was, so let's see what you think. What's that?

SENATOR COGHILL: Credible benefit.

MR. STUART: Yeah (affirmative). We want to know -- light on, name, even though you've done it before.

SENATOR COGHILL: Senator Coghill. Credible benefit is what I was thinking.

MR. STUART: Yeah (affirmative). You want to have some trust in the benefits, if they're reported. In other words, you can't predict what will happen to you as a patient, if you don't have reliable information about the amount of benefit or the amount of harm. That's exactly right. So it's a credibility thing. And if I just say this is the way we do it, this is the way we treat your cancer, and this is the way

we do it here, that's totally inadequate, and a patient is put in the position of not being capable of informed consent. This is an atrocious situation because people die because they go along -- just, they jump into the river and go down it, and the doctors haven't really done their homework or they can't do their homework because they don't know how or their group doesn't have the resources to find someone to do it for them or the FDA has done a bad job or it's published in a journal that was peer reviewed, but the peer reviewers didn't pick up all the problems or -- and it just goes on and on and on and on.

DR. BROWNSBERGER: Or Congress hasn't authorized money to NIH to really study it and so we leave the studies to the drug companies instead of doing really first-class studies, which the NIH could do, so.....

MR. STUART: Or the NIH doesn't do a good study. Or the University doesn't do a good study. So it's -- the "or's" just go on and on and on, but the bottom line is all these things need to be addressed at various levels, and the starting place -- what's the starting place? It's those four areas of bias, the selection, performance, attrition, assessment, and there is a fifth one, other biases and problems with the study. Yes?

MR. RILEY: John Riley. Are you going to address the confounding effect of placebos and the reason that we do

placebo control?

MR. STUART: No. We didn't, but I can, very briefly, say there has been millions of words written about placebos. There is a lot understood about placebos. There is the contextual sort of halo effect around the placebo. There is the placebo, itself, that -- well, let me just give you a case to show you how this, the placebo, might work.

If you take people with Irritable Bowel Syndrome and you recruit them into a study, one-third of the group gets what -- I would say, the whole enchilada. They get a new drug. They get visits with their doctor, and the doctor has been trained in communication, understands Irritable Bowel Disease very well. The next group just gets to pick up the drug, and they either get the active drug or the placebo, but no human contact. And let's see. The third group gets the -- one group gets the drug and the contact. The other ones get either the placebo or the agent without the human contact, without lots of support and conversation and visits. Well, it's been shown that, if you get the contact and the right drug, you're going to do the best. If you get the contact, at least in this one study -- if you don't get the contact, but you get the placebo, you do better than those who don't get any contact and just pick up the placebo.

The placebo is a control that we use for comparing to the active agent, but in fact, there is a placebo response. Then

you have to use judgment to ratchet, depending upon what the study is, how much difference there really is in the real world. So I think the answer with placebos is we could spend a whole day on them. I don't think we want to do that with this group, but.....

MS. STRITE: No, but there is a very important point I think we should make.

MR. STUART: Sure.

MS. STRITE: And that is you want no difference between the groups except the thing that you're studying. So if you're studying a pill, you're interest is in the active agent. Placebo control becomes very important because taking a pill now becomes equal between the groups, and with research, what we're looking at is, what is the difference in outcomes between the groups? So we use the placebo when we have a pill -- I mean, this is the ideal in research -- because the placebo effect is going to come out in the wash.....

MR. STUART: Right.

MS. STRITE: .....because the difference in the groups in the valid is going to be because of the active agent.

MR. STUART: In other words, the experience that I was talking about, the human contact element, is going to be the same in both groups in a study and that sort of cancels that problem out. Yeah (affirmative)?

DR. BROWNSBERGER: One of the problems -- I'm Keith Brownsberger. Sorry I'm talking so much, but.....

MR. STUART: No. It's good.

DR. BROWNSBERGER: One of problems with the thing you selected, Irritable Bowel Syndrome, it's probably not one disease. It's probably five or six different diseases. So your selection of patients to go into each group is going to be a real problem. And that's -- it's more of a symptomatic than a real, definite disease. So you really have to be careful that you're treating the same disease with the placebo and the medication.

MR. STUART: Yes, and they can try to get around that by constipation predominant or diarrhea or pain predominant, but you're right. I accept that, and I think that's true. Irritable Bowel Syndrome is probably not a very good example to be using for much of anything because of that, actually, but that's the first one I thought of when this question came up. So let's continue on here. Or should we quit?

MS. STRITE: Well, actually, we can take another question or two before.....

MR. STUART: Yeah (affirmative). Let's take some questions, then break.

MS. STRITE: Yeah (affirmative).

MS. EDMISTON: This is Gena Edmiston. But how do you assess validity of a study where the recipient of a treatment,

it can’t really be blinded so that the treatment is obvious? Someone is getting a massage or manipulated.

MS. STRITE: (Indiscernible - simultaneous speaking) for that study; yeah (affirmative).

MR. STUART: For surgery?

MS. EDMISTON: Yeah (affirmative).

MS. STRITE: Well, the -- I mean, here’s the bottom line. Science is not going to be able to answer all questions for us. Some topics are just not going to lend themselves to it, and if you have a study that’s not blinded, you have bias in there. And Mike and I, sometimes, will have scientists show up and -- or actually, people that are sympathetic to how difficult it must be, and they’ll say, oh, you know, but the investigator tried really hard. So you try to make the study as clean as you possibly can, which you should anyway. Be aware that there is going to be a bias because of blinding. And then it may be, at the end of the day, you have to say, and because of that, I can’t trust the results. So you’ve got to think about what the impacts are in terms of it not being blinded.

Let me give you an example harkening back to the Boise neurosurgeons and orthopods where -- oh, where was I going with this? Help me.

MR. STUART: Well, her question really has to do with ethics, in a way, because she said you cannot do it, but in

fact, many times, you can. For example, sham surgery can be done. Now, you have equipoise. You don't know what the right thing to do is. The ethical group, on one end, is going to say we cannot ethically do a sham operation on people and make an incision into their skin because it's unethical.

The other extreme -- and I tend to be more there -- other people tend to be there, but I tend to be more here. This group says we have equipoise. We're never going to know, and if we can get people signed up and they agree that yes, I will be getting a sham incision maybe, but I might be getting the real operation. At the end, I've contributed, and I've made a difference. And that ethical perspective says we'll never know then because, in certain cases, we're not going to do it. We're not going to randomize people to smoke or not smoke, but we can randomize people to get sham surgery for knee debridement and real debridement for lavage of the knee, for osteoarthritis so they don't have to have a total knee replacement, for example. That study was done with a sham and assured that that operation did not work. So without that, orthopods would be doing this a lot more than they're doing it.

So in a way, I'm answering it that, really, it depends. It depends upon your ethical perspective about how you think we should behave as an ethical country and how much it matters when you don't know something. We're going to have to treat

some patients with orphan diseases without really knowing whether we're benefitting them. As long as the patient knows this and we know that and the drug has been FDA approved, then it becomes patient-centered care, and it's up to the patient.

MS. STRITE: So that's one -- but that's one issue.

MR. STUART: That's my answer.

MS. STRITE: Yeah (affirmative), but part.

MR. STUART: Yeah, yeah (affirmative).

MS. STRITE: Because what I was going to raise was the SPORT trial, and the SPORT trial was, again, trying to deal with low back surgery -- sorry -- back surgery for sciatica with radicular symptoms. Did I do that right?

MR. STUART: Uh-huh (affirmative).

MS. STRITE: And what happened -- and this trial was funded by the NIH and was probably, possibly the best study we're going to have on this, and what happened is patients were randomized to low back surgery or to watchful waiting. Well, what happened was a number of patients ended up having their back pain resolved who were scheduled for surgery. Well, at that point, many patients are pulling out of the study, saying, well, I'm not going to undergo surgery now when I don't have any symptoms.

Likewise in the other group, there was about 25% of the population that was in so much pain that they said forget watchful waiting; I don't even care if it works or not. Just

do something. And so they underwent surgery.

So now, what kind of study design do you have when 25% of your randomized patients have chosen the opposite of what they're assigned to? What kind of study is that?

MR. STUART: They have choice. They have choice.

MS. STRITE: It's an observational. We've got observational stuff mixed in with the RCT. So we've got a big problem there. So the fact of the matter is that there are going to be certain topics that are not going to be answerable this way and then that doesn't mean people don't practice medicine. It means, again, that you inform the patient that you don't have good evidence on this. You know, maybe it will work, whatever choices you make for that.

So the one thing though that I see people do, especially that are new to critical appraisal, is trying to grant extra validity points, you know, because the investigator was in a tough spot, and you can't do that either. You're not grading the investigator's efforts. You're grading whether or not you feel that you can trust those results or not.

MR. STUART: Another question?

COMMISSIONER URATA: Bob Urata. So what percent of all medical knowledge that exists is good or meets your criteria? Because the production of medical knowledge now in the journals that we all read, and reputable journals, and all our guidelines and standard of care, based on evidence-based

medicine, what percent of that is valid and usable in your estimation, or if you actually know, what would you say?

MR. STUART: Well, we know some about it. As.....

COMMISSIONER URATA: Can you give me a percentage?

MR. STUART: Yeah (affirmative). I can. The Institute of Medicine and others, (indiscernible - voice lowered) and others have really shown that, if you look at the studies that are published or the guidelines -- let's start with guidelines. If you look at a -- this has been done with many guidelines, many times. Fifty percent of them are going to have major problems. So we know that 50% of the recommendations and guidelines are problematic.

Let's move to a primary study, a randomized, controlled trial. The McMaster Group regents who looked at their own evaluations for their formulary over three years, was it, something like that?

MS. STRITE: I think it was a year-and-a-half.

MR. STUART: Yeah (affirmative). And others and ourselves estimate that it's less than ten percent, and we don't talk about medical knowledge so much because that, to us, is a different -- maybe slightly different. We're talking about the quality of studies. So I just want to emphasize that.

I'm answering in terms of quality of the studies rather than -- for example, I wouldn't want to say that, in a

situation where a patient is making a decision, there is no good evidence, and they decide to do it anyway that that's a problem because I don't think it's a problem. I think it's a problem if the patient doesn't know that there is no good evidence. That's all. They could make a decision based on beliefs on their own preferences that have nothing to do with the evidence because they're under so much pressure to have something down that they're going to do it anyway. And I'm just saying that, from the perspective of the studies, it's less than ten percent.

COMMISSIONER URATA: So my efforts to keep up-to-date and do the best for my patient, there is no evidence for that.

MR. STUART: No. I never.....

MS. STRITE: (Indiscernible - simultaneous speaking)

COMMISSIONER URATA: No evidence for me to base.....

MR. STUART: No. I wouldn't say that.

MS. STRITE: But it's like you're drinking from a fire hose and so what we hope is that, with these skills and some of our more reliable sources maybe, that that may help you sort of get away from the fire hose and maybe have some opportunities for evidence that's maybe going to be more likely to be valid and helpful.

MR. STUART: Yes.

MR. STRITE: We have to do a lot of grief counseling, unfortunately.

MS. TSIGONIS: Jean Tsigonis. I guess I've relied on AAFP essentials and kind of their review things, thinking that they've gleaned everything in giving us, you know, the best at that time, realizing things do change decade-by-decade. But did you find any difference in that kind of stuff when you were -- I noticed you quoted from AAFP there that wasn't really good.

MR. STUART: Right.

MS. TSIGONIS: So and they give us credit for.....

MR. STUART: All specialists.

MS. TSIGONIS: .....evidence-based medicine, you know, based on these articles. You know, if they have that little star, we get a little bit more credit. So have you found them to be any better at all?

MS. STRITE: The remedy, I think, is, first of all, to get familiar with the concepts yourself so that you can help yourself answer that question, number one. You're going to get variability in there. In our experience, primary care docs tend to be more embracing of an evidence-based approach, and I think that there are a number of reasons why that's the case. And so I think that primary care, overall, tends to do a bit of a better job in this area, but if I were a practicing physician, I'd be looking closely at that information.

So probably what I would do is, if there was something that I thought was going to change my practice -- and I'm

talking about as an individual practicing doc, not if I were part of a team of something -- I would look at those articles that I thought -- that they starred and might I change my practice on that and then go look at the original article, apply the critical appraisal checklist and then see what I thought.

MR. STUART: I would just add to that that, for me, the ideal would be journals would be -- they'd kind of take a different approach. Instead of the abstract, they could continue to have the abstract the way it is, but then they should have a checklist with the critical appraisal done for you, and it should be graded, and you should be able to see how they critically appraise it. There is no reason they couldn't do that, but they wouldn't be accepting as many papers because they're not -- they're in a double blind. If they publish bad studies that they've rated as low, then they're going to get whacked, and if they don't, they don't have a job. So it's -- you see, it's -- actually when you get into this, there are all kinds of problems with information. We should be able to get, without paying, almost any study that has ever been done, but we can't. Lay people have to pay. You and I can get anything -- well, mostly anything we want, but other people can't. So they don't even have a chance to do a look at the literature without help from somebody from inside. Yeah (affirmative)?

MS. STRITE: I do have to give you though a little cautionary end to the story to our Salpeter review, and this has been a number of years, and maybe this would be different, but they rejected our letter, and they rejected it on the basis of the fact that our letter was not speaking to the original research -- no; let's see. Let me say this differently.

The original research was published like, what, I think a year before.....

MR. STUART: In the *Journal of* (indiscernible - simultaneous speaking).....

MS. STRITE: .....the commentary came out, and they said, oh, no; you should be sending your letter to that journal, not addressing the commentary.

MR. STUART: And they know fully well that the time limits had expired for that journal.

MS. STRITE: Right. And so people are allowed to read that commentary without any kind of rebuttal of that, and I thought -- I have a lot of respect for AAFP, but I thought that was very unfortunate and doing a service, so.....

MR. STUART: Ward?

CHAIR HURLBURT: Part of the response to Bob's comment was, you know, I remember what you taught me more than ten years ago now that, in trying to be disciplined and conscientious and keeping up, when you look at the abstract --

like if I look through my monthly issue of the *Journal of the American College of Surgeons* and I look at the abstracts, I don't learn anything. It tells me what article I want to look at.

MR. STUART: Right.

CHAIR WARD: And I think then, with the skills -- and I think you're trying to make it simple, so that people like me can use them, that you can look at the article and learn something, but I would say, before Mike taught me that, I would go through the abstract and say I learned quite a bit about what's in the journal, and I realize that that's really not the case.

MS. STRITE: So one more before we break for lunch and then we will resume with a Q&A.

MR. STUART: Well, Ward might have some closing remarks, but let's have the one more question. Then Ward, you can close out.

MR. CAPISTRANT: So Todd Capistrant. This has got my brain thinking about, when we're driving towards this quality measure for physicians, that you're going to be marked and graded on the quality of the product you get for giving good healthcare, but it seems, to me, that, if only ten percent, or pick the number you want to quote, is of value for us, as a practicing physician, we're tremendously behind the curve on trying to provide quality. So rather than target the

physician and have all these hoops to jump through and make sure you're.....

MR. STUART: Right.

MR. CAPISTRANT: Why aren't we targeting the huge dollars that go to research at universities, NIH? I mean, it seems like we're trying to dam the stream at the end instead of at the headwaters.

MR. STUART: I agree. I agree with you.

MS. STRITE: We want to get everywhere. We want patients engaged. We want the researchers engaged. We want the journals engaged. We want the payers engaged. And then again, I think, what I said earlier, a big, big, big problem is that it's not gotten into the educational system in the first place, which, if that were to happen, that would be correcting a lot of these other areas.

MR. CAPISTRANT: Right.

MR. STUART: Ward, do you want to.....

CHAIR HURLBURT: Yeah (affirmative). Just very quickly, and we'll break for lunch. You've heard me mention David Eddy. You've heard Sheri and Mike mention him, and he really was a guru and continues to be to this day.

When he was Vice President at Kaiser with that huge system and when Group Health was kind of affiliated with Kaiser, not a part it, David Eddy would come up to our Technology Evaluation Committee meetings, and he would say,

almost every time he came, you guys do it better than we do in Kaiser.

So this is really, I think, good stuff for quality of care in the world, not just in the U.S. This is a worldwide issue because it has been so challenging there.

The second thing I did want to mention, and you saw Sheri read it, Mike and Sheri have published a little book that they don't have copies here to sell, but it is kind of a manual, I think, that also will be something that can be used. I just did want to mention that, for docs particularly probably who are available, that is something that has been published.

So we'll break for lunch, and we're due back at.....

MS. STRITE: According to my watch, it's ten after noon, so 30 minutes from lunch. So we will reconvene at 12:40.

COMMISSIONER ERICKSON: And lunch is boxed lunches right outside the door in the hallway here. Everybody should help themselves.

12:08:53

(Off record)

(On record)

12:42:28

MS. STRITE: So if everyone can get cozy and take your seat? So for anybody here who is eligible for CME, there is a general evaluation in your binder that anyone can complete, but those of you who are eligible for CME who want it -- and

why wouldn’t you -- there is a special evaluation form that is manila-colored, size of this yellowish paper. This is the CME evaluation form that you need to complete, and when you complete it, you put it out on the table outside, and you will get -- ta-da -- a CME certificate. So that’s the process for that. Deb, anything else you wanted me to mention for that? And this award goes to Deb. You’re welcome.

UNIDENTIFIED MALE: Do we have to wait until the end of the day?

MS. STRITE: I’m not involved in that part of the process, so I don’t know.

MR. STUART: Trick question.

MS. STRITE: I don’t know that, but you are psychic, I believe, right?

MR. STUART: So I’m going to re-comb a couple of the things we said earlier and then introduce some new things. We do re-comb things so that people who don’t get them the first time may get them the second time. And so I apologize to all those one-passers, but we’re going to -- we’ll do the best we can to keep introducing new stuff, but also making sure you get the key foundational stuff.

So Research Bias again. We’re going to do just sampler because of our time, and it’s not what we would do if we had two days. We would give you the full picture, but we’re going to just sample so you get the idea. But still, at the end of

the day, we would like to have a discussion about what you, as leaders and policymakers, commissioners, et cetera, we think you should have in your backpack in terms of helping others do their jobs while getting people in the quality arena to continue to help you in ways that will help your decisions also.

So we'll be talking some about P&T Committees, about Technology Assessment Committees, and about any group that's doing quality improvement where evidence may have a role and that really has to do with anything in healthcare, to tell you the truth, or any role that people might have.

So Sheri already asked this, does it really matter? I asked it, does it really matter? And the answer is yes. We want to know whether the intervention caused the outcome or whether the outcome was biased or distorted through bias or whether it happened by chance. And then once we've done that, we can then, if we say that the risk of bias is low enough to accept this, then we would look at how much benefit there is. So that's just a repeat.

Remember: validity means closeness to truth. Truth changes depending upon our scientific knowledge, so it's always a moving target. So we don't say truth; we say closeness to truth because it's really at this time, and it might change dramatically. I never thought we would suspect bacteria for causing ulcers. I mean, that was the farthest

thing from my mind, and I think, didn't he have a problem because everybody attacked him, the guy that -- the H.pylori investigator had emotional problems because everybody thought he was nuts. It's just that information changed.

Does everybody remember those five areas? Mortality, morbidity, function, quality of life, symptom relief. Very, very good. And Sheri put a shrimp in there for some reason. Why did you put that shrimp in there?

MS. STRITE: Because jumbo shrimp is an oxymoronic -- so the size of the benefits matters; 60% relative risk reduction may, in fact, be very shrimpy.

MR. STUART: A little inside joke. Thank you.

MS. STRITE: A little inside joke.

MR. STUART: Oh, there it is. So where we started was the first stage of a study, which is getting people into the groups, and you want the groups to be the same. So here, we have the control group and the out of control group. I mean, the control group and the comparison group, and they should be the same in all prognostic variables, clinical, co-morbidities, et cetera.

Remember, I said there are two main study types. We start with, what's the appropriate study type, and then how well were the methods done, and if it's valid, then we look at the results. And the results are the clinically meaningful. So reliable and clinical usefulness. So observations are

where people choose things.

Experiments are when we take choice away, and the investigator rules the world. The investigators create the protocols. They direct traffic. People don't choose which drugs they have. They don't choose much of anything, if it's done perfectly, which it's never done perfectly, of course. In fact, we've never seen an "A" study. Wait. We did see one "A" study, but most of our studies, our good studies end up with a "B" because they usually have some problem. They're not just perfect, but a "B" study is a really good study.

The important point is that, with experiments, you have this idea that people agree to participate in something that they're not controlling. So it's different than those who just go out and get their usual care and then an investigator grabs all the charts and writes up a study on what he has observed that they chose to do.

Does everybody see the difference between observations and experiments? Because that's a very simplified way of getting around all of the different types of studies that you will run into. So when you look at case control study, then you basically say, well, it's not a randomized, controlled trial. It's an observational study. It's not an experiment. And we want experiments.

Remember, we're creating equal groups with randomization, and Sheri gave you the example of the vitamins and decreased

heart disease, but it wasn't the vitamins. It was that the people who took care of themselves, as part of taking of themselves, also took vitamins.

Here's a -- you're now going to do a critical appraisal. Sheri's going to help you through it.

MS. STRITE: Well, actually, I'm going to save time, and I'm going to do it for you because it used to be the case that, occasionally, we would have somebody come and want to argue with us about observations can be good as RCTs. So I had to invent this little case story to explain no; when we're talking about therapies, they are not. So I'm just going to review this really quickly. So I'll get myself oriented here. Which side is which?

So here, we've got an observational study, and in this example of our observational study, we've got two concurrent groups for study. Sometimes, people think that that's what creates an experiment that you have two groups. Not necessarily, because you can do two case series next to each other. You can find groups to compare. So that's not the key. The key is whether or not they chose their treatments or not or if their physicians chose their treatment.

But in observational studies, you cannot help but have varying baseline profiles. Therefore, you've got two groups that are different no matter what, and frequently, on the basis of the thing you're interested in studying. A lot of

people say, well, you know, if we do propensity scoring, you know, if we do adjustment for those groups. Well, how many characteristics do you think you're made of? Probably billions, right? I mean, it has to do with where you grew up, what you eat, you know, genetics, I mean, a lot, a lot of things, and to adjust the baseline characteristics properly, you'd have to have access to adjusting everything and that's not possible. I mean, in the first place, you can only report on what you can measure. Much of what we're made of or have experienced is not measurable. And so they can only adjust things that they have measured and then you're trying to do a lot of guess work. So that does not solve the problem.

Again in observational studies, we have patients or their physicians making choice, and we've showed you how that can end up confounding things. There is no blinding. Everybody knows what treatment they got. And we'll talk a little bit later about blinding and how important that is, but if you have an unblinded study, right off the bat, you have some fairly significant problems. There are no formal treatment protocols; therefore, people are treated differently. There is varying medication use, measurement methods, duration. So you may have many co-interventions in one group as compared to another and that may explain your outcomes.

So observational studies. We're not trying to bash observational studies. They have many good uses and purposes,

but trying to prove cause and effect of a therapy is not one of them, unless you have what Mike referred to earlier “all or none results.” So he gave you an example of that and that is very rare. Whereas in a well done randomized control trial, you have your comparison groups, and they’re taken out of the same pool in time. And the baselines, if randomization has gone well, should be very, very similar, and you should not adjust that is our feeling. Even if you think there is something to adjust, we think it’s fine to look at it, but to report the outcomes unadjusted also because, by adjusting something, you may have caused some other imbalance in some way. If you have a computer-generated random number table, that’s great. That would be an example of a method that would actually determine randomly the sequence for assignment. Ideally, no one knows who got what treatment. That includes blinded assessors. We actually like to see that called out in the report. There should be a formal treatment protocol, controlled medication use, uniform measurement methods, uniform duration, et cetera.

So what we’ve posed to people is, if you were going to do a critical appraisal -- and if I hadn’t first told you this was an observational study -- what grade would you this get of our system, A, B, BU, U? That would get a U right off the bat. Well, you don’t get extra points because you’re an observational study. This is as good as an observational

study can get. This is the supreme. So again, useful for a lot of things, like safety and hypothesis-generating, helping us understand things like adherence, that sort of thing, but not cause and effect.

This particular slide is done by a very famous physician researcher who is the head of something big at Stanford, I believe, and he did a very complicated formula, trying to get at the positive predictive value of various study types. And basically, the bottom line is, he says, if you've got a well-one RCT or a meta-analysis of well done RCTs, that you have about an 85% likelihood that you've got valid results, but then when you look at the, quote, well-done observational study, we drop down to 20%.

Now, Mike and I are not capable of evaluating his methodology, but there is a consistency in the order of these studies and their predictability that is compelling, and even if we cut these numbers by half, to me, that's very meaningful.

As Mike pointed out, we tend to think of there being four stages of a clinical trial. So there is how the subjects are selected, who they are, how you got them into the study, how you got them into their study group, those kinds of questions. And then there is what you are interested in studying, along with other things that happened to the patients along the way in the context of care.

And we do our sort of line up at this a little bit differently. Some groups talk about attrition bias at this point, but I think attrition bias is just one piece of a whole data collection issue. So there are a lot of data collection facets that you have to look at, and then what ends up being missing in terms of what data you didn't capture and -- or people ended up dropping out of the study, that whole area, and then there is the assessment arena. So assessment bias comes into there. And again at the end of the day, we're comparing the groups to look at the difference in outcomes between the groups.

Now, one of the things that I want to mention is Mike and I have our own biases, if you will, and Mike tends to like an orderly procedure. So if he were doing an exercise right now, he might invite you, after calling out a bias, to ask you where it fits into these groups. And Mike likes the ordering principle. I must have been in the out of control group that he showed earlier, and I just want to find a bias.

And so again, bias is anything that systematically leads away from the truth. So if an organizing principle is helpful for you, great, but you don't have to worry about it. If you find anything that you think may have distorted those results, that's a bias, and you can stick it in another category if you need to.

The other thing that I want to mention is that we spend a

lot of time talking about anything that you find that's different between the groups, except for what you're studying, that is automatically a bias, and I want to stress that because, when we watch people do critical appraisal work, over the years, I've seen people miss that a lot because they're reading along in the article, and you know, everything is written up nicely, and they may have logical reasons why they did something. Like for example, we have people coming in for screening. We have people who didn't come in for screening. So the date we chose to follow people for the non-screened group was an arbitrary date we picked. Those who were screened, it started at the date of screening. Well, now that sounds logical because they had to make a choice there, but what they did is they created a difference in the follow-up periods. And in that particular example that I'm talking about, the difference ended up giving them an extra two years to find prostate cancer in the unscreened group.

So what was our conclusion? Prostate cancer screening saves lives. Well, no. You had two extra years to find cancer was one of the problems with that.

So we harp on the difference between groups because, again, you get sort of lulled into reading the logic and then miss that, oh, my gosh, that was a difference between the groups and that really mattered. This group, we had an extra visit for some reason.

So Mike gave you this story of my friend at UCSD, who, as a medical resident, was unsupervised at night. What was one of the flaws in what Mike described, besides the fact that they were directing traffic to which study arm a patient got assigned to? There's a little clue in there that there is another big problem with that study. Why did they do that? Why did they -- when they saw an intervention group, why did they make that the next patient eligible for randomization did not get that assignment?

COMMISSIONER HARRELL: The amount of work.

MS. STRITE: The amount of work. Okay. So what if I told you that the study was not looking at the amount of work you did with the patient? Therefore, that means that those patients, they were getting extra work up, maybe extra contact with their care provider and so that could do something in the course of maybe discovering something about the patients that got more time with the doctors or maybe they felt they got more supportive care from the nurses or whatever. It could have made a difference in their outcomes. So anytime you notice anything that's being done to the groups that is not identical, other than the intervention in the study, you have a problem.

The other thing though is that bias could be something that's affecting everything. So you might have a bad measurement instrument, for example.

So we're going to talk about, who do you get and how do you get them into their groups? And I'll just ask you, Mike, did you want to say something about inclusion and exclusion criteria, based on one of the questions that came up earlier?

MR. STUART: No. I think we got that clarified.

MS. STRITE: Okay. So we're not going to focus on that, but it is important to look at the inclusion and exclusion criteria, but that's actually, to me, not as important -- well, I mean, it is important, but especially important is looking at the Table of Baseline Characteristics because not only can that help you get a sense of whether or not the groups were balanced through randomization, but it tells you who actually got studied. So inclusion/exclusion criteria, you know, they may say, well, everyone was eligible between 18 and 65 years of age, but the fact of the matter is all the 40-year olds showed up. So it's kind of like the invitation to the party versus the PICO part who actually got studied.

So what's happening with the whole randomization issue is that a sequence is being generated that's going to determine which group a patient gets assigned to. And so I'm going to use a three-arm study because it's a little easier to illustrate when you've got more than two elements, but let's say we had a sequence generated like this. It's randomly done, A, A, A -- sorry, did I do it three times? Three A's, B, A, C, C -- you want something that's not predictable and

then randomization, ultimately, should channel people equally so that groups are balanced in terms of their prognostic characteristics. If we do not see a description of what they did to randomize, then we mark that as a threat against validity. And it's enough for us just to see the word "randomization" in the title, they've got to tell us something like this, and it can be very simple. This is all we need to see, that they used a computer-generated list of random numbers.

Mike talked a little bit about Concealment of Allocation. Do people -- this is kind of a unique one where a lot of people haven't heard of this before. Can someone explain what this is? Do they recall, what's Concealment of Allocation?

COMMISSIONER HARRELL: You don't know which group they've been assigned to.

MS. STRITE: You don't know which group they've been assigned to; right. So it's like -- and the word "blinding" is really used more downstream, but it's a kind of blinding, and in fact, blinding really starts with randomization because, with randomization, you have something unpredictable. So that's one important point. And then the next thing that helps blinding happen is not directing a patient into a group that you favor them to go into, so that helps with blinding as well as selection bias issues. So you're avoiding channeling and then you're at the start of keeping it unknown what a

patient is taking.

So here are some things that are acceptable methods of Concealment of Allocation. Mike had mentioned, if you have a call-in center, interactive voice response system, if you had sealed containers that were all identical and locked and in the control of a pharmacist who is otherwise not involved in the study, things like that -- so you want to try to get a sense of whether or not they've done a good job trying to hide that. And then I will just tell you that we see Concealment of Allocation reported in probably fewer than 75% of the articles that we look at, and yet, The Cochrane Collaboration has identified this as one of the key indicators of quality for a study. So again, we've got a big dissidence in terms of what gets reported and what we need to know.

Then there are the issues of blinding; can someone unblind it? And in terms of blinding, we tend to see critical appraisers doing a really good job at looking at who is blinded and how they were blinded. So they would look for -- we have a placebo. It's matching, matching in taste, et cetera, but then completely neglecting what was the success of blinding. So we've seen critical appraisals where somebody did, you know, a nice job blinding the medications when it was a highly toxic chemotherapeutic agent versus placebo. And then they'll mark, oh, it was well-blinded. Really? You've got a bunch of people losing their hair in this group, who

otherwise wouldn't lose their hair.

So one of the things that's really important is to try to get a sense of whether or not blinding is likely to be successful or not.

And then Mike had mentioned some other performance bias examples. You know, things like, was it likely to be successful, did the patients actually take the medication?

A couple things that I want to point out about that. Again because you're looking at the difference between groups, sometimes, some things that seem like they might bias outcomes may, in fact, not matter. So if you had patients having some co-interventions -- you know, taking some other medications or other care experiences -- if those are balanced between the groups, then that comes out in the wash when you're comparing the outcomes between groups. Yes?

COMMISSIONER HARRELL: It's Colonel Harrell. Just a quick question for me. Crossover, is that a performance bias or a blinding issue?

MS. STRITE: You're talking about crossover not -- to make it clear, he's not talking about crossover design, which is a different thing, but you're talking about crossover, like migration, like I talked about in terms of the SPORT trial with the surgery?

COMMISSIONER HARRELL: Uh-huh, yeah (affirmative).

MS. STRITE: So ask me your question again. So is

it.....

COMMISSIONER HARRELL: So when you have -- I have been assigned to a group, and I'm not doing well, and I, the patient, say, you know what? Not so much. And that's what you mentioned. I'd like to go to this other group by virtue of my own choice. Does that count under performance bias?

MR. STUART: That's what I would say, but Sheri would say it doesn't matter what you call it. It's a difference.

MS. STRITE: Yeah (affirmative). I would say it doesn't matter what you call it.

MR. STUART: She'd say it's a bias.

MS. STRITE: But it can happen for different reasons that matter though, too. It may be that I -- and I was going to show this later, but this is a perfect time for it. It may be that I know that I'm on placebo and my doctor knows that, and I end up feeling -- you know, I'm not getting better. I know I don't have the active agent. And so I'm going to take myself out of the study or see if I can get on the other agent. Now, we've got a blinding problem that resulted in a crossover problem in attrition. So some of these things are very interconnected as problems and relate to each other. Is that helpful?

MR. STUART: Yes. In my mind, I think like you do. I'm not doing -- in the performance bucket, I'm not doing what I should have been doing according to the protocol of that

study. In other words, I'm having experiences that were not supposed to be experienced in that group. So to me, that's a performance bias. In a sense, we don't really care what you call it, as long as you call it a bias.

MS. STRITE: So Data Collection Biases. Some of this is, you know, obvious in terms of how is data collected, et cetera. If you do, like, a mail questionnaire, you're going to get a different response for satisfaction than if you have a nurse asking a patient walking out of the doctor's office, you know, how did you feel about this experience, and the patient isn't going to want to say anything bad. So some of those things, you know, are important to look at, obviously, that are more obvious, but the thing I want to mention is that people get very excited when they see the words "validated instrument." And "validated instrument," what does that imply?

UNIDENTIFIED MALE: (Indiscernible - away from mic)

MS. STRITE: It implies it's valid, right? Just what we've been talking about all day. And it can actually be very easy in certain circumstances to earn that title. And so here is one example of some validation methods where they've got three questions. Face validity: Does it make sense? I designed it, so yeah (affirmative); it made sense to me. A matter of judgment, right? Content validity: Does it include all the right stuff? Well, we've got the same situation here.

Construct Validity: Is it accurate and dependable? Well, frequently, we don't have something that's already known to be accurate and dependable to compare it to, so big question. So I just caution that, if you see "validated instrument," that raises some questions, and the way that Mike and I approach is, if we see that they've used a standard instrument, we usually stop there, but if they've used something that's unusual, we may take a look at the references and then identify the validation study and then go do a critical appraisal of that. So state your name and turn on a light somewhere.

MR. BLAS: My name is Leo. I'm more involved in the statistical financial studies, but it's the same methodology. You don't mention much about the numbers of the people in the sample, the number of the population. To be valid, you have to have so many people in the study as compared to the size of the population for whatever you're studying to get the validity for (indiscernible - voice lowered), right?

MS. STRITE: It depends. Yes. So I mentioned earlier that a small sample size is more prone to chance and then Mike is going to talk about power and what that's all about, which has to do with frequency of events. Actually since people have just eaten lunch, is the time to announce that my boss in San Diego, who is one of the few big researchers in family medicine, said, you know, we're living here in San Diego and

are full of beans, and I wonder if Beano really works? So we had a clinically meaningful question to all of us who enjoyed Mexican food, and one of his partners in family medicine was very playful also and so they got a statistician roped into this activity, and they said, yeah (affirmative); we're going to study Beano, and we're going to do a power calculation. And they got 19 people participating in it, and it ended up getting published, actually, and we're happy to report that Beano works. So we did a lot of things to make it a valid study, and the effect size must have been fairly large to be reported as such for such a small population. So there you have it.

MR. STUART: In general, 100 above that -- well, let's put it another way. Below 100 is usually considered a small study, and then when you start getting down below 50, a very small study. I showed you that study with 5,000 in each arm. That would be a big study, but 100 is -- the number 100 is where people start to raise the antenna, like this -- we might be recruiting some outliers into our study because it's too small. And so you're right; numbers do matter for a number of reasons, and we'll slowly cover them as we go along.

MS. STRITE: Yes, but Mike is going to show you how you could have a study of 10,000 in each arm and that may not be enough people.

MR. STUART: Still be not enough.

MS. STRITE: Yeah (affirmative). So Attrition Bias. Attrition bias is one of the areas that is really very complex. On the chart that Mike showed you on the evidence-on-the-evidence, that's an area where there is less solid information and kind of more inconsistencies and inconclusiveness over the attrition area. And I think the issue is essentially this; attrition may not result in attrition bias, i.e. something that distorts the results. And so how you try to identify that can be a bit on the tricky side.

MR. STUART: For the lay folks, what we're talking about there is that you randomize 100 people to each group, but then during the study, people drop out from each group and.....

MS. STRITE: Or may not complete a final test. So it really is all about missing data in some way. And so this is a tricky answer, and here is what we see some groups do. Some groups get -- want to apply some kind of standard and so they're critical appraisal approach is to say, oh, well, if there are 20% of people missing in each group, then we're just not going to look at the results of that study and that may be a fine way to do it. I think that it's not because I think that you may have studies that you are excluding that are otherwise valid. And so what we do is we try to look for clues. Differential loss is usually pointing to something about the agent that's a problem. We take a look at whether

or not they've reported a comparison of baseline characteristics and then reported characteristics of completers and are they the same. So there are a number of things that we try to take a look at.

Now, I'm going to show you a very different kind of slide. I am a digital artist, and I thought this was actually kind of a fun piece for Alaska because it's got mountains and bears; oh, my! But the reason why I'm showing you this here is because the next thing I want you to realize is this, too, is digital art. This is not meant to be understood.

I did a project where the issue was, I think it was, like, 50% loss of data for a study. And so I did exactly what I described as taking a look at a lot of contextual elements, trying to get a sense of whether or not attrition bias actually could have distorted the results or did we just have a case where now we have a smaller sample size. And the point of this slide -- and at the time I created it, it really did have logical meaning. These lines are not really made up, but if someone asks me now, stepping away from it, to try to explain it, the best thing I can say is that there is an interrelationship between a lot of contextual elements and study performance outcomes and different things that affect each other in different ways. And so this is really just to illustrate that these areas can be fairly complex to examine.

And then I harken back to one of the examples that I had

given you to your question is that, here, we have an example of how blinding can relate to attrition bias. So this was the example I used where we had a patient on placebo. Physician/patient knew that. Patient became encouraged to leave the study because their drug isn't working. Now, we've got attrition bias because, had they not known that they were on placebo, they would have been more likely to stay in the study and then we would know what would have happened to them along the way, and it might have been that, in fact, the other drug does or doesn't work, but we're not going to have that revealed by a lot of attrition when it happens in this way. So attrition bias is rather complex.

Some other assessment bias issues: Invalid assumptions for modeling, models are not truth, and then problems with certain analysis methods.

So what I want to do now is direct everyone's attention to page five, not the slides, but in the first part of your book, page five. And again, what we gave you -- Mike and I spend a lot of time when we're training people in the methods. We've given you just a few snippets of some of the key things, just as kind of a sampler, but what I want to do right now is give people an opportunity to read page five and then ask us any questions that you're interested in.

UNIDENTIFIED MALE: Which page five?

MS. STRITE: It's near the front.

MR. STUART: It’s got a little underline under it.

MS. STRITE: It’s got a little underline, and it’s over to the -- it’s not fully the bottom, but the far right bottom. And at the top, it says, “Basics of Evaluating Evidence in Superiority Trials for Therapies.”

(Pause - group reviews document)

MS. STRITE: Do you have a question? Sure.

COMMISSIONER HARRELL: This is Colonel Harrell. I just would like to you to expound upon the second bullet in Section IV under “Assessment of Outcomes, analysis methods.” I don’t know that we talked too much about that.

MS. STRITE: Right. Mike mentioned Intention-to-Treat analysis, and is that coming up or not?

MR. STUART: No. (Indiscernible - simultaneous speaking)

MS. STRITE: Intention-to-Treat analysis is when you have dichotomous variables. So everyone, dichotomous would be -- it’s like either/or, dead or not dead, two choices. When you’ve got dichotomous variables, the Intention-to-Treat analysis has been agreed as being the preferred method as the primary method, meaning that you can still look at just who completed the study, for example, and that’s interesting information, but Intention-to-Treat requires that you look at everybody, and you look at everybody in the groups to which they were assigned.

Now a problem with that is, if you lost people, how do

you look at data for everybody? And so with Intention-to-Treat analysis, it requires that, if you've got missing data, you do something, trying to fill in the missing data. And we've talked about bias meaning anything systematically that leads away from the truth. Intention-to-Treat analysis is a little bit bizarre, in a way, because you are creating a bias, because you're making something up to fill in the blank information. But Intention-to-Treat analysis became -- developed as a method to try to deal with keeping randomization intact. And so it was considered to be sort of the best attempt to try to do that by accounting for all patients randomized and in their groups as they're randomized. You only do this for efficacy.

So I'm going to step away from efficacy for just a minute. For safety, you do not want to analyze people as randomized. You want to analyze them as exposure happens to the drug. But for efficacy, you're trying to give the drug a hard hurdle to prove itself.

The two methods that are used to try to fill in missing data when you don't know what happened to somebody, there are methods where they try to guess what happened to people and so they may come up with models where they look at all the patients remaining in the study, was this person similar to these people, what were their outcomes, and so might we predict that that what's would happen to them.

And then another method is to do a high bar, which is what Mike and I do. So we say, okay, what if -- and we usually don't use it this extreme, but this illustrates it. What if everybody who is missing in the Drug A group gets counted as a failure, and what if everybody in the comparison group, say placebo, gets counted as success? Now, we've made a really high hurdle, and in an otherwise valid study, if we still get statistical significance after raising the bar that high, then we say, okay, we can probably conclude that it's true that it's efficacious. Is that helpful?

So that's a primary method, and to look at just those who complete the study is great. Just those who followed protocol, that's great, but those should be complementary to ITT analysis as a primary study.

Mike is going to talk some though about a study method, an analysis method that we see all the time, which is something he mentioned before, which is Time-to-Event analysis. So I'm going to stop there with that, let him mention that.

The other thing in this bullet though that I'll mention is this predefined groups for analysis, and we're not spending time talking about chance today, but when we talk about pre-definition, the concept behind that is that you want to identify your outcomes for study before you start your study. You want to identify what you're going to set as statistical

significance. You want to identify, before you start, the sub-populations you're going to look at, a few things like that, and the reason why pre-specification is important is because it is reducing the likelihood that those things that you called out in advance are going to be due to chance because you've narrowed them and pre-specified as compared to, I'm looking at a sea of data and now I can make a lot of associations, and boy, this one is statistically significant, and oh, look at that, but they're statistically significant just because you're looking at a lot of data and you have a lot of chance effects.

So when you look at the P-value of less than 0.05, what that really means is you just have a less than one-in-20 chance that your finding is due to chance. When you start looking at more outcomes -- if I'm looking at two outcomes -- and I went three. If you're looking at two outcomes, you now have as high as a two-in-20 chance that something is going to be due to chance. By the time you're looking at 20 different outcomes, you've got, potentially, 100% chance that one of those is going to give you a statistically significant finding, merely because it's a chance effect. So pre-specification helps reduce the likelihood that that thing that you called out is going to be a chance finding. Is that helpful?

MR. STUART: And there are many other examples, last

observation carried forward. I mean, we just can't get into all the details about the appropriate analysis methods, but it's a very big area, and it's a complex area because of the Kaplan-Meier curves that you're going to run into, and we'll give you a little taste of that, but basically, we're just giving you a very quick glimpse of some of the problems.

I will say that some groups try to make it very easy on themselves. For example, ACP Journal Club won't review a study if they've lost more than 20%. That's really an attrition thing. So they won't even do an analysis. So we would say, well, that's a problem because there are studies that have lost 20% that we've seen that are not fatally flawed studies.

So you will see that Guyatt's group pretty much accepts up to 20% of lost data. Other groups, 20%. We do not say that. We don't have a number because we don't know a number. We go by the study and the contextual elements in the study and the other biases when we grade a study. The assessment part of it is a difficult area. We'll just say that that's kind of advanced work, wouldn't you say, to really understand it?

MS. STRITE: Yeah (affirmative). Well, yes and no. I mean, I think that some of the methods that we've talked about definitely learnable and understandable.....

MR. STUART: Sure.

MS. STRITE: .....just a bit more complex than some of the things that we've talked about. The other thing, too, is that Mike and I are going to talk about a few statistics right after this discussion, but you will see not many. We do not spend much time with statistics at all, and for the most part, we don't feel that we need to. So where we really concentrate our time is bias and chance, and if we need a good statistical concept, we'll get one, but I think that there is a lot of emphasis that's paid to that in the study, itself. And so generally, I feel that the statistical part is probably well done, and we don't need to go there.

MR. STUART: There's another question in the back. You had your hand up. Remember, turn on the light. Your name again, please?

MR. BLAS: (Indiscernible - away from mic) pay much attention to the statistics?

MS. STRITE: No. We pay attention.....

MR. STUART: No. We didn't say that.

MS. STRITE: .....to a few key statistics, but we don't spend time -- I don't want to get too beyond this group, but for example, there is a thing called Stopping Rules -- oh, no. (Indiscernible - voice lowered) is where I was headed. This idea that, for some outcomes, you may increase -- no; I'm sorry -- lower your P-value because you've already tested an outcome, tested an outcome, and there are different methods

for doing that, like (indiscernible - voice lowered) and blah-blah-blah. We don't need to know. That's -- to us, if we really are concerned about something that seems irregular, we find a statistician and consult with them.

MR. BLAS: The question is, when you finish the study, you have the biases, you have all the (indiscernible - away from mic) that you mentioned. The result is that you cannot extrapolate from the group that you tested. I mean, (indiscernible - voice lowered) the result of the conclusion. The conclusion is about the group (indiscernible - voice lowered), but you cannot say anything about the rest of the population. If you have all those flaws in the testing, you (indiscernible - voice lowered) 200 people participated in the study. The conclusions apply to them, but it doesn't apply to anybody else.

MR. STUART: That's a judgment.

MS. STRITE: Well, those are two different questions. First of all, if there are a lot of flaws in the study, we don't think it applies to those 200 people. So if we see a lot of flaws in the study, that study is going away, and it's hypothesis-generating, at best. If we have a study that we think is valid, then the issue, after that, is how you apply it. And then that becomes an issue of looking at the inclusion criteria, the exclusion criteria, and the Table of Baseline Characteristics to get a sense of, who are those

people? And so that was part of the PICOTS that we mentioned earlier, is paying attention to the patients and then trying to assess whether or not it's generalizable to a broader population or is it generalizable to just those people. Yeah (affirmative). So those are important.

MR. STUART: Other questions? Yeah (affirmative); go ahead.

MR. HOPE: Chad Hope. I do have one question or maybe something you can help me clarify. One of my pet peeves is when a study is set up to measure whether or not there is a difference between the study drug and placebo, say, for instance.

MR. STUART: Right.

MR. HOPE: The null hypothesis being there is no difference; the alternate being that there is.

MR. STUART: Right.

MR. HOPE: That is not a superiority study or is it?

MR. STUART: It is.

MS. STRITE: Yes.

MR. HOPE: It's my understanding that a superiority study is only looking at one-tail difference.

MR. STUART: No. No. You can do a one-tail, but we would not like to see that. We would like to see two-tailed because we don't know the results, but we're now.....

MR. HOPE: So (indiscernible - simultaneous

speaking).....

MR. STUART: .....we're getting kind of deep here.

MS. STRITE: Yes. Absolutely. Superiority just means we're looking at which is better than.....

MR. STUART: Which, of the two.

MS. STRITE: Which, of the two, is better?

MR. STUART: Yeah (affirmative).

MS. STRITE: And the problem -- so what he's talking about -- a one-tail t-test, as I understand it, only gives you a chance for one thing, for one of the two to show superiority. Whereas, a two-tailed gives you an opportunity for either to show superiority.

MR. STUART: Right.

MS. STRITE: And another problem with the -- and like Mike said, we want to know. I mean, what if we get a surprise and the placebo is superior.

MR. STUART: So we always two-tailed.

MS. STRITE: Yeah (affirmative). So we want two, but then the other thing is that the one-tailed t-test -- we had, actually, a study we reviewed with a statistician, and she looked at it, and she said, why did they use a one-tail t-test? She said, this is very strange to me in that context. And so I asked her, I said, well, could you do a two-tailed test for us, and she said, sure. And it turned out that the one-tailed t-test, we believe, was chosen because it gave a

bit of an advantage to be statistically significant; whereas when the two-tailed was applied, it ended up being a harder test, and suddenly, it was barely statistically significant.

So we always want to see two-tailed, and I know that this is going over the heads of a number of people here right now, but just for you and a few others that may appreciate this, for non-inferiority and for equivalence trials, we still want to see two-tailed tests and that's actually advocated by the European version of the FDA.

MR. STUART: Yes? Question?

MS. PENWELL: Vicki Penwell. I wonder if there are -- this is probably very elementary compared to this other conversation.

MS. STRITE: No; elementary is important. Please.

MS. PENWELL: Okay. You cannot watch television in the evening without seeing commercials for drugs and joints and all kinds of stuff, and I'm wondering if that exposure is going to make it increasingly difficult to find groups for studies that are not -- I mean, you know, like finding a jury pool that hasn't already been exposed.

MS. STRITE: Right.

MR. STUART: Uh-huh (affirmative).

MS. PENWELL: I'm wondering if we're creating a population of more biased than usual consumers?

MR. STRITE: Well, I say biased in a different sense than

we're using the bias here, but let us say, people may be more reluctant to participate and that may be affecting the generalizability of participating in research.

I'm going to answer your question a little bit sideways here. All research has flaws, and people just willing to participate in studies and allow themselves to be randomized makes them different than people that would not make that choice. And so the thing that you're describing, I would say, yeah (affirmative), that probably contributes, too, to a change of the kind of person who is willing to expose themselves, you know, randomly to drug choices.

Just for fun, I do this, sometimes, in some groups, and this is a fun group. So we're going to have some fun. Let's say that I'm a researcher and I've got a beautiful new molecule, as the pharmacists love to appreciate -- I don't understand what that is, but they get very excited about it. So my beautiful new molecule is myoceptimab and then we've got placebo. Okay. How many of you -- okay. You're all going to be randomized. You don't get to do informed consent here, but what you do get to say is, would you prefer to be on my new drug or would you prefer to be on placebo? So I'm going to ask first, how many want the new agent? How many want placebo? Don't be shy. Don't lie.

Well, I'm very impressed because you know what I usually see? Guess what I usually see? Way flipped. And we've been

in groups where we've spent, like, two days with, you know, a bunch of doctors and clinical pharmacists, because those are the groups we usually teach, and all of them but me, you know, want the active agent. Oh, wait. I know what the set up was, it's because you knew I manufactured the drug. That's what made you suspicious about it. Well, anyway, that's pretty impressive because, usually -- and it is in just group we train, and I'm, like, did they not hear us? It is human nature to want the new thing, you know, the latest and the greatest and all of that, and if you've got a lot of people exposed on the news to the latest and greatest, that's what they're going to want. That's what they're going to want.

But then the big issue with direct-to-consumer advertising -- and I have to confess I don't watch TV. So I actually don't get exposed to this, but I know, for a fact, that, when that became allowed, a lot of things in medicine changed. A lot of pharmaceutical dollars started moving away from directed to physicians and directed to patients. That changed a lot in terms of -- the doctors can speak to this much better than I -- patients showing up, about asking that the docs may not even have heard of yet, let alone have any chance to know if the studies were valid are not. And so the patients are not, you know, I feel, getting helped by that, personally.

MR. STUART: Question?

DR. BUTLER: I think you just did a great social experiment in kind of the recruiting side because, if you had presented a study of the side effects and adverse events associated with the drug, I would bet you would have more people choose the placebo.

MS. STRITE: Well, that's why I've been really surprised in groups we've trained because we had spent all this time. So this is not a training. This is more discussion, right? In the training, we're going through case-after-case of these awful things and then everybody at the end, again physicians and clinical pharmacists, are saying, oh, I want the active agent. (Indiscernible - voice lowered) randomize to, so.....

MR. STUART: Culturally, we are that way. I mean, we want the newest because it's the greatest. And in fact, science is not an intuitive thing, especially when it comes to cause and effect. So our observations, as doctors, we just believe there is a cause and effect because we saw that patient say I'm better, but we don't see the science part that we're not even thinking about. So you're absolutely right.

MS. STRITE: And I think it's okay to want, you know, the latest and the greatest, but we want good information about the likelihood of benefit and harms for that, too. So any other questions? Okay.

So now, I'm just going to mention that -- bear with me here just a nanosecond. I'm going to talk some about some of

the statistics that we do use, and feel free to follow along with the slides as you want. And so we'd be restarting up on slide -- so again, the second part of your book -- 118. But just to let you know, you have this information in a one-page condensed form, along with some other information on page ten, in the front part of your tools and your one-page summaries.

So this is just going to be a quick tour through the key statistics. So we've all gotten extremely good at naming the five clinical outcomes, but then there is this other key thing and that is, what is the size of the outcomes? And so that's what I'm going to be addressing right now.

But before I do that, since we all, I guess, watch TV or read newspapers or get exposed to some information in other ways, if you read in the newspaper that global improvement scores were 30% higher in people with fibromyalgia who exercised compared to patients who did not exercise, what should your first question be? And again just looking at this statement, not questioning other things outside of this -- so we'll just make it a given that you'd question if the study is valid, but just on that basis there -- and let's say you have fibromyalgia -- are you going to say, well, now I go exercise? What might be something that you might look at in that statement and question?

UNIDENTIFIED MALE: Randomization.

MS. STRITE: Okay. Let's say people were randomized, so

that's good. Uh-huh (affirmative)?

MR. KNAPP: Well, it sort of depends on the scale of either -- there are sort of relevant scales. If only one percent of the people got any better anyway, then it isn't very significant. If it's very tiny, 30% above something very tiny is not very important.

MS. STRITE: Okay. So somebody paraphrase what he just said. How come he's talking about one percent when I don't have one percent up there, I've got 30%? Why did he just say that? You're absolutely correct.

UNIDENTIFIED MALE: (Indiscernible - away from mic)

MS. STRITE: Well, no. That's not what he's getting at.

COMMISSIONER HIPPLER: Relative versus absolute.

MS. STRITE: Exactly. Exactly. So it was the same issue that we saw with the Vioxx situation where it was 60% benefit, but it was 60% of less than one percent. And the reason why I have that up there is that, see how natural it is to look at that 30% and say wow? And so if I want to sell you a newspaper, if I want to sell you a drug, if I want to sell you on, well, anything, if I want to sell you on wearing your seatbelt, all I have to say is -- and I'm making this up -- 65% better off if you wear your seatbelt when you get into an accident. But the question that you want to ask is, what is that 65%? Is that an absolute number or is that begging the question 65% of what? And so I'm going to show you how to

calculate the absolute risk reduction, and I'm going to show you then what happens with the relative risk reduction.

So now, you have to be skeptical of everything you read when you see those numbers. The question is, what is that measure of outcomes?

So I've mentioned before there are various measures of outcomes and that's what's detailed on page ten there, and we're going to go through any, but a couple. But when I showed you that Forest Plot, we saw odds ratio. That's a measure of outcome. Relative risk reduction, that's a measure of outcome. So a measure of outcome is just a term for some way in which the difference between the groups is described. So you can take the same data and just describe that in different ways. That's what a measure of outcomes is all about.

So up here, we've got a study in which, in our control group, 15% of the people died, and in our study group, 10% of the people died. And then I'm going to just sort of stipulate that everything we're going to talk about now we're going to assume that it's a valid study. So now, we get to look at and evaluate the statistics.

What I want to pose to especially anybody who is helping patients make decisions is, as a patient, I want to know those four numbers. That information is the absolute most meaningful to me, and anything that is other than that is not

really giving me clear and complete information.

Imagine, if you will, if I told you that 85% of the people died in the control group and 90% people died in the study group. Now setting aside which agent is better or less safe, my point is, is that, if I'm a patient and I know that I have a five percent chance of some improvement and if I know that, if I don't take anything, that I have a 50% chance of dying, I may say that is such a small difference to me that I'm not willing to take the drug because there will be side effects, possibly. There may be unknown side effects. But if I'm in a situation like this where this is flipped and this is the mortality, my personal choice and preference might be, wow, it looks like I'm going to die no matter what. I'm going to try anything, hoping that I'm going to improve my odds.

So my point is that those four numbers may put a patient in a very different relationship to what may happen to them. And so this measure of outcome is what's called risk with and without benefit.

Now, the next valuable piece of information, I think on top of that, is we keep mentioning the absolute risk reduction. The absolute risk reduction is simply taking these two numbers as percentages and simply subtracting. So when we talk about absolute risk reduction in this example, it's five percent, but again, that five percent may feel very different to me, as a patient, depending upon what those numbers

actually are.

What relative risk reduction does is it takes the relative difference in size between the outcomes. And so in this instance, ten is one-third smaller than 15, one-third equals 33%. Ah, now, I can sell a newspaper because, instead of saying, oh, it's a five percent benefit, now I get to say it's a 33% benefit, and we all know what's better, right?

So with relative risk reduction, it really is not -- I don't know. I don't know how important people find it. So I honestly don't know if it really has a use other than selling something, but to me, if I have this information and this information, I'm very happy. Some people may find this a benefit, but also one of the features of relative risk reduction is, mathematically, it can't turn out to be equal to the absolute risk reduction, but it can never, ever be lower. So almost always, it's going to be higher, and almost always, it's going to be a lot higher. Any questions on that?

So it's the equivalent, again, of my saying that I've got something 90% off and then not telling you what the baseline price is.

MR. STUART: Okay. I'm going to come back to the Kaplan-Meier curve again because you will see them, and I just want to make a couple more points about time-to-event, just a couple. I could spend all day.

Here, we have a time-to-event analysis. Sheri put a

thumb between those because a doctor, at one point, said, yeah (affirmative), I love them, if you can get your thumb between them. But the problem with that is that, if you look at the scale, you can see how distorted the difference between these groups are. You go from zero to five percent, and it should go up to 100%. So that's just -- that isn't the punch line of this slide, but I mean, in this particular one, that's just something that you can't help but see.

So in a time-to-event study, let me start by why a lot of doctors like them because they like to look at the results for a patient. Let's say I'm an oncologist, and I have a patient on a new drug, and I want to say, at six months, we have whatever this comes across -- I'm going to just make it up to be one-plus percent dead. I'm just making this up now, but in this group, it's almost -- well, it's over two percent. So they like to be able to go from the time up to what the likelihood of mortality is, and in this case, it happens to be -- Kaplan -- time-to-event curves can be anything that's dichotomous. So it could be -- in this case, it's ulcer complications. It could be time to a reinfection, time to death, time to pregnancy, time to anything, but they're time to an event. So that you can look at the time on this scale and look at what the likelihood of the event is on this scale. That's why healthcare people like them is they can look at the results at different times. And there are a lot of

assumptions that these things stay stable, the differences stay stable between the groups, but that isn't really true. You have to get into hazard ratios and things and understand them. We're just not going to do that.

What I want to show is that we start out with 4,000 people here and we end up with 500. We start here -- same thing in the other group. As the head of the DOD Pharma -- what's it called, the pharma.....

MS. STRITE: Pharma/Co-economics Center.

MR. STUART: Pharma/Co-economic Unit said, where did the people go?

MS. STRITE: Where are the patients?

MR. STUART: Where are the patients? Did they all die? Did they all (indiscernible - voice lowered)? No. They got censored. So I'm back to censoring, but not in the way we talked about. This is censoring for Kaplan-Meier time-to-event curves. All right.

Censoring is, essentially, removing people from a study. I'm the investigator, and I can say they broke the protocol. I can say they had an event 14 days after they stopped taking Vioxx. I could make up any censoring rule that I want. What's really odd is that we, as consumers of information, don't demand that they tell us what their censoring rules are. Has anybody in this room seen a study where they actually lay out the censoring rules?

MS. STRITE: And how many people have seen Kaplan-Meier curves?

MR. STUART: Oh, they've all seen them.....

MS. STRITE: I know.

MR. STUART: .....but they may not know they've seen them, but they've all seen them. And so we have a hard time understanding lots about censoring, so we called the statistician from George Washington University who has probably written more about this than anybody and said, please teach us; we don't understand it. He said, I'm going to give you an easy way to understand censoring. Censoring is the way of -- the reason censoring got started in time-to-event studies was that people are enrolled at different times into a study, and the study may go two years, but I might enroll somebody six months before the study is over.

Now if that person has not had an event by the time the study is over, we take them out of the study and that's called administrative censoring. Some people call it right-sided censoring, and we don't need to spend details on this. And that doesn't really necessarily create a major threat to validity, but this one does. This is not administrative censoring, and this is where the investigators invent rules, like I showed you with that previous slide with the Vioxx where they took the people out if they had had an event after 14 days. Now that's not administrative censoring, not just

late enrollment and taken out of the study. That's actually patients who I make up a -- I made a decision that I'm going to take patients out for whatever reason I want, and if I don't see that in the protocol or in the study, I have a real problem with that Kaplan-Meier -- that time-to-event and that's my biggest problem with Kaplan-Meier curves is I can almost never get the censoring information. So we have to then contact the authors, et cetera, et cetera. So it makes it very difficult.

Censored patients are assumed -- did you see the censored patient there? Censored patients are assumed, on average, to have the same risk as patients who are not censored. Does anybody believe that that is true? I mean, we don't -- there is nothing that convinces me of that. So there are these assumptions that go into time-to-event studies and analyses that are questionable.

Time-to-event studies are also called life tables because they're actually, historically derived from tables. Now computers do this, and investigators never see these life tables. They just see what comes out of the computer. They just throw their data in, out comes their curve, and their data has been analyzed for them.

Here is the way it really works in time-to-event in a Kaplan-Meier curve. I'm going to start with eight patients, and I have a time interval, the first time interval, which is

one, which is up here, time zero to one. I start out with eight patients here. That's 100%, which is eight. I don't censor anybody, meaning you don't see any vertical lines that represent censored patients here in this first time segment. Somebody dies. That triggers the computer to calculate, and it calculates seven out of eight survival, which is 0.87. So at this time place right here, there is 87% of the patients surviving. I just keep doing this until another patient dies and then the computer calculates again, and I multiply that 0.87 -- I'm sorry I said 78, but it's 0.87 -- times this 0.8, which I get from this segment, but in this one, I censored two patients.

So I just want to give you an idea about how this is done. I'm taking patients out when they -- for any reason I want, and I do a calculation when somebody has an event. So there are two things going on at the same time.

Now just to show you what happens, the red line shows what happens if I censor, which is the black line and what happens with the curve when I don't censor. So you see, I get different curves, and in a very long study, I'm going to have quite a difference depending upon whether I censor or not. This is probably a little too much for most of you, and I'm going to drop it at this point, but I just want to show you that, when you read a study and you see it's a time-to-event analysis, you may want to get a little help from one of your

colleagues if you don't understand the censoring because it's really important to look at a study that's a time-to-event study or a Kaplan-Meier study with the thought of censoring in mind. So that's the bottom line.

A couple things about safety. I said I'd come back to it, and here it is. One, it's difficult to assess. The practice -- we almost have to always rely on observational data because the events for safety are frequently so rare in randomized control trials that we don't have enough power to show a difference between the two groups.

Safety population is not the same as the efficacy population. Sheri told you that, if they're randomized, they should be analyzed for efficacy. Not so with safety. They should receive, at least, one dose of the drug or the intervention to be in the safety analysis. So it's a different set, data set for the safety analysis.

Doctors tend to not report safety issues, at times, because they're really instructed frequently to focus on the efficacy and that's just a problem in studies because it's not a formal safety evaluation many times, even though it might be in the protocol.

And the duration issues. What that means is that there wasn't enough time to find the safety issues. I saw a study a couple years ago. I don't know if it was any good, but it just illustrates this point. The lithotripsy for

ureterolithiasis, kidney stones, somebody looked at the number of diabetics who had had lithotripsy and matched them -- this is an observational study -- to those who had diabetes but didn't have lithotripsy, and it looked like lithotripsy might be shaking the pancreas and the adverse effects of diabetes didn't come out for ten years. I'm not saying I know that because that's the problem. It's an observational study, but at least, it's something that we would want patients to know about. For most safety issues, you should know that we don't have strong data, but there are some signals or there is some slight evidence or weak evidence or there is a suggestion that there might be a safety issue. I think that's all I need to say about safety.

Power. What I told you about power is that it's a way of stating that you have enough people in your study to find a difference, if a difference actually exists.

MS. STRITE: And when Mike "if a difference actually exists," that's shorthand for a statistically significant difference.

MR. STUART: So let me give you a hypothetical. I have decided to do a study. I'm an investigator. I do a power calculation, which responsible investigators should do before they start their study to see how many people should be in the study because it's not a very good thing to ask people to come into a study, if you know that there is no chance in finding a

difference between the groups because you're not going to have a budget to get enough people in it.

So anyway, you do a power calculation to find out how many people should be in the study. So I've done the power calculation. I, now, have done my study, and I see no difference between the groups. Does that mean -- well, let's do it the other way because it's a lot easier. I don't get as many people as I need as stated by my power calculation. I can't recruit that many people. So I have a smaller study than I wanted. I didn't meet my power calculation number, but when I did a well-done study -- it's valid -- I see that there is a difference, and it's statistically significant, and my new drug is better, and it's statistically significant, and it's clinically meaningful, and it's a well-done study.

What can you say about the power? I didn't need my power calculation; you could say that because I already told you that. What else can you say about the power? Was the study powered? I have a no. Can I get a yes out of anybody?

MS. EDMISTON: If you've reached the statistical significance.....

MR. STUART: You got it.

MS. EDMISTON: .....it is enough people.

MR. STUART: Very good. You want to grab that microphone, turn on that light, state your name, and say that real loud.

MS. EDMISTON: Just your third bullet there, if you reached statistical significance, by definition, you had enough people. So yes; you did.

MR. STUART: So I don't need a power calculation as a reader. Okay. So one of the things that we've seen is critical appraisers write studies down because they didn't do a -- they didn't meet their power calculation. Well, that's a mistake, if there is a statistically significance difference, and she got it exactly right. By definition, it was power. I had enough people to show a difference. Does everybody see that?

But if I have a study and I don't see a difference between the groups, what are the possible explanations for there being no reported difference? I've got to remember Sheri will correct me if I don't say "reported difference" because we don't know that the real difference is the report difference. So if I see no report of difference, what are the possible explanations for that?

One is I didn't have enough power. Great. Another one.....

MS. STRITE: Which means people. Power means people.

MR. STUART: I didn't have enough people. There is one more. Yes?

UNIDENTIFIED FEMALE: There's no difference.

MR. STUART: There's no difference, and it's true there's

no difference. So that's -- if you realize the things we just said and you understand that the confidence intervals can help you with the results much more than the investigator, you have more knowledge than the investigator because he doesn't have the study. You have the study, so you don't need the power calculation. All you need to do now is, if it's a well-done study, go to the results and see if it's statistically significant. Okay. That was good.

MR. KNAPP: Can I just make an observation?

MR. STUART: Yeah (affirmative).

MR. KNAPP: So my name is Gunnar Knapp. I just caution that that doesn't mean that statistical significance means that you might not have been better off with a larger study that could have shown you more.....

MR. STUART: Oh, sure.

MR. KNAPP: .....about the scale of the effect.

MR. STUART: Thank you. Good point. Okay. Well, Sheri illustrates this, this way. If we have 100 people here and we have one person who is going to have the event, we can see that in this diagram up here with the little frown face. So that's one in 100. But if we have a small study, we may miss this frown face because our study is too small. Does everybody see that? Okay.

Now, confidence intervals. If you have a well-done study and it's valid, we've showed you those Forest Plots and we

showed the point estimate, but in fact, that point estimate may not be where the truth is. It could be anywhere within the brackets of the confidence interval. So you should know that it's statistically more likely to be where that point estimate is, but it's not necessarily there. It's a bell-shaped curve, and it could be out further in the confidence interval.

So if I have an absolute risk reduction of two percent in a study with a 95% confidence interval ranging from one to three, the absolute risk reduction might really be one, but it could be as high as three. Does everybody understand the confidence intervals?

That's really why we prefer confidence intervals to P-values. P-values tell you whether something is statistically significant. Confidence intervals tell you the range within which the true value may lie, but it also tells you whether it's statistically significant. As we said, if it's touching that line of unity, then it's not statistically significant.

And in an absolute risk reduction, which is a percent, what would the Line of No Difference be? Would it be one or zero?

COMMISSIONER HIPPLER: One.

MR. STUART: Why would it be one? Is it a ratio or is it a percent?

COMMISSIONER HIPPLER: Percent.

MR. STUART: If it's a ratio, it's one. If it's a difference in percentage, it would be zero. Does everybody see that? But I don't have a Forest Plot here. I have to use my head to say it's a percentage, absolute risk reduction, and I just look at the parentheses because I don't have that Forest Plot with that Line of No Difference. Does that make sense to you, what I just said? Yeah (affirmative). Okay.

This comes into play. I was at a meeting in Salem, Oregon. An orthopedic surgeon came up to me and said, okay, I do kind of like this evidence-based stuff, but I've got to tell you I read an article about fonda.....

MS. STRITE: Fondaparinux.

MR. STUART: Which -- fondaparinux versus enoxaparin, and it said there was really no bleeding difference. And to an orthopod, a bleeding difference for a total knee or a total hip is a big deal.

We did a guideline with Kaiser Hawaii, and this bleeding issue was -- we learned very quickly that a one or two percent difference in bleeding is a big deal to an orthopedic surgeon because you get blood in the knee, you can't move the knee. You get a longer hospital stay. You get fibrin in there. You get a stiff joint. You get all kinds of problems, a higher risk of infection. So the amount of bleeding into a joint after surgery is very, very important to orthopods.

And so he said, I saw the study, and the study said there

was no difference in bleeding between these two drugs, and I'm telling you, every time I used fondaparinux, I find I get more bleeding. And I said, well, you know -- I didn't want to get into a discussion about how observations can be misleading. And so I said, well, I'll go over the literature. And I said, okay, here's the study. It's *Lassen, et al.*, and they say -- the authors say the two groups did not differ in clinically relevant bleeding. And I submit the authors mislead that orthopod, and here's why.

The actual bleeding rates were 47 out of 1,140 in one group for -- that's 4.1% of the parinux, and 2.87 for enoxaparin, but the P-value was not statistically significant. So the author said, no difference. Well, that's no statistical difference. That does not mean that there is no clinically relevant difference. So you have to look at the confidence intervals to see what could be.

So we -- and what's really interesting about the orthopedic literature is almost no study that we ever looked at, for total knee and total hip for this guideline we did for Kaiser, gave the confidence intervals for the bleeding. So we had to calculate them, and it was interesting. In Hawaii, what we did was we had everybody in a room like this at a big, huge table, and we had groups at each corner of the table with computers, and we had a stack of papers to critically appraise. We taught them how to critically appraise the

literature. They'd grab a study, go their group, do the best they could to critically appraise and grade it. And if they needed a consult, they'd ask the rest of the groups and us, and we would do consults just in time to get them through. And we critically appraised how many studies?

MS. STRITE: Forty-seven studies.....

MR. STUART: In how much time?

MS. STRITE: .....in one afternoon and one-and-a-half days.

MR. STUART: In one afternoon. Yeah (affirmative). Okay. One-and-a-half days.

MS. STRITE: So we spent half-a-day doing really fast training in this and then just had that. And then we came back for another two days, and there were a lot of phone meetings before we got to this point though. But we came back and then did the clinical recommendations, graded it, and created an algorithm over the next two days. It was one of the fastest guideline development activities we've ever been involved in.

MR. STUART: We had already done the search for them. We did a lot of backroom stuff for them. So here are the confidence intervals. The absolute increase in bleeding, ARI, was 1.3, but the confidence intervals went from 0.21 favoring fondaparinux to almost 3 favoring enoxaparin, and they said that's clinically meaningful. And since it could be as great

as 2.8%, everybody agreed the authors' conclusion up there about no difference is totally misleading. So it's just another one of the little items I wanted to show you about confidence intervals, how they can be used, in this case, for safety.

And we've talked about Cochrane a number of times. For those who would like to know more about how Cochrane thinks, you know, we have our little manual. They have their little manual, and in theirs, they point out -- I'm just taking one thing from them, just to show you it's not just us.

"A common mistake when there is inconclusive evidence is to confuse 'no evidence of an effect' with 'evidence of no effect.' When there is inconclusive evidence, it is, quote, wrong to claim that it shows that an intervention has 'no effect' or is 'no different' from the control," which is what they did. So they -- Cochrane would say that was totally inappropriate for those authors to do that.

Let's stop here and see if there are other questions about anything we've covered today.

MS. STRITE: And we're getting close to break time, too, just FYI.

DR. BUTLER: This, I think, is a question more related to the systematic analysis. Could you go back two slides?

MR. STUART: Sure.

DR. BUTLER: Was the lesson analysis included in the

systematic review?

MR. STUART: Was the -- oh, the bottom bullet is a -- I didn't go through this, but I put that last bullet in there to show you this is confirmatory evidence that fondaparinux may, in fact, be associated with more bleeding because, when you look at a lot of studies, there was a statistically significant increased bleeding rate with fondaparinux.

DR. BUTLER: That's the root of my question because.....

MR. STUART: Yes. It was included.

DR. BUTLER: So they contributed a third of the patients going into the systematic review.....

MR. STUART: Yeah (affirmative). I believe.....

DR. BUTLER: (Indiscernible - simultaneous speaking) the bleeding in the fondaparinux group was lower than in the enoxaparin group in the *Lassen* study.

MR. STUART: Yeah (affirmative).

DR. BUTLER: It seems like there is a lot of mix in different kinds of studies of different patients.

MR. STUART: Well, you are right about that also. You are right about that, and it's one of those issues with PICOTS. You weren't here when we talked about PICOTS. PICOTS is population, intervention, comparator, outcomes, timing, and setting. And when you do a systematic review and you include lots of variation in those PICOTS items, you're going to get different study results and that's one of the problems we have

with systematic reviews is that the one thing that we don't get told is the quality of the studies they include in the review or they use the Jadad scale. And I don't want to go over everybody's head, but you know what the Jadad scale.

MS. STRITE: Jadad.

MR. STUART: Oh, okay. Well, it's not a good scale and so it's quick. It's easy, and it's flawed.

MS. STRITE: And it's fast.

MR. STUART: So essentially, what we're saying here is that that's a little bit of confirmatory evidence, really, for what this orthopod was saying, but that wasn't the bottom line. The bottom line is the confidence intervals tell you the range. And I now believe that it's possible fondaparinux is associated causally with more bleeding. Other questions about anything?

MS. STRITE: Then let's take a ten-minute break. And I need to consult with you. Turn off your microphone.

2:05:54

(Off record)

(On record)

2:19:01

MS. STRITE: So what we're going to do -- you can even get a head start on it, if you turn to page 48 and that's in the first part of your book. And page 48, Mike and I told you that we love Myoceptimab, and we use it for absolutely

everything. And so now, you're going to get a chance to do a critical appraisal. Now again, we covered topics -- critical appraisal concepts in an extremely high and general level. There are certain things that we didn't discuss, but we can all participate in doing the critical appraisal.

So what we're going to do is Mike and I are going to give you a little time to read, I believe, it's three pages on page 48, 49, and 50. And when Mike and I are doing critical appraisal work, if the study is going to be a Grade U, we want to get there as fast as possible. So what we don't want you to do is to look at this like you might if this were in practice. We want you to look at this in terms of critical appraisal opportunities. And so in this instance, even if you found a lethal threat to validity in which you would say, oh, I'd throw that study out, what we want you to do is actually look for as many threats to validity as you can, and you are, again, encouraged to use a tool, so either page five or page 25, and you are encouraged to actually confer with your table mates, if you like.

One of the things that Mike and I feel is that critical appraisal is best done as a team sport. It is a process of discovery, and different people will discover different things along their journey. We find that it's great to team up people from different professions because they see different strengths that people with different training and backgrounds

can contribute. So what we're going to do is we're going to leave you alone to work this study over and then we will come back, and we will debrief it. Anything, Mike, you want to add?

MR. STUART: Yeah (affirmative). Just find as many biases, or if you want to name good things, we will accept that, too, but we want you to find as many things to comment on as you can, not just it's a good study or it's a bad study. We want to know the details. What is the flaw that you're talking about? How big a flaw is it? Or they did this important thing correctly, and it's very important that they did. We don't care how you do it because this isn't a real training day.

Normally, we would show you how we do it and kind of model our different approaches, but since we have different approaches, you know you might, too. So it's however you want to do it. Just make sure you look for the important things to comment in this study, and we'll debrief it. It will be putting everything kind of together, and the debrief will bring out lots of discussion, we think.

MS. STRITE: And also, too, since everyone is here now, I have an announcement, and my friend who told me -- help me out here -- Dowling Street -- Dowling-something-or-other is closed. And so people might be helped by knowing that and think -- while you're doing a critical appraisal, think of

alternate routes, if that would be important to you.

2:22:43

(Group session)

2:48:15

MR. STUART: All right. So let's just do this open-endly. What do you think about the study?

COMMISSIONER HIPPLER: We like it.

MR. STUART: You like it.

UNIDENTIFIED MALE: What's the name of the company, so we can (indiscernible - away from mic)?

MS. STRITE: Delfini. And how often do you see a study that you like?

MR. STUART: Look how sure he was. He's confident. He knows what he likes. That was good. Is it kind of a surprise that we didn't give you something you would hate, I mean, that you would find full of holes? Well, let's just go through it and see what we've found. Tell us about the study in detail. Why did you like it? We'll start with you. And just give us, however you want, any specifics about the study that are good or bad, but be as specific as you can, and when you tell us about a place in the study, please tell us the page and the column and the paragraph. So if I were talking about methods, I'd say page one, left column, second paragraph. So red light on, name, and away you go.

MS. STRITE: A lot of instructions.

MR. STUART: Yep.

COMMISSIONER HIPPLER: This is Allen Hippler. We represent Table 2.

MS. STRITE: Oh, what's behind door number one?

COMMISSIONER HIPPLER: So the first thing we really like was we felt like there was no selection bias. It was very well randomized. On page one of three, the second paragraph on the right side, the allocation was made using computer-generated sequences. So that's the first thing we liked.

MR. STUART: Very good. Do you want to keep going? We'll let you finish the selection bias section, if you want, Table 2.

COMMISSIONER URATA: Well, we didn't think there was selection. This is Bob Urata.

MR. STUART: Right. I'm talking about the evaluation of the selection bias issue. So basically, we've already got the fact that it was random and randomly assigned using a computer-generated sequence. So we've got -- in my book, I'm a little more pattern-oriented than Sheri is, for people who are just learning us. So I'll just say generation of the sequence is good, and it gets -- lowers the bias rating. There are three other things we -- well, at least, two other things we look at for selection bias.

COMMISSIONER URATA: Yeah (affirmative). It was blinded to the person that was giving out the medication. The

medication was also identical between placebo and active drug. And the people were blinded to -- the assessors were blinded to who got what and also the pharmacy personnel were, except for the head pharmacist. They were also blinded to whether it was placebo or active medication.

MR. STUART: Excellent. Now, what we did was we, first, got part of selection bias. So I consider the selection bias checklist not complete, but performance bias, what you did was you picked blinding, and you actually, I think, might have even talked about concealment of allocation when you were talking about blinding. I'm not sure if you did or not. But basically, yes.

MS. STRITE: Not yet.

MR. STUART: I agree that the non-study drugs were similar and the dosages were similar. The -- what else did you say? The intervention seemed to be equal in both -- is that how you said it?

COMMISSIONER URATA: Yeah (affirmative).

MR. STUART: In both groups, yeah (affirmative). Anything else you want to say?

COMMISSIONER URATA: (Indiscernible - away from mic)

MR. STUART: And the assessors were blinded, right?

COMMISSIONER URATA: Yes.

MR. STUART: Okay. So we've got a little bit of assessment bias. We've got a little bit -- I mean the rating

of that. I'm not saying -- I'm saying we're covering these areas, not that there is a bias, that you're addressing that bias area. And the way we'll do it is any way you want to do it. So we're collecting this information because, at the end, we're going to rate the quality of this study. All right. We had some other comments. Yes?

MR. KNAPP: This is Gunnar Knapp, and for our table, at least, some of us were confused about whether the I think what you call -- was -- were co-interventions the same.....

MR. STUART: Yes.

MR. KNAPP: .....and you just characterized the use of other drugs as being the same and that wasn't the way I read it. It said -- the way I read -- and I'm referring to page one, column -- or the fourth thing -- or bottom, left paragraph. Patients were allowed to take their usual medicines sort of in any way they wished. And I would just wonder, might they possibly take medications in response -- might their use of those medications be related to -- how they use them be related to whether or not they were receiving the placebo or the drug and might that, in turn, have an effect on the study? And so I'm -- that raises a red flag, in my mind.

MR. STUART: Right. So let me see if I've got it right. You're saying that you're not sure that the dosages and everything were the same in the co-interventions, the non-study drugs, because they were using their usual care, which

may have differed between groups.

MR. KNAPP: And in particular, they had the ability to say, well -- they had -- the patient had the ability to choose dosages of these other drugs and might conceivably say -- might -- what might have gone on was the fact that I am taking this drug that we're supposed to be testing might affect my use, the likelihood of using other drugs. And so the effects I observe might be not due to taking the drug at test, but the -- my use of the other drugs.

MR. STUART: Right. Okay. Got your point. Now, let's go to page two, left column, results. I'm going to start with the second sentence. "Baseline characteristics (table provided) -- although we didn't put it in here -- including medications, dosages, clinical and demographic variables were similar in the two groups." So what we were attempting to do is remove your objection there, but you didn't read it the same way that I tried to write. So I may have to rewrite this. Your point is well-taken.

MR. KNAPP: I still wasn't sure whether statistically -- whether the statistical concern is actually removed.

MR. STUART: Okay. That's a very good point, and I'll accept that. That's a potential threat to validity. Yes?

DR. MANDSAGER: So with.....

UNIDENTIFIED FEMALE: Name?

DR. MANDSAGER: Dick Mandsager. So with patients that

are this sick, the baseline characteristic was achieved, but there is no mention whether, throughout the study period, the other.....

UNIDENTIFIED FEMALE: What's your name?

DR. MANDSAGER: .....the other drug stayed the same, and for 36 months with these sick patients, that's pretty -- it seems unlikely to me that everything stayed the same at the end of the study, as stated as it was at the beginning.

UNIDENTIFIED FEMALE: Did you say you're Bruce.....

DR. MANDSAGER: Dick Mandsager.

UNIDENTIFIED FEMALE: Oh, Dick. Sorry.

MR. STUART: Okay. The -- let's see. There was 98% adherence to myoceptimab and placebo. We don't really have the information downstream from the baseline characteristics as what you're saying. What you're saying is the groups could have become unbalanced. They could have become unbalanced during the study, even though they were balanced through randomization. When you speak, please turn on the light and your name also.

MS. WRIGHT: Sorry. Peggysue. The last sentence under "Results" on page two, bottom, left corner, unfinished sentence, "Antihypertensive agents, individualized therapy," blah-blah-blah, "supplements taken during study were similar in both groups." So you did address the continuation with the medication over the time period of the study.

MR. STUART: Yes. That's what we tried to convey was that things did stay the same, but he's making the point that he's not convinced. The first point was that it was may be statistically different, even though they were similar. We said similar. We didn't say -- and in a way, this is a judgment call because I'm not sure I know whether statistically significant differences in this setting are meaningful or being similar is good enough. So there's that word "judgment." He made a judgment. You made the opposite judgment, which is more like what I did, which was to say things really were similar. And I will try to rewrite this, so it's clarified that they remained the same throughout the study. But I agree, in the real world, it would be very hard to do that, to keep things the same, but they might be able to, if they really recruited people and got them to agree to the protocol and said, now, it's really important that, if you're in a study, you try really hard to get all these things that you want to be the same to stay the same. So it's a really good point. Both you guys, it's a really good point because, in reality, that's going to be one of the biggest problems in the studies you see. So it's really well -- your point is well-taken.

MS. STRITE: And I'm going to make a comment here, too, because this is an area that tends to be one of the harder areas to assess in studies. We're always really happy when we

see that investigators have provided us with a list of the medications by group and then give us more information about that. Ideally, you'd get dosing information and duration information. Never do you see that. Sometimes, if you're really lucky, at best, you get the percentage of use between the groups. So again, that's lacking the dosing and the duration, but it kind of gives you a hint.

A couple of things that I look at -- and remember my expression is -- there's a slide I had there on missing this in attrition. Here's where I look at some other study elements contextually and say, well, okay, how well do I think this study was blinded because, if there was a blinding problem, then it might be the case that we really do have differences between these groups because, if I know I'm taking placebo versus myoceptimab, I may be more likely to take some other agents. And when I look at this, I think, well, okay, how would they get this information? It would be self-report by patients. And if I'm a patient that maybe was biased to take more drugs because I'm on placebo, maybe I would be more likely to hide that and not report that so well.

So when you look at some of these things, sometimes, there are some other clues in here that might be meaningful. So I'm just kind of showing an approach of sort of how you sometimes can put various other study elements together to try to determine whether or not you think that it's likely that

you can depend on the information or not, and it's sometimes a guess and.....

MR. STUART: Sometimes, it's just a judgment based on the information you're given, and you just have to do your best, and you're not sure. Yes?

COMMISSIONER URATA: I just wanted to point out that -- it's Bob Urata. I wanted to point out that you did say four percent of subjects in both groups failed to complete the trial.

MR. STUART: Right.

COMMISSIONER URATA: So that would suggest, perhaps, that the issue of, you know, the illness (indiscernible - background noise) might not have affected either group unequally?

MR. STUART: Right. Right. Anything else?

SENATOR COGHILL: Senator Coghill. As a policymaker, the two things that I picked out that I thought were negative, right away, or at least, questions mark in my mind, were, why 55? You know, if you look at what the purpose of the study is, it's exercise-related issues, not age-related issues. And so that was question number one.

And then looking at the very end of it, I noticed who paid for and consulted on this particular study. So those two things were the first flags I saw.

MR. STUART: The first one is a really interesting one.

Your second point I want to comment on. The *Annals of Internal Medicine* maybe two months, three months ago did similar exercises. They published that they had given to, I believe, attendings -- it might have been residents and attendings -- and the Editorial Board of the *Annals of Internal Medicine* came down with the opinion, in their discussion about these cases that were slightly different than these, that the drug support by itself, unless there is clear involvement, should not be a major threat because you should go by the actual biases that you find in the study rather than the fact that they got money.

Now if they're doing the analysis, which they do, if they're doing other things where they are in control of data, I would agree that becomes more worrisome. But this is not -- there is agreement on this. I would say, at least the *Annals* and other people, including us, go by what we find in the study, and sometimes, that includes a bias because the sponsors participated where they shouldn't have, but sometimes, it doesn't.

SENATOR COGHILL: I agree, and I see a lot of studies like this. So it makes you suspicious, number one. So you look for critical things. And then on page two, you had missing values were imputed. You know, you can think the direction of that implication is going to be.

MR. STUART: No. They tell you. They tell you how they

impute it.

MS. STRITE: Tell us where that is on page two.

SENATOR COGHILL: That's the top of the right-hand, second page.

MR. STUART: Yeah (affirmative). So go ahead and read that.

SENATOR COGHILL: "Missing values were imputed using a mixed effects model."

MR. STUART: All right. Now, most of.....

MS. STRITE: We didn't talk about that.

MR. STUART: We didn't talk about that. We welcome you to talk about that, but that is considered -- well, first of all, a four percent loss in a study this long is very low. As you said, low. Low. And so if it's low, it's got a less likelihood of distorting the results. So that's the first part.

Now imputing the data with a mixed model is, to us, totally acceptable in a situation where the likelihood of not losing people is about zero in a study, although it happens. So we would say yes; we can't disagree with you that that is not perfect, but it's not a lethal threat. That's how I would answer that. How would you react?

MS. STRITE: Well, I'm going to recomb things a little bit because, again, today, our goal is not to teach you critical appraisal because we knew we couldn't do that in the

time we have.

MR. STUART: Although, we're seeing that a lot of people know how to do it.

MS. STRITE: Yeah (affirmative). So -- but I am going to recomb one of the things we talked about, which was Intention-to-Treat analysis.

So I want to go to the top of page 49 on the right side and start with the first complete sentence there and have people first react to that. "All...patients were included in the analysis and analyzed for efficacy in the group to which they were allocated." What is that sentence telling you that we talked about earlier?

COMMISSIONER HARRELL: This is Colonel Harrell. There was no censoring. They studied everyone.

MS. STRITE: They studied everybody. It's an Intention-to-Treat analysis. In Intention-to-Treat, the reason for that title is, if I intended to give you placebo, I analyzed you as I intended to treat you. So if I accidentally got the active agent, but I was supposed to get placebo, I still get counted in the placebo group. So the fact that they're saying that I'm analyzed in the group to which I'm allocated, and this is for efficacy, that's a positive thing. So we like that.

Mike's point about the attrition is -- you know, a low percent of attrition is good.

The two ways of imputing data that I mentioned before are

trying to guesstimate truth and then the setting of the high bar, where I gave the extreme case analysis of missing people in the placebo group would be counted as successes. In the myoceptimab group, missing people would be counted as failures. They didn't choose that method. They chose the guesstimating of truth, and in the guesstimating of truth, mixed effects models are considered to be one of the best ways to do that because the mixed effects models take a robust look at the various patients and their various characteristics.

I'm going to contrast that though with something that we haven't talked about, but that you see a lot because what, most of the time, happens is Last Observation Carried Forward in this, quote, attempt to guesstimate truth. People recall seeing that, LOCF. It gets abbreviated a lot. A lot of researchers use that because it's easy, and what it means is that, if I have a 12-month study and I dropped at month six, they take the last observed value and then just carry forward that, as if that would be my ending value upon completing the study. That's not considered very robust for a lot of reasons. So mixed effects models tend to be favored because they're looking at more data in more complex ways that, again, try to look at variables at patients that have stayed in the trial and then compared them to the patients that are missing. So -- and it is a model though, and models are not truth.

MR. STUART: It's a model.

MS. STRITE: And when you have -- you know, with Intention-to-Treat analysis, again, that is creating a bias, too, but it's an attempt to try to minimize some other problems that result from other kinds of biases.

MR. STUART: Yes?

COMMISSIONER HIPPLER: We were happy with the sample size. In fact, I would have been happy with the sample size a fifth the size used, 5,300 people.

MR. STUART: Okay. Well, (indiscernible - voice lowered) statistically significant results, and it is a fairly large study, and they may have been able to get that with a smaller study. We don't know. I suspect that that's quite possible. It's actually a good point. Other questions, comments, issues? Yes?

MS. BERNER: Barb Berner. I had a question on the -- in the second paragraph under "Methods," I thought there might be some bias in selection where it said something about "optimal medical therapy as determined by their physicians."

MR. STUART: Right.

MS. BERNER: That could have been a bias by the physicians who might have had different determinations.

MR. STUART: Okay. Good. And what kind of bias would you call that? And it's okay to say -- when somebody asks you that, you can say, it's just a bias. And you could also put it into selection, performance, attrition, or assessment. It

doesn’t matter how you do that. I just point that. So if you have further comments about what you think about the type of bias, that’s okay.

The reason I’m saying that is I can’t show you this, but -- because it’s on Sheri’s computer, but I actually put these things into categories, and maybe I can plug in the computer to show you or maybe we can find it.

MS. STRITE: Oh, I probably can.

MR. STUART: Actually, at the very end, I would like to show them how I do it, just so the people who want to know where they fit into the biases. If I don’t use that thing on pages five and 15, maybe, the one on the checklist is on.....

MS. STRITE: Five and 25.

MR. STUART: Twenty-five. The one on 25 and fill out every one of those boxes, I’m likely to miss something. Oh, okay. I thought there was a ghost in the room pulling this mouse around. Thank you. Other comments? Yeah (affirmative)?

MS. HUDSON: I’d like to point out -- Laura Hudson. I’d like to point out some questions with regard to the conclusion on the end.

MR. STUART: Okay.

MS. HUDSON: They stick to the positive statements they could make. They talk about their good outcomes. And then they talk about angina exercise tolerance, but if you look at

those specific paragraphs of detail on the page before, on the right-hand column, fourth paragraph down from the top, angina episodes and exercise tolerance, those addressed a 24-week period, and it doesn't indicate how it progressed or trended for the rest of the 36 months, which is a significant amount of time.

MR. STUART: Yeah (affirmative).

MS. HUDSON: So maybe they're pinpointing one point of time as an effect to draw a positive conclusion to combine with their other 36-month conclusions, and they make no mention at all where they had no significant changes in postural hypertension and symptomatic bradycardia or death.

MR. STUART: Okay. That's a good point. I think the answer from sort of the other side would be yes, but they randomized. And so you would expect things to be similar, since they have a protocol stating when they're going to measure what, and they measure both groups at the same time. It's not like they selected that after the study got started.

Now if they had stopped the study by looking at the data, that would be another issue, but with this one, it's conceivable that that could have happened, but again, I don't see a high risk of bias for that.

MS. HUDSON: It just seems like they're sticking to the positive statements they can make.

MR. STUART: Okay. All right. Yes. Other comments?

Yeah (affirmative)?

DR. BUTLER: Do they overgeneralize, going from the patients who are having three episodes of angina a week to, it looks like, all patients with symptomatic angina?

MR. STUART: Well, that's.....

DR. BUTLER: Because the market just got a lot bigger for this drug.

MS. STRITE: Who wanted stock in our company?

MR. STUART: (Indiscernible - simultaneous speaking), and I can bet he's seen this before. You go from the data to the conclusions, and you don't have -- you have a disconnect.

DR. BUTLER: Put it in the water.

MR. STUART: Yes. Okay. Okay. Yes. And you see that time-after-time. There could be no doubt that this.....

MS. STRITE: Exquisitely designed argument and executed study.

MR. STUART: (Indiscernible - simultaneous speaking) Avastin is a superior drug. There could be no doubt. And then he goes on to explain this, and.....

MS. STRITE: There could be no argument.

MR. STUART: Oh, no argument.

MS. STRITE: Yeah (affirmative).

MR. STUART: Then he goes on to say, the reason for that, and then he gave a reason that was so bogus that I couldn't even remember. It was an editorial written about one of the

worst studies I've ever seen, and this guy gives this a glowing review, and all the oncologists are reading this. And I used to use it as a teaching case when I had oncologists in the room because they get this disconnect between the data and the conclusions that you're pointing out, and it's very common.

MS. PENWELL: This is Vicki Penwell. I have a comment and a question. On the CI and relative risk and absolute risk, all that stuff, I would really love to see that as in this -- in the narrative form, but also as well in a table form.

MS. STUART: I agree. I agree.

MS. STRITE: And what you, as a critical appraiser, you have to look at both because the frequency of them not agreeing is really big.

MR. STUART: I think that's a good point.

MS. PENWELL: And then the other thing, I only realized, going through this in discussion, that I was thinking of safety as relating to the safety of the drug. And so I sort of get through now, when you're talking safety here, you're actually talking about the safety of the administration of the trial; is that right?

MR. STUART: No. No. The adverse events, they list postural hypertension.....

MS. PENWELL: No. I'm talking about this, where it talks

about the mistakes in the administration. Page two, second paragraph.....

MR. STUART: Can you give me the page, the column, and the paragraph?

MS. PENWELL: It's page two, the right-hand column, the second paragraph.

MR. STUART: Oh, the safety population. Okay. The first point is what we said earlier. When you're evaluating safety, it's not like the efficacy where you do an ITT on everybody; that's number one. They have to have received the drug.

MS. STRITE: Let me explain why though. ITT analysis, generally speaking, is very conservative, meaning that it's an analysis type -- again depending upon on how they impute the data, but it's an analysis type that tends to be harder on the drug. And so it's more conservative, forcing a greater likelihood of decreased statistical significance to no statistical significance. If we did that same way of approaching for safety, we run the risk of making the drug look more safe than it is, and we want anyone who has had an exposure to be counted in there. So what this is saying, actually, is a good thing, that they are actually spelling out that they cut the population differently to analyze safety as compared to efficacy. So that's the first key point.

And then the second one is they're letting you know that they had a couple of accidents when they were dispensing the

medication so that we have a little bit of a difference in what we would call migration or what some people refer to as crossover, not to be confused with crossover design which is a different thing, in that we ended up with some people out of their randomized groups. So that's nice reporting. And then they're emphasizing to us that they analyzed them as treated. So that paragraph, as a critical appraiser, is telling me something good that makes me more comfortable about how they assessed safety. Is that helpful?

MS. PENWELL: Yes. And then finally, I don't know what that "alpha was set at" thing is.

MS. STRITE: Yes. What that is -- way back when, when we were talking about statistical significance and pre-specification, some of you will have seen the term "*et priere* (ph)." That's just Latin for pre-specification. That means that I called statistical significance for this study at less than 0.05.

Now some studies, someone may say -- and I'm not sure what drives these choices. I would think you would want to pick this; it gives you a better chance of not having chance effects. But I could say, okay, I'm going to set statistical significance, and another way to say that is, I'm going to set alpha at less than 0.01. So as long as I'm doing it at less than 0.05, that's legit, but once I call that, I can't now claim something to be statistically significant if I set a

lower alpha. So that's a good thing that they're telling you what they did in advance.

MS. EDMISTON: Is that -- real quick. Is that a mistype on alpha? It actually says set at more than.

MS. STRITE: Oh, I want to kill myself. Yes. Thank you. Sheri has a problem with greater than and less than, right and left. North and south I can handle. Up and down I can handle. East and west I'm all confused. And Mike failed here because he's supposed to know that about me and catch me at that, and I have failed at this because I looked at this 17,000 times. Yes. That is supposed to be less than. Thank you, and you should be very proud of yourself because, in all the people that have looked at this, you're the only person who has noticed that. So everybody else has my disease, too. Say what?

UNIDENTIFIED FEMALE: Table 1 is very proud.

MS. STRITE: Yay!

MR. STUART: Yeah (affirmative). There you go. Let's go to you.

DR. THOMAS: Hi, Kim Thomas. Our table just had a comment about the primary endpoint. I know it's common to often group morbidity and mortality, but it's often useful to see death. Even though MI is still a significant outcome, most people would consider death more significant, just to be able to see if the death rate was higher in the intervention

group.

MR. STUART: Good point. Good point. Well-taken. It's a threat. I mean, it's not ideal. It's not an ideal way to do it. Okay. And actually in the cardiology literature, when you look at the composite endpoints, a lot of the cardiology literature is driven by choices to hospitalize for angioplasty, for example. So your point is very well -- there should be a table with the combined and the individual, both for safety and for efficacy, if you're going to do it that way, if you're going to do combined. Okay. And then Tom?

COMMISSIONER HARRELL: Just one more statistical question because this stuff makes my brain hurt. The second page, right-hand column in the "Results" section, the relative risk confidence interval is below one, 0.5 and 0.73. The absolute risk reduction though, the confidence interval is above one, but doesn't contain one. Can you describe what that means?

MR. STUART: Yes. Yes. That's a good question because this is where a lot of people have trouble. The -- when you're talking about relative risk, it's 0.58 or 0.58 of one and that's a ratio. So in that case, statistically significant confidence intervals would -- I'm sorry. In that case, they give the relative risk reduction as a P-value, and it's statistically significant confidence intervals are 0.51 to 0.73, which do not cross one. So.....

MS. STRITE: And let's make sure that everybody knows

that relative risk is shorthand for relative risk ratio.

MR. STUART: Yeah (affirmative), but it does not cross one because it's a ratio. Therefore, it's statistically significant. Okay. That's point one. Are you with that one?

In the next sentence, absolute risk reduction -- by the way, which is a percentage, not a ratio; therefore, the Line of No Difference would be a zero in that case. The absolute risk reduction is four percent, 2.53% to 5.36%. It does not cross zero. So it's statistically significant. Does that make sense to you? Okay. Other comments? Yes?

COMMISSIONER MORGAN: This may not be relevant, but -- oh, my name is Dave Morgan. I'm on the Commission. I'm not a physician, like a lot of people at the table and around. But would the social economic location of the 18 centers have an influence?

MR. STUART: It could.

COMMISSIONER MORGAN: Like, if these 18 centers are in Mississippi that has the heaviest, most obese people probably on the planet.....

MR. STUART: Or is it Oklahoma?

COMMISSIONER MORGAN: I thought it was Mississippi.

MR. STUART: It might be; they fight for that.

COMMISSIONER MORGAN: Well, hey, depends on who had the most chicken-fried fritters that week. Or was the.....

MR. STUART: Or pickles.

COMMISSIONER MORGAN: Or were the 18 centers in Denver, where everybody's job is 28 miles (indiscernible - voice lowered) or something?

MR. STUART: Yeah (affirmative).

COMMISSIONER MORGAN: Do you, in these studies.....

MR. STUART: Yes?

COMMISSIONER MORGAN: As you can see, I don't know a lot about these.

MR. STUART: Actually, you do.....

MS. STRITE: No; you do. Yeah (affirmative).

MR. STUART: .....because it's -- you're critically thinking. And were you here in the morning when we talked about PICOTS? You just talked about setting. PICOTS. The "S" is setting. And if the setting -- when you're taking into consideration all elements of a study, the sophisticated reader is going to do just what you did. He's going to say, let's take a look at this settings here. Does this represent our people? And if it doesn't, because I'm from Mississippi or Tulsa, then it may not be representative of your people. This study may be misleading. That's right.

COMMISSIONER MORGAN: For the record, I'm from Kentucky, not.....

MR. STUART: Okay.

MS. STRITE: We love Kentucky.

COMMISSIONER MORGAN: University of Kentucky. Okay.

Let’s get that straight.

MR. STUART: We’ve been there; we love it.

MS. STRITE: We’ve been there; we love that.

MR. STUART: We love that school.

MS. STRITE: Yeah (affirmative). So I just want to make just a little definitional thing here. There are two kinds of validity. There is internal validity, and then what you’re addressing is called external validity. So what internal validity is, is how true is the study within it’s own context. And so when we do critical appraisal, that’s kind of our first approach at that. And then once you determine whether you think this study is true or not, it then becomes the external validity assessment, whether or not you think that those results are going to apply within your group. And it may be, given that they’re human beings, yes; it will. But it may be that we can’t expect as big a difference between the groups, for example, as many people benefitting. So the external validity assessment, PICOTS is used to help evaluate that and then that’s a huge, huge set of judgments to be made.

MR. STUART: Other comments?

MS. STRITE: Do you want me to get it up for you?

MR. STUART: No. I know where it is. Okay. What I’m going to do now, if we’re done, is pull up.....

(Pause - whispered conversation)

MS. STRITE: I like this. Mike’s got New Folder 2, New

Folder 3, New Folder 4, New Folder 5. I know what disease causes that.

MR. STUART: We'll go to your files, okay? So what I'm going to show you, if I can find it because it's on the wrong computer, is the way I do it because there are a couple things -- you've got almost everything that we wanted you to pick up in this study. You missed -- nobody actually -- did anybody comment on Concealment of Allocation?

UNIDENTIFIED FEMALE: Yeah (affirmative); we did.

MR. STUART: Did you? Okay. I thought it came up with blinding, and I let it go, but for.....

MS. STRITE: Let's have someone point to where that is because I think that's really important.

MR. STUART: Okay. Could you show us where that is?

MS. EDMISTON: It's in the second -- the right side of page one, second paragraph.

MR. STUART: Okay. "Allocation was concealed through use of identical, locked metal containers maintained under the control of the Central Pharmacy Director who was not otherwise involved in the study." What that means is that no one can predict -- this is the formal definition of Concealment of Allocation. No one can predict to which group the next patient will go. And the way they did it here, it would be impossible to predict, unless there was something that I don't know about, that nobody could direct traffic at all. So what

you see here is a way of organizing everything you said, and this is how I summarize what I do. Sheri will take the study, and she'll make notes in the margins, randomization or selection bias or however. I can't remember, exactly, her method, but at the end.....

MS. STRITE: No, not the bias (indiscernible - voice lowered) randomization, RAND, blinding, BLID.

MR. STUART: Yes.

MS. STRITE: So I have little abbreviations for critical appraisal concepts for places where I want to go back and look at, like, blinding all of the (indiscernible - voice lowered). So they might have mentioned something about blinding assessors way over here, how they masked the study medications. By doing these -- not reading the study, but just doing these quick abbreviations, I can now go through and spot check various areas extremely quickly and efficiently. And then if Mike wants to have a conversation with me about it, it's very flexible for me to go to where I need to go to, to discuss my findings.

MR. STUART: So what I wanted to show you was the fact that I can now, at least, show you how I evaluated the study. Now, we don't think there is one way to do this. Sheri copies and pastes the abstract of the study and makes bullet comments in any order that she feels like putting them. That's fine because we can see the problems with the study. I try to

organize them into these different bullet sections. Like, let's describe the population for the study. This is, in a sense -- it's not a formal critical appraisal, but it's just the way I tend to approach it. Then I do, what's the issue in the selection bias arena, and I have three things, generation of sequence, concealment of allocation, and similar baseline characteristics. I don't read the study. I just go find those three things. If they're all there, there is low risk of bias for selection bias. I then move to the second phase of the study, which is the performance phase, and I have given this low risk of bias because of these points here, and we've covered all these. I then go to attrition bias. I've given that low risk of bias. I stick my neck out on every study, and Sheri and I will usually go over them, and almost invariably, we will agree, and if we don't, somebody will convince somebody else, and I don't know where we've ever come to a situation where we can't agree, have we? I don't think so.

So I do attrition bias. Then I do assessment bias, and I say low risk of bias because they did an ITT analysis, and they did a conservative imputation of missing data, but that's my judgment because I kind of think mixed models are a reasonable way to impute data. I mean, it's been studied, and it's not a really bad way. That's for sure. Then safety, this is where it gets marked down. Uncertain risk of bias

because they assessed only those receiving the treatment. That's good. They had similar adversity and that's good, but the power may have been inadequate to show a difference in safety, if there was a difference. Even though it was a long study, it may not have been powered to find the differences.

We did a systematic review on biologic DMARDs for five drugs, and even with meta-analysis and systematic reviews, we did not feel we had enough information to know the difference between the comparative safety of these drugs, which is what we were looking at. So we looked at registries. We looked at cohort studies. We looked at case control studies, and we tried to fit it all into a conclusion that made sense to us. But you see, I almost never have found a study where I could be low risk of bias for safety, and I said that, like, seven times today. So be careful about the safety and be cautious with how you word the safety issues in saying there were no significant safety issues. Well, you didn't report any, but you don't know. So be really watchful for that.

And then I have this category "Other," and I say the study size is sufficient. The secondary outcomes support the primary outcomes. The primary outcome and effect size are clinically meaningful. That's my critical appraisal. It's not long. It's very thorough; at least, in my opinion, it is, and it's adequate. And I can now be transparent in telling people what I think about this study, rather than walk into a

P&T meeting and say, well, there is good evidence that this drug is effective, which you will hear and which we've heard from so many P&T Committee members in so many different organizations, and they can't show us. And we think that you need something to summarize -- we don't care how you do it -- something that summarizes the biases.

We worked with one of the largest group's in the country P&T Committee, and when we went to the first meeting to audit it, it was a very interesting experience because it was a giant room in a hotel. There was kind of a gladiators' arena at the bottom. Then there were specialists all around the first ring. Then there were primary care docs hiding up in the secondary ring. And then the staffers were off at another table, and they would sneak up and present their data and creep away. And then everybody in the committee would state their opinions. And I was saying, well, what are they doing with the evidence? And they were actually ignoring the evidence, to a certain degree, in certain cases and going with the expert opinions. At Group Health, as Ward knows, we.....

MS. STRITE: Couched his evidence.

MR. STUART: Yes. We separated it by having two committees, eventually, in technology assessment where the first committee did only this. The second committee were the medical directors. I went to those meetings, too. We tried to get some process started in there, but it ended up always

being a discussion, and well, let's go ahead and do this, and they would make better decisions because of the first committee. But still, they would pass drugs that a purist evidologist would not, and they were based on other considerations, which we're going to get to next. But the point being, we think that P&T Committees in any quality improvement need to do the evidence, lay it out this way for the outcome, so that, eventually, it gets packaged in a way that everybody gets to see the reliability and the clinical value of whatever we're talking about.

MS. STRITE: That's a nice segue.

MR. STUART: Okay. So we're now going to do what, Sheri?

(Pause - whispered conversation)

MR. STUART: Thanks. There are a lot of good things to say about specialty societies. They certainly create a culture, and they create a forum for discussion. They do a lot of good things, but really when it comes to the evidence, sometimes, they don't do so well, and they're -- that doesn't mean they can't in the future, but right now, we've seen evidence that some of the stuff coming out of the specialty societies, like guidelines, are somewhat problematic, and I mentioned this earlier.

MS. STRITE: And we're -- sorry. We're at the second section in the back, slide 139.

MR. STUART: The problem with clinical guidelines is that

they're combos of opinion, evidence. There is a lot of pecking orderism. Some guideline groups don't even include evidologists. They just include specialists. We don't think that's good. That's evolving. It's getting better.

Guidelines are almost -- as David Eddy said, they're notorious for never projecting (indiscernible - voice lowered) outcomes, if you followed them, or projecting clinical outcomes, if you followed them, or projecting anything, if you followed them. They're just recommendations, and in his opinion, the homework hasn't been done for most of them until you do projections about what would actually happen in our population.

And so he created, 20 years ago, a thing called a balance sheet, but he called it a clinical balance sheet. It did have the regular cost -- the economic items, but it also had the clinical outcomes, and we did that for all our guidelines because he taught us how to do that because we didn't want to be spending millions of dollars on PROPRIO suction therapy and physical therapy when we had no appointments that would have been available and no evidence to show that it really helped people. Actually, we had some evidence, not much. But the idea of projecting outcomes allows you know what you're getting for your money, and we think that there is no need, frankly, to do cost-effectiveness analyses or even use that literature, in some cases, because the literature is so flawed

in the cost-effectiveness world. They don't do the effectiveness part very well. They just go right to the results and start plugging in assumptions and make trees and all kinds of models and things, but really without good evidence, you don't know where you -- you don't know how to make the projections.

So many guidelines are also outdated, and there are people that come to these meetings who are going to have biases because we all have biases, and we want to be able to see both a rating of the strength of the evidence and a strength of the rating of the recommendation, which is what we really need. Then we want to be able to go back and look at the studies upon which they built those ratings.

This is just one example of a clinical practice guideline study that came out -- let's see. This was pretty recent.

MS. STRITE: 2012.

MR. STUART: Last year. So 114 guidelines the IOM looked at, and the median number of standards satisfied was eight of 18. So they were not too happy with that because it's no different from a decade before. That's what it was when they looked at it ten years earlier. They did not include information scientists, as I've mentioned before, frequently, and almost never included the patient perspective. Although, I have -- I'm not so sure about that one because I don't know how to include the patient perspective. I've seen many

attempts at it, but when you have a patient on a committee, you have a patient on a committee. Now that patient is representing all patients; it's kind of tough.

The methodological standards from the societies, in their evaluation, were not high. The study that they did showed that this change -- staying static. They're not rapidly improving. They're evidence-based guidelines. So there is some work to be done there.

This is just another one from *Brito* that evaluates a systematic review supporting the practice guidelines that, essentially, says the same thing. I don't want to read the slide to you. It's just caution when you're looking at guidelines because there are enough problems looking at primary studies. It's more difficult to look at secondary studies because now you've got to look at how they do heterogeneity and some other things. And then it gets even more complex when you look at the guidelines because you've got opinion. You've got lots of recommendations, lots of interventions. So it's a difficult area. Now -- yes, comments?

MR. CAPISTRANT: Todd Capistrant. *U.S. Preventive Services Task Force*, the big book people get, how do those pan out as far as credibility because it drives a huge amount of resource (indiscernible - simultaneous speaking)?

MR. STUART: They're considered to be transparent. So

you can go look at how they did it. That's point number one.

You can make your own judgment about how they did, but in general, the feedback has been that they do high quality work. That's sort of out there in the atmosphere, but there's no gold standard. All you can have is transparency and see if they did what you think they should have done or whether they didn't. And so I would say they've done a better job on some than others, but they do, in general, a really good job.

MR. CAPISTRANT: Yeah (affirmative), because the specific example is the recent change on the CT scan recommendation for smokers with 30-pack year smoking history. That's going to drive up costs and utilization.

MR. STUART: Yes. Yes. I'm not on top of all that literature. I looked at it a couple years ago, and I was very nonplused by the literature. Now, I don't know about now. Has anybody critically appraised that literature because I don't feel comfortable commenting on it right now? But yes; that's an example of where the 300-pound gorilla is going to change costs.

MR. CAPISTRANT: Yeah (affirmative). They just released the recommendation to scan, so.....

MR. STUART: Okay. Thank you. I (indiscernible - voice lowered).

MS. EDMISTON: This is Gena. I have questions on the science that guides what most of us would refer to as core

measures out of CMS, et cetera, which are really, you know, process measures.....

MR. STUART: Right.

MS. EDMISTON: .....and we're all very much guided by them.

MR. STUART: Yes. Yes. Well, you notice we haven't said a lot about regulators. That's not our -- that's not where we live. We live with -- well, let's.....

MS. STRITE: But we do know something about the CMS.

MR. STUART: We've worked with the CMS, and we tried to find a way to work with them where we could use semaphores, red danger performance measure, yellow being somewhat cautious, green, good one, and.....

MS. STRITE: But we also had -- what? We evaluated -- I don't know -- ultimately, like, 70 measures. I think this was for SOW 2011 or something like that. Seventy measures. They had no time to give us to do it. And so I had to go present to them, which was really scary, and the guy said to me -- the physician in charge said, look, I developed these last year, and I know nothing I did was evidence-based. Anything that you do is going to be a step up from what we did.

So the bottom line is that you don't have a lot of evidence-based measures going out there, unless something really radically has changed since we were involved in working with them, which I suspect has not happened.

MR. STUART: I think it's going to evolve. Things will get better, but there are people that point out that we all have a ways to go, all of us, and CMS certainly does, and everybody does. All of the electronic textbooks up-to-date. Everybody has a ways to go. So it's not like there is somebody who is perfect out there. It's just that we're all in this -- in a large spectrum, we're all kind of bunched together or we're not where we want to be and that includes CMS.

Performance measures that penalize doctors and reward doctors, that's probably -- you know, that's really controversial that that's valid. People try to adjust so that it becomes something that's useful, which I believe it is, and I believe that we have to pay attention to regulators. So I'm not saying anything, except that, if a regulator has made a ruling, at Group Health, we would, at least -- if we thought they were wrong, we would try to talk to them and say you've got to change this, and sometimes, they will do that, too. So it's not where we want it to be.

MS. STRITE: And not in your binders, but on our website or you could write me, we've got a one-pager on performance measures and then we have a really big, very comprehensive that we created that gets at things about evaluating performance measures. Very different from anything that we had ever seen. We did a three-day program for AAFP, looking

for trained family care docs to sit on specialty teams and so that work emerged from that, and it's quite comprehensive, and we're happy to share that and spread that around.

MR. STUART: Okay. Secondary Studies and Secondary Sources. I don't want to read this list. It's in your book. We use different sources for different things. We're -- unfortunately, we're unable to tell you, if you go to this source, you're going to get great stuff, but these are some that we like, and rather than take time to go over these, I'll just say they're in your books.

MS. STRITE: Yeah (affirmative), but I'll mention that you have it as a slide, but then you also have the list in a more, probably, easy-to-read fashion on page 27.

MR. STUART: Okay. The other thing is that we have given you the critical core competencies for primary studies. We also have tools for secondary studies and secondary sources. And so again, they're not in your book. Or are they?

MS. STRITE: No.

MR. STUART: No.

MS. STRITE: So those, you can.....

MR. STUART: You can get from our website.

MS. STRITE: Yeah (affirmative) or write me.

MR. STUART: And remember, secondary studies are studies of studies, so the roundups of primary studies. Secondary sources are any that use primary studies or secondary studies

to create something like a compendium or some other directive work, and we have tools for those.

MS. STRITE: Yeah (affirmative). And just as a reminder, so again, to contact us, our website is listed in your materials, and there is a little button to contact us. You can email us. Also too, I sent around the sign-up sheet, if people want the sort of four-times-a-year update. So I just want to make sure that that's circulating and not stuck someplace, and ultimately, gets back to me.

MR. STUART: So again, these are online, if you want them, and if you're doing an appraisal, this might help you evaluate a guideline. Questions about what we've covered all day so far? I'll just stop for a second and see if there are comments or questions.

MR. MONAGLE: Mike Monagle. Just one of the things you alluded to, but you hadn't talked a lot about was what policymakers or regulators need to know about evidence-based medicine.

MR. STUART: Great. Thank you because that's where we are.

MR. MONAGLE: Okay.

MR. STUART: And that's exactly what we want to do the rest of the day is really just a discussion. We'll put some prompts up for the discussion, but that's what we want to talk about now.

So what do need for an evidence-based approach? You need leadership. You need an evidence-based approach, which includes the right work components, meaning you've got to resource it. You've got to provide structures, methods, processes, tools. You've got to coach individuals. You've got to get somebody trained so they can train the trainers and then the trainers train others. And you've got to get a culture where people actually use evidence-based approaches to evaluating science. That culture is really important, and it can easily come and go. We've seen it in organizations go up. People leave. It drops back. New people come in. Somebody is really good; it goes back up. It goes back down. So you have to have a culture that supports this, which requires leadership and the resources and structures that people need to do their work, so they get trained and can do their role.

Sheri, earlier today, said that implementation is all local. What she meant by that is that we could have a conference webinar with people all over the world, and we would pretty much not have too much difficulty agreeing maybe on the quality of the evidence, but getting it implemented in our region is going to be very different than in somebody's region, depending upon what our resources are and what our skills are and what we can do in our setting.

The first step in finding out where an organization is, or any group is, is to assess where they are, and we have a

tool called a Health Care Quality System Assessment Tool and that's the starting place to see where you are. Whether it's your P&T Committee, your Commission, whatever it is, your hospital, it's assessing, do people have, within the organization, some skills and understanding of what we've been talking about all day?

MS. STRITE: And that we did include in your book, by the way, on page 28, and Mike's going to, right now, just detail the contents of that. So just be aware you have it and now he's going to review it at a high level, probably.

MR. STUART: Right. So you've got to have some appreciation within the organization. Number two, you have to have the systematic process for doing what we talked about, doing critical appraisal. Number three, you have to have people who understand the strengths and weaknesses and appropriate application of study types. You also need people who are able to quickly critically appraise something and they can be called upon; could you take a look at this study for me? Could you take a look at this guideline for me? You need to have these core competencies within your world, whatever that means. Yeah (affirmative)?

MR. KNAPP: Just to be devil's advocate, why do I need all those things in my organization? Why can't I simply subscribe to some outside provider that reviews the state of the literature and uses these techniques you've been talking

about all day to come up with it?

MR. STUART: I think, if you can find one, an organization that can actually meet your needs, then you audit them or someone in your group audits them, and you say we agree, they are -- they can't be beat, and they are able to get what you need done or they already have it on the shelf, great; do it. But even with that, you have to have some skills because things are going to come up within your group that are not going to be in the published literature, and you're going to have to do some homework with your own databases, your own MEDLINE searches, et cetera. I think you need to have the skills within your group.

MS. STRITE: And you also need the organizational understanding because, without that, you know, again, the leadership, you're going to end up with what Mike described as, if you've got an individual that's in leadership that's been aware of these things, if they leave and you haven't set up an appropriate structure that supports this, it's all going to fall down, and it's going to erode. And you're going to end up with opinion cast as evidence when, in fact, it isn't, people that are unaware of how to evaluate the difference. So it's -- in my opinion, if you're going to do quality patient care, you need organizations that are attending to that as much as they're attending to process steps, for example. And I think this is equally, if not more, important, depending on

the question.

MR. STUART: So that's the science. The application of the science, if.....

MS. STRITE: Actually, I want to ask Ward; do you want to add to that?

MR. STUART: Yeah (affirmative). Let's get comments on -- this is the first part -- the way I framed this is, what does it take to have an evidence-based organization or an evidence-based approach? What we started with was the scientific evaluation capabilities? You've got to have something there. Anybody want to add or other comments? That was a good comment about getting an outsource to do it for you, if you can, but we don't think that's adequate. It's what a lot of people do though.

CHAIR WARD: My experience, being both in the public and the private sector, on that is you do need, within your own organization, an understanding of and appreciation for the concepts of that. And maybe before we finish today -- and that may have been one of the things that has gone with the dire budget situations that Washington and Oregon got into, but if you could comment on your all's relationship where, on an annual basis, you used to go in and talk with folks in both those states for their state employees' coverage, Medicaid questions or whatever, how that worked out.

So I think that, within the organization, you need an

understanding of that, and the bigger you are, obviously, the more you can build that expertise, but I would say both in the public and the private sector. Then if you have a national organization, you may have some more expertise at the corporate level. For states, for example, on the Medicaid program -- Alaska is a part of the Med Group out of Oregon Health Sciences University that -- I don't remember -- 16 or 17 states or so collaborate in. But there still are going to be things that come up.

And then beyond that, getting to the point that you made, I think that you don't have to invent the wheel all the time, and you can look to outside entities. Like Sheri mentioned, the Cochrane Collaborative, I think they're a good source of information. They're well-respected. Certainly, a lot of the cookbook kind of standards, but they're so common in the industry, using InterQual standards or Milliman standards related to levels of care, related to pre-certification processes, related to concurrent review of inpatients are evidence-based. And so you don't have to do everything yourself, but you need to understand the process, and there are always going to be issues. We're all different, as people, so we've got seven-plus billion different people in the world, and we're all different, and you need to come up with some expertise yourself.

MR. STUART: Right.

MS. STRITE: And when it comes to things like outside sources, we always audit them, and some people may choose more of an audit light than we might, but that's been very important and helpful to us because we have had some surprises, such as when we were first trusting Cochrane completely and then started auditing and then started finding, you know, variation. So I think that you need people that are able to take a look at what's out there as sources and evaluate how solid that's going to be and how that might be best utilized, too.

MR. STUART: As far as applying it, there are implementation skills. There is a need for access to quality information -- MEDLINE to study -- so that staffers can actually get information and evaluate it and then put into recommendations and implement it, but we have a whole checklist I'll show you for -- or maybe we didn't include that -- for implementation options, and it's just a checklist to say, would this work for us, would this work for us, and you might pick a third of the things because you can only do a third of the things. But you should have some checklist for implementing change, once you've got a recommendation for change or you're going to make a change. And then you've got to get into information for patients and that's really -- we have a long ways to go in terms of that, and we have the incredibly big problem of performance measures and the fact

that we're regulated and we have to do them, and sometimes, they're the wrong thing and then we have to be responsible and give feedback that they're the wrong things, which is what we should be doing, but we have to play ball, too.

Then as part three, the organizational commitment, we really think it works better if an organization has something in their mission statement about evidence-based clinical improvement. We've not seen too many groups do a good job without the leaders being encouraging, supportive, and working very hard to get the resources, where quality gets cut first. When there is budget cuts, guess what goes first? Quality improvement.

So there has to be some sort of systematic understanding and systematic approach to quality improvement within any group. And so there needs to be some skills and competencies within that group, and I think there should be some expectations because it's morally right to do the right thing for patients. It's pretty simple. But to do the work, it's not always simple, but the idea of doing the right thing for patients by including the evidence, by not forgetting or not doing the evidence, that's not okay. And it needs to be nurtured and taken care of, and people need to be supported in their jobs doing this.

So we have a ten-phase -- ten steps for what we do. This is kind of what we did with Kaiser. We assessed whether we

could actually pull off, this PTE Prophylaxis guideline. (Indiscernible - voice lowered) it's going to be quite easy to pull this off with these wonderful people, and the support we got from leaders was just incredible. And we had all these things done, and we got it all done, and we kind of left before all the measurement was done, but they had a measurement plan, at least, but our time was up. And I think -- we don't care whether you use these ten steps or any other steps, but it's just including consideration of the things we've been talking about. Are there other considerations that people think we might have missed here? Yeah (affirmative)?

MR. KNAPP: I also apologize for talking so much. Gunnar Knapp. I think that, when you give sort of lists of sort of who needs to do what, you think, well, who are you giving your advice to, and I think that we need to think of actually sort of three different groups of people the way -- or different groups, how they need to think about evidence-based medicine.

One is medical providers, people that are in the business, ranging from hospitals to private practitioners. You know, how can you incorporate this in what you're delivering? And I think that's been the main focus of your talk today.

You haven't really talked very much about how patients do this in -- you know, on their end. And so you know, one point you've made is that the providers should provide the evidence

to the consumers or patients so that they are able to add this -- you know, have the information they need, but ideally, we would think of sort of what level of, you know, evidence-based approach to their own healthcare can consumers get and sort of how can you educate people into, if the doctor -- you know, if the doctor says this, you know, be sure ask about, at least, this, this, and this. You don't want the patients to go home and try to -- you know, you can't get every patient to get your course, much less understand it, but there are a few basics. You know, how do you evaluate those ads you see on TV or whatever? So that's a whole question for consumers and gets into things, like all the information you find on the Internet and how do you evaluate that, which is sort of a whole bit of secondary information of.....

MR. STUART: Yeah, yeah (affirmative).

MR. KNAPP: And then there is a third question, which is kind of what the Health Care Commission is about, as I understand it, which is, what's the role of the state in this? What should the state be doing in this area? And I'm completely new to this, but I read over the pink page with the Health Care Commission's Finding and Recommendations, and they seem, to me, based on everything I've heard you say, to be very, very sensible recommendations, based on -- in line with the kind of reasoning you're talking about.

I would only suggest one more thing, which is that I

think that government -- one thing government can do is help to both require and sponsor further and better research, you know, sort of do not just the original trial when these, you know, drugs and treatments are originally tested, but you know, longer term trials. Are they mandated? Are they reported? Are recording systems set up that sort of mandate the ability to improve our understanding over time? And you know, to what extent can that be done, even within Alaska or is that something that has to occur nationally would be, I think, the one additional thing that I don't see in these recommendations, most of which seemed spot-on for addressing the kinds of the things you're talking about in terms of what the state could do.

MR. STUART: Great. Good comments. Other comments or comments on that comment? Okay. The staffers need just what we've already said. They need to be supported, and they need a culture where they can thrive and do the work, but they also need to support the state. They need to support hospital systems. They need to support managed care groups. They need to support what we call messaging to basic patients and that's where we're going to go shortly, which is somebody should do a really good job, and we think Healthwise does a fairly good job of summarizing information so patients can make decisions. But we've not seen it to the point where we can actually recognize groups that really do the quantitative aspect about

how much benefit, how much harm, and what are all the options that should be considered by the patient.

So all those things you mentioned are true, and there are a lot of different target groups, and you just have to take each target group and look at what the requirements should be for that target group, and I think it's all based on the same principles as far as our work goes, which is, is it valid and is it clinically useful. You have the clicker.

MS. STRITE: I have the clicker. Yes? And there is a.....

MS. PENWELL: I'm not sure if this is the right time to kind of bring this up, but what I really have found missing from this entire discussion is the role that's -- I see this as being a really important science-based part of understanding your own treatment, and you know, being a quantifiable thing that you can look at when you're trying to make the decisions about your treatment. However, I feel like it misses a huge part of what makes up a patient. You know, patients are spiritual. They are influenced by their culture, by their traditions, and those, often, are not quantifiable things.

MR. STUART: No.

MS. PENWELL: And so I have a huge concern, when we're talking about policy things, that we don't allow evidence-based medicine to become the where-all/end-all of how doctors

and patients make decisions and that we don't allow it to morph into a synonym with cost-effectiveness because a traditional or cultural approach to medicine, you know, some of those things you can't put a price tag on, as far as patient well-being goes.

So while I strongly -- you know, I'm very disturbed in, you know, some of the current discussions we have in our society about our distrust of science and about our distrust of scientific fact, I completely support the idea of evidence-based medicine and the quality of the studies and stuff you all are talking about, but I guess I would just really like to see a piece of this that remembers that we're a whole system. You know, we're not just numbers and percentages. So thank you.

MR. STUART: Right. Right.

MS. STRITE: Well, let me do this.

MR. STUART: Yeah (affirmative).

MS. STRITE: I'm going to skip ahead.

MR. STUART: We're going to skip ahead because of your comment, I think.

MS. STRITE: Yeah (affirmative). Then we'll come back to this because one of the things that I want to do is spend a little time to help leaders understand where failures can happen, if they're not paying attention to evidence in their culture in the right way. So we're going to leap ahead though

a little bit to go to some patient-centered stuff, and what I want to do is I want to remind people what we started out the day with, and we said that evidence is just one piece of patient decision making or payer decision making or provider decision making. The thing is though that it's been so neglected. There are so many problems with it. That needs to get straightened and polished first and then you throw it over into wherever your target group situation is or whatever, and you end up having to make decisions -- again, different target groups in different ways for different purposes -- but when it comes to patients, we have what we call patient requirements, and patient requirements are things like you're talking about. What are my values and preferences? What are my needs? What are my limitations? What resources do I have or not? What value judgments do I make?

And so one of the things that ends up being a big problem is exactly what Mike was describing when he was talking about our friend and colleague Tim's mother who just got told this is how we treat this cancer. There will be some side effects, and we'll start you on Monday. She had no opportunity to understand anything about the evidence and was not invited to participate or even take control of a decision that should have been hers, based on her values and preferences. And I think Mike said something about the bottom line; I was probably multi-tasking at the time. But she chose not to take

the therapy, and she knew she was going to die. She was probably going to die anyway in a much worse death and that was her preference.

MR. STUART: Well, she opted for symptomatic relief, period.

MS. STRITE: Right, and to spend as much quality time with her family as possible. So Mike's going to -- but then we'll go back, if we have time, to cover some the other places we were headed, but we'll do patient-centered decision making at this point.

MR. STUART: Okay. So we've already done this slide, the five things that mean something to patients. We've already done this slide, The 8 Key Questions Patients Want Answered. What do I have? Why? How did I get it? What could it do to me? What do you know and don't know about it? What choices to I have? What's the evidence for each choice? What's the cost to me? What's the cost to society? In some cases, if a patient wants to know that, they should have that information. What do you think, as my physician or other healthcare provider? And then, how do I get it done? So those are the eight questions that we -- that's our set of what we think patients want to know so that they could be at the center of the care.

MS. STRITE: Do you want to trade microphones with me? That one's going in and out so much.

MR. STUART: Yeah (affirmative). You know what I'm going to do? I'm just going to change positions yet again.

MS. STRITE: No. No. Take mine.

MR. STUART: All right.

(Pause - switches microphone)

MR. STUART: So again, to be respectful, you need to give people the right information. You need to treat them respectfully. But our arena is the information, but we do see these other things that go beyond the evidence. We don't stop with the evidence.

And I'll go back here a couple slides, but these are the things patients deserve. They need to know their prognosis. They need to know their options. They need all these things, and they need to be able to be comfortable in asking questions without being, in a sense, sort of -- I don't know -- "there-there." They need the information. They don't need the "let me worry about that for you." They don't need that. It's good to be supportive, but not directive, if the patient doesn't want you to direct. So they should be able to get individualized care that meets their needs.

And so we came up with -- we were sitting around in my office with some Kaiser folks, and this is a model that we started thinking about years ago. It goes something like this.

The first thing you do is you get your question that you

want answered and you review the evidence. You also need to be prepared to create information decision aids, any kind of aids for patients to be helped in understanding the prognosis, the benefits, risks, alternatives, costs, uncertainties, but then that's at the heart of this, but in fact, at times, there is no reliable evidence.

And when you fall into this column, you have to information the patients again about their choices, but you just tell them we don't have good evidence, so we're going to have to make a decision based other factors and that may be that you choose what you want, period, because we don't have evidence. We've got my experience. If you want to hear my experience, great, but that's not evidence. Well, people could argue it's a type of evidence, but it's not critically appraised literature. But the patient may say, you know, I would just like to talk this over and understand more, and can you answer these questions, and they work in a partnership, and they don't just have a patient make a decision without any help. But then there is the situation, if somebody has a cardiac arrest, I'm not going to say, what's your preference? I think that, in that case, there is a recommendation that's very strong that we, unless you've told us not to, would rather have you alive.

So that's where the prescription of information comes in and the support through information aids and reassuring

everybody that everybody understands the issue. So checking for understanding is a big piece of this and then a decision can be made, but we have gone further than this.

The University of Washington faculty that teach at the ACPE courses -- that's American College of Physician Executives -- use our model called the Glide Model, which is greet, listen, inquire and exchange information, determine and decide, and effectively end the conversation so that it's a happy visit. That's the Glide Model, and I didn't really -- because of time, I didn't really get into all the details of the I, D, and E, but just for an example, the greeting. People should be cordial. They should be prepared to give the patient full attention. They should be -- the literature is very clear that warmth, empathy, and respect are really the absolute requirements for any contact with the patient.

The listening has to do with what's been written about as the patient story. Frequently, it's not about a science issue. It's about an emotion. It's about a fear. It's about loneliness. It might be, you know, a chronic disease. I'm losing my voice; pardon me. It may be something that you're only going to get by listening to a story.

So you start out listening to the story. Then you can come back and do the biomedical questioning about what the medical diagnosis is, but don't forget the patient's story because it may be that they just want to be reassured. They

don’t really want you to give them a drug, but this has been studied, and machine gun questioning a patient, when did it start, did it run down your leg -- you know, there is this anecdote that we ran into in the communication literature years ago where a resident, who is really quite green, is seeing a patient, and the patient has back pain. And the patient comes in, and the doctor says, okay, nice to see, Ms. Smith. Can you tell me a little bit about what’s bothering you today? The patient says, well, I have this terrible pain, and it’s running down my right leg. Or, no. I have this terrible pain, and the doctor says, where is it? Well, it’s in my low back. When did it start? Well, it started the day my husband left me. And then the resident says, does it go down your leg? And the idea here is that he totally missed the patient’s story. He just went right to the biomedical science issues and forgot about the patient because he’s the customer. He is the resident trying to learn medicine, and this is what happens is that, at times, the story gets lost.

So we want the story told to keep it patient-centered, and we want to say, is there anything else? Is there more to this? Did I miss anything? And most doctors are never taught how to respond to people’s emotional reactions to things. So all you have to do is say, I’ll help you; we’re in this together. You don’t have to own their emotions, but many healthcare professionals are never trained how to deal with

emotions so they just ignore the emotions, and the patient is going I just told you something. My husband just left me, and the doctor could, at least, say, oh, I understand how that could be kind of tough, just a supportive statement.

So the listening is not just biomedical listening. It's the story. It's the emotional story, and it's the information and the medical stuff that comes after that, once you know where the patient is in their own context. Anything you want to add to the Glide Model there? I think you came up with this model.

MS. STRITE: No; it was a team effort.

MR. STUART: Team. So we're going to go live to an example of what you can do for patients. Let's see what we've got. Let's go to hormone replacement, maybe.

MS. STRITE: So what this is, is this is our website, one of our pages.

MR. STUART: Why don't you do this?

MS. STRITE: Well, I have to have a clinician participate.

MR. STUART: Okay. I'll participate. I'll pretend.

MS. STRITE: I've got a clinical component here.

MR. STUART: I will.

MS. STRITE: Yeah (affirmative), but I will try to do your voice as much as possible. So this is on our site, and it's a page called "Same Page" because we're really good at

naming things. The concept was that we wanted something that really expressed that we want to have information that both is useful to clinicians and is useful to patients.

And we actually had a beautiful experience when we were in Tulsa, and several times there, we've taught physician-patient communications -- and I should say clinician-patient communications.....

MR. STUART: Right.

MS. STRITE: .....because there weren't just doctors there. And unfortunately, this is a funny width, but I think it's not going to matter. But the thing that was wonderful is we had some of these messaging scripts that we're going to show you, and what the messaging scripts are all about -- we had a consultant pharmacy group say that they were helping with long-term care issues, and the only way they can communicate with the physicians, the attending physicians for the patients was through the medical record, and they said, we can see them making some drug choices that we want to give them different information on, and we want it to be evidence-based. How could we do that?

And so we created this sort of idea of a script that gives a little evidence-based information and then some quantified information and then provides, you know, transparency in terms of what are the resources. And what was wonderful in Tulsa when we did our physician-patient

communications was we provided some of these we provided some of these to the participants, and then as an exercise, we asked them to use that information and how would they discuss these things with the patients. We'll show you an example. But anyway, the one that was the best and the most popular, I think, actually was a nurse practitioner's.

MR. STUART: She's a nurse practitioner.

MS. STRITE: She was unaware of the evidence. She was able to take the messaging script and then she modeled sitting down with the patient after going through the steps that Mike enumerated, to get them to the, what do we want to discuss in terms of a potential treatment, and said, now, let us review this together and then review the evidence.

And so what Mike is going to show you is what a couple of these look like, and there are very, very concise ways to, sometimes, pack a lot of evidence-based information in, in a very quantitative way, in an attempt to try to do it in a way that's organized, such that both clinicians and patients can get an eyeful of the information and then make some choices. And I think hormone replacement therapy is one of our favorite ones to show.

MR. STUART: Well, they could be laid out anyway you want. On our website, there are templates that can be modified, but essentially, you have to look at what the condition is and how you want to set it up, but it's almost

invariably, for us, text and tables. We don't have the capability of doing, you know, high quality graphic displays and things like that. So we, pretty much, use text and tables.

And you could say, for hormone replacement therapy, the best available, valid, and useful evidence indicates there are benefits and risks of estrogen progestin compared to placebo within the treatment period below. Out of 400 women in a three-year period, 133 of them will be able to get rid of the hot flashes, and we just go right down the line with the important outcomes to patients -- hip fracture, any fracture, colon cancer, stroke, heart disease -- and it gives them quantitative idea, out of 400 patients, this is how many will expect this, and we have a plus for good, a negative for not-so-good, and people are going to have different choices depending who they are and how they value the different outcomes. So that's just one example.

I'll give you another example. Down below, we may have the note prescribing information. You can go to PDR. You can go to drugstore.com or other sources, and we happen to like MedlinePlus because the government pays for this, and it's available to anybody about various drugs or doses, et cetera, or side effects. We say -- we may do -- depending upon what the issue is, we may do it different, but this is just one I'll show you. And then we have the evidence here.

Here's the next one. I'll just do this one, too. This, you'll see, is quite different. This is for sciatica. We, first of all, start with the grade of the evidence, meaning we're uncertain about all this, but in most cases, your sciatica is gone with doing nothing in four months, and 80% to 90% will be better at six months, and almost everybody by a year, but not everybody. That's surgical treatment. Sorry.

Non-surgical treatment, the rates are lower, but at a year, at -- before the end of a year, but at a year, it's getting close to -- this may be a little high; it may not. We don't know, but it may be about the same results without surgery. So people have to know that, but there are some risks of surgery that I don't believe most surgical patients who have back surgery get this quantitative information. They don't know that one in 100 is going to get a pulmonary embolus. They don't know any of this. They don't even have a clue that they may tear the dural sac and get a really bad headache or be in the hospital, require a blood patch, blah-blah-blah. And so these are the kinds of things that we like to lay out for the patients. And again, the references.

I could show you many of these that are different, but you get the idea that the evidence should go into something where that nurse practitioner could sit down with a patient, show her this, have a discussion with her, use her preferences, her values to make a decision.

MS. STRITE: So to put kind of a summary capping on this, we think that, when it comes to what you do with patients, it's a matter of information. It's engagement. And then it's a model that allows a patient to make decisions even about how they want to make decisions.

And so we start with the information and that's the looking at the science and the piece that's so missed and then putting that into usable information, decision support, information aids, and then having, ideally, clinicians who are trained to engage with patients and then help facilitate their decision. So I hope that was helpful.

Because of time and because we want to make sure people have a chance to discuss more, what I'm going to do is I'm going to leap to slide 156, and I'm going to -- except that I'm not in the slides.

So we're going to go backwards here to 156, as soon as I can figure it out, and I am, basically, going to review slides, and I'm going to talk really, really fast, just to get a few concepts out there, mainly so you know what you have in your book in the rest of these slides and what the meaning is and then we can have a broader discussion.

So what this particular segment is addressing, it's addressing leaders, and what I want to highlight for leaders are what it is that they need to do to help staffers and teams be successful. We've worked with a lot of staffers and so

we've seen some things, and we've got some tips that I think are important to be aware of.

And so the subheading of this is, "How teams need you, as leaders, to foster a culture of attention to evidence, if you are going to promote an evidence-based practice in your organization." And my first bullet in here is, "Foster a Culture of Attention to Evidence." That is not duplicative. That is actually the first step in trying to do this. So it's kind of a topology. The second one is, "Support Attention to Evidence." And then the third one that's extremely important is, "Be Realistic About the Evidence."

What these break down to is that leaders need to understand that, to really be a steward to evidence-based practice, you need to be able to understand the importance of the use of evidence in approaches to providing best patient care, and they need to ensure that the use of evidence becomes a part of the culture. And then Mike showed you our ten phases of clinical improvement because implementation tends to be local.

In terms of support of attention to the evidence, this comes down, in a large way, to understanding that, to have an evidence-based culture, you are going to have to spend some money to support it, and the work is labor-intensive, and it takes a lot of time and a lot of time to do it well. And if you are going to walk your talk, you've got to put resources

behind it. The resources aren't just in terms of money, but you need to have structures, methods, processes, and tools.

Mike and Ward -- he had modeled for you one of the structures that they did at Group Health that was very successful. And then the need for roles and skills, and this is going to be very important.

When it comes to the roles and skills, I'm going to show you a slide, in a few minutes, about something more about that, but before I talk about that, I'm going to say that one of the things that happens -- and I had a conversation with someone recently. So if I've already said it in this group, I can't remember if it was here or elsewhere. It might be deja vu.

Staff tends to defer to physicians. And so you can have a system where you've got really great staff doing really great evidence-based work and then a physician who isn't as aware of the evidence speaks up and says, in my opinion, there is good evidence that, and the staffer just shrinks. And all of that good support you've done just goes out the window. So staffers need to be given a culture where they are encouraged to speak up and given opportunities to help them get the evidence out there and get around some of those cultural things. In some groups, that's not much of a problem. In most groups, that is quite a big problem. So that's very important.

The other thing I think that leaders need to recognize is something we've been saying all along, that this is a process of discovery to do critical appraisal. There is a lot of judgment. There is no perfection in it. And so it's important, I think, to not penalize people who are doing a reasonably good job at this if they have missed something, for example, because things will get missed, and you're doing the best you can.

Mike told you about our work for one of the projects that we did with Kaiser Permanente Hawaii, and I just want to mention that, on our website, we've got the entire case study. And then the Guidelines International Network taped a webinar of ours where we got into some of more nitty-gritty tips. And so this is now not so much for leaders, but for people who are actually facilitating clinical improvements. Some of those tips are really, really helpful, and you might enjoy the webinar. It's, like, 90 minutes, and there are supportive materials in here. So there is a lot of good information that we have, just for case studies and examples, and Mike and I are always happy to chat on the phone about that and what we thought worked.

One of the things though that was really important, using that as an example, is, in their culture, they had an especially difficult time with communications. How many have lived in Hawaii? Because you're all up here up north, right?

Well, I'm going to tell you the culture is really, really different, and we have experienced people really, really friendly here, and we've experienced many friendly people in Hawaii. And there are, like, cast systems that are very strong and entrenched there, and if you are a pharmacist, you do not speak up to a physician. And you have seen how Mike and I interact together, and we have a great time, and we're buddies. One of the clinical pharmacists walked up to me, and she shook her head, and she said, we never talk to a physician like you do.

MR. STUART: Well, Sheri corrected me in front of a large group, and to me, it was just a normal day, but to them, it was talking stink.

MS. STRITE: Oh, my. Yeah (affirmative).

MR. STUART: You talk stink.

MS. STRITE: You don't say anything bad about anybody, and the whole reason why we had this project to do, in the first place, is because there was great disagreement among the clinicians about what they should do for prophylaxis for total hip/total knee replacement, and the pharmacists, in fact, were kind of freaking out, and the hospitalists, they were very unhappy, but nobody would talk to each other. So it started out with a contentious issue and then everybody gave everybody their (indiscernible - voice lowered) over this topic, and years went by, and it was a big problem.

So we had a cultural situation, and we needed to be aware of the various barriers and potential incentives to try to get over that. And the message in this instance was, after the doctors all got used to me and actually ended up liking me and the pharmacists could understand that they actually could speak up, what I think was actually one of the biggest things was that evidence ended up being a neutralizing force.

So we got in there, past the opinions and past the stresses, and we sat people down and said, okay, here's critical appraisal. Let's look at the evidence. And now, they can all engage in a discussion over what they are actually looking at in terms of the strength of the science. So a little culture lesson there.

The other thing that I want to mention is Mike mentioned that, if an individual leaves an organization, sometimes, they've been the standard bearer of an evidence-based approach. They're gone, and now suddenly, all of that is lost. And so what we want to stress is we've given you some tips on how you help prevent that from happening by having educated leadership and then providing support and fostering an evidence-based culture, but then you still want to find individuals that really are going to do this work well because some people naturally take to it and will do a really great job, and a lot of this really gets driven by personal power, especially because, as a broader culture, we don't have a lot

of support for this. So if you find people that engage with this, boy, take advantage of them, put them to work, and use them.

What you want to do though is be really careful of people who will tell you, I know how to evaluate evidence. You need to have a method to actually evaluate that. This is one of our sort of post-pretest questions. I guess what I meant by that is it was more informational, not really pretest. And what we did is we asked people -- and most of these are physicians that are taking our course, a lot of clinical pharmacists. The question is, do you feel comfortable, with your skills, evaluating the medical literature? The vast majority of them said they did not. So a small subset of them said that they were very confident in evaluating the medical literature.

What this slide shows is -- that light blue -- these confident people missed two or three of those critical appraisal questions that we showed you. So there is a big discrepancy sometimes between a person who thinks they know how to evaluate the evidence and whether or not they, in fact, do. So we think it’s important to have a method to really evaluate whether a person knows that or not because, otherwise, you’re going to get a lot of sentences like this, “There is good evidence that,” when, in fact, there is not.

And then quickly, we have just a little bit of help for

you in terms of doing team selection for clinical improvement projects. Stakeholders should be involved. And then you want to make sure that you've got a multifaceted leadership that includes structural -- again, that's part of that support -- that you've got folks that have evidence-based expertise, they help facilitate that in the group, and then subject matter experts in the clinical realm. Of course, we think that around ten team members is the ideal. If it gets too big, it gets too unwieldy. And ideally, they are respected by others, good communicators, hard workers, motivated, enthusiastic, and willing to be worker bees. So this just gives you some of the other requirements that they need.

And then we have a tool that we've included in here because, sometimes, you've got QI Teams or Clinical Practice Guideline Teams. Sometimes, you've got committees that are making decisions on therapies. And so for these groups, which tend to be pharmacy and therapeutics groups or medical technology assessment groups, we've got a tool that outlines various considerations for teams to make decisions and so that's included in there. I'm not going to go over that because.....

MR. STUART: Stop there. Oh, that one's fine. Stop there.

MS. STRITE: Okay. You want there? Because they can read the tool on their own, now that they know that that is

there.

MR. STUART: Just go to the next one because I don't want to take up a lot of time. I'm reacting to that statement that science isn't everything; that's right. We do the science first because it is something and it's important, but then the next thing is the patient perspective about benefits, harms, risks, uncertainties, et cetera. It's all written down here. And then the clinician perspective, which is going to be different in various parts of the country. But this is where I wanted to get to, this "Other considerations."

You have to pay attention to accreditation, to standards in the community, to cost, to ethics, to legal issues, to marketing, to media concerns, to public relations, to purchasing. You've got to pay attention to regulators. And then there is the realities. What can I do? I'm not sure I want to take on a big project that we have absolutely no evidence at all when we do a quick look. And then the impact on our organization, if we do it, this thing about David Eddy saying, we never project what's going to happen; we just do things and then we see what happens rather than saying, this is what we think is going to happen, so we, at least, have some idea of what's going to happen when we make a change.

I think going through the rest of these details is not going to be helpful, and I think what we should do is -- because we've, I think, covered as much as we can get through

today, we should open it up for questions, answers, comments, suggestions to others, suggestions to us, discussions among yourselves, just an open-ended discussion to end the day.

MS. STRITE: Yeah (affirmative). Although I do want to just say one more thing.....

MR. STUART: Okay.

MS. STRITE: .....because we want to make sure that you can cover the objectives. Objective 3: Identify, at least, three factors needed for an evidence-based approach. And so I would suggest that you write these down because I will tell you. You need organizational support, even if you're a little organization of one. You need skills to do critical appraisal. You need critically-appraised information moved into usable information, like clinical recommendations, evidence-based summaries, et cetera. And then you need -- so I'm going to give you a fourth -- some kind of channel for distribution of that information so it gets in the hands of patients and providers, so you can engage in patient decision making.

And then I'll just mention Objective 4, since that may be something you need to fill out on your forms. "List, at least, four elements of patient-centered decision making from the viewpoint of the patient." And there are many answers you could give for that. One of them that we would include though is information based on high quality evidence, quantified

information on probability of benefits, harms, alternatives, impacts, including costs.

MR. STUART: So, Ward?

CHAIR HURLBURT: Yeah (affirmative). Maybe come back to what I mentioned before. Government, in general, is a huge buyer of medical care services, and all states are, and Alaska, specifically, is very big buyer because we're a large state with a lot of people. State government has a lot of employees. There are a lot of retirees from school districts and so on. The Medicaid program is big everywhere -- Corrections.

Could you talk a little bit about in the past? And as I say, I -- you may -- probably are not doing that now because I that they've had to cut back on everything they could, but how you saw our sister states to the south adopting some of the things that you talk about, as you worked with them.....

MR. STUART: Yeah (affirmative).

CHAIR HURLBURT: .....on a recurring basis and apply it to their role in either buying services directly or buying it through others and vendors that they contracted with?

MR. STUART: Yeah (affirmative).

MS. STRITE: I don't think we can really say much about Oregon because our experience with Oregon, as a government entity, has been episodic and has been.....

MR. STUART: Complex.

MS. STRITE: .....based on some individuals.....

MR. STUART: We did projects for Washington, but were paid for by Oregon, for example. We, essentially -- well, you already mentioned that there are 17 states that have kind of signed up for the output of the OHSU -- or.....

MS. STRITE: Which we're impressed with the work that they do.

MR. STUART: Yeah (affirmative). So I would say, you know, you will hear criticism, lots of it, from industry because the criticism is everything comes out being equal to everything else. It's always -- but that's not really true. They just do the evidence, but they have a problem in condensing the evidence. So you've got to have somebody on your staff to maybe interpret it, sometimes. But we've been very impressed with what Oregon's done, and I've been impressed with what Washington's done.

Now, I don't know as much as the P&T Committee there. I just know about the state groups that come -- the individuals that come from whether it's Medicaid or Corrections. What's the department?

MS. STRITE: Labor and Industry?

MR. STUART: Well, also L&I. And every year, folks that work in those divisions come to us. We actually go to Olympia, and we train those people, and they go back and do their jobs, and they do appeals, and they do all these other

things. But essentially, we don't get involved so much in that, unless they call us, and they do. They'll say, what do you think about this study, sometimes, and we do that. So we play a very limited role, but we also have done large reviews for the Technology Assessment Committee, and I've been impressed that Washington has done a really good job of assessing technologies using our approach. And it's not our approach; using an evidence-based approach, I'll say.

And so I think that the only thing I can say is that we don't get the full picture. We just get the fact that they really are hungry to learn the basics, and we stick to the basics.

MS. STRITE: Well, the individuals that show up are hungry. We don't know. So it's.....

MR. STUART: Yeah (affirmative).

MS. STRITE: You know, we get a little pinprick of a view. One of the things that I will mention that's one more item that's in your notebooks -- in the first of part of the section on page 38, there is an example of some decision support that we did, and the probably key one that I would mention is on page 41, where Mike mentioned that we did a systematic review for the state of Washington, and this was on MRI, women at high risk of.....

MR. STUART: This are BRCA1s and other high risk women for breast cancer, and this is a tough position for women to

be in. And so -- but the Medical Director said a lot of people really want a semaphore. He didn't call it a semaphore, but Sheri said, well, what you want is a semaphore. And you can see, on page 41, a semaphore.

I asked the question, what -- the first column is, what's the likelihood of the outcomes, and column three is, increased detection of breast cancer, MRI, a high level of confidence that we can find more breast cancer.

This is a -- you're going to hear a review in 20 seconds that took us hundreds of hours. Is there a decreased need for other tests? No. We have low confidence that you will save money by stopping doing mammograms, for example. Changes in treatment plans. High confidence that you're going to do more surgery, you're going to do more biopsies, you're going to do more of everything.

So you see, you go right down the list. But when we see mortality, we have low evidence that we're going to save anybody's life, but that doesn't mean we that shouldn't be doing MRI for this high risk woman. It just means that's what the evidence is. And in fact, I can understand why, given the evidence -- we have no good evidence that it's going to save your life, and you get more surgery, and you might get unnecessary surgery, but we really don't have the evidence on mortality, and it's going to be a decision that is going to have to be made without evidence. And the state accepted

that, and they did cover MRI for high risk women. But I presented the same situation to the University of Washington Certificate Course students, and they rejected it.

MS. STRITE: And not as much was weighing on them.

MR. STUART: So they were not holding the bag. They didn't have the skin-in-the-game, and the people at the state who were meeting had this peanut gallery full of radiologists and everybody harassing me and everybody else in there about we want this, we want this, and it was a very interesting meeting, but in fact, when the state made the decision, the committee made the decision to cover, but it is interesting that they went with kind of the other factors. The University of Washington sort of post-graduate medical leader of the future group said no.

MS. STRITE: And so my point in showing this is for a couple of reasons. One, to show you here is an example of the way decision support can be constructed to express some very complex decision factors.

MR. STUART: In a meeting.

MR. STRITE: So that's just an example, but the other thing, too, is going back to Ward's question about the state of Washington. In this instance, we saw that they were very willing to consider a lot of considerations along with the evidence. What we don't know though, however, in terms of medical technology assessment, is they are getting information

from a lot of different vendors, and my guess is that probably -- the quality of the evidence review on that probably varies greatly.

MR. STUART: Yeah (affirmative). And they put out bids this year, and we just couldn't do it because there was so much work, we couldn't do the work that they wanted done. So we didn't even bid. We didn't even respond to it.

As far as the Commission goes, I think -- Ward, do you -- I mean, I think a discussion -- anything that anybody wants to say that -- in regard to the evidence as it affects your role would be of interest to the group, too. Anybody in the room that has opinions about that?

MS. STRITE: Yes? There is a hand over there.

DR. GRAHHOLM: Marin Granholm, family doc here with the AFP. But I just want to speak to the Commission and their sort of future role in wading through all this stuff, more from a personal level. I'm in Anchorage the last couple years, but the eight years before that, I practiced in Bethel where a lot of this evidence doesn't necessarily apply really well. And as you're sort of reviewing this stuff, I just want to make sure that you recognize that some of our populations have very, very different -- what else -- genetics, physiology, bacterial populations. Like going to your UTI example, our patients out there had different bugs, and we only know because we cultured. So anyway, just wanted to

share that.

MR. STUART: Thank you. Other comments to the Commission or from the Commission?

DR. TSIGONIS: I had a question. Jean Tsigonis, Tanana Clinic. And I guess I came here hoping to get some how-to's on a practical side and just wondering what your experience is with maybe computer systems, where you've been giving advice.

Like if we had a goal now with the idea that we really want to do evidence-based medicine, we want to apply it, applicable to our culture or whatever -- so you have your maybe four things you want to start with -- how would go about doing that? How can you -- like, perhaps, as you're going through and you've got your person with congestive heart failure and you want to have them on their ACE and their beta-blocker, what can you put in your system that then you've got a check that you can check you've done, so that you get your points you need, so you get paid for what you do, and also that you know you're doing quality care, and then maybe bringing in -- like, you've got these articles that -- to easily bring articles into the chart, you know, or something like, this is what we discussed? I don't know.

I'm just trying to -- how do we do it? I mean, it's all nice and lofty here, but we've got a lot of demands on us that we have to do this, and we have to do it accurately. And we don't have a good way at our clinic, right now, when a person

comes in and then goes, to be sure that everything is checked off before they leave, that it's all done.

MR. STUART: Well, it's a big arena, and first of all, you have to determine are you talking about a process change, or basically, something that has to do with a drug or an intervention to a patient because, if it's a process change for efficiency, there is a lot of debate about this, about what you should be doing.

Certainly, Kaiser uses very creative ways. They always do the evidence first, if there is any. If it's a process redesign, that's going to be quick. But I mean, they may have -- for example, in nephrology, we did a project with them where there is a nephrologist on-call, and he kind of lives in a cockpit, and he's a consultant for chronic kidney disease, and he comes, whatever day it is, and he looks at the charts and says, this patient is (indiscernible - voice lowered) is a little bit high, and they have an electronic way of kind of querying the doctor. Maybe you want to refer to this patient or maybe you want to not be using that nephrotoxic drug, and it's a very elaborate system that's individualized for each patient. So that's one extreme.

The other one is the rules. You know, here's what we want to do checklist. And there is very -- the literature is, I would say, conflicting about this because there are some studies that show, if you do too much of this flagging on

computers, you just disrupt everybody's day, and they get confused, and they don't do well.

MS. STRITE: And also focus just on those things and neglect other things they should be paying attention to.

MR. STUART: Yeah (affirmative). They want to gain it maybe. And so I think it's a very complex situation where what do I do in my group becomes understand the basics, know whether there is (indiscernible - voice lowered) or not and then use your creativity and all the options that you have -- and we are, on our website, have an implementation checklist -- and see if any of those fit for you, as far as implementing. That's not what we're not talking about today. We're talking more about upstream where we're talking about where evidence does come into play, and when it comes into play, you should know the quality of the evidence and the things we talked about today.

As far as putting it together, I'm not going to be as helpful without knowing everything there is to know about your group and that's why I say evidence is global; implementation is really up to you. It's local, and it's takes all the creative powers of everybody who is interested in the quality improvement. Anything you want to add to that?

MS. STRITE: I think that looking at our implementation tool and then a couple of our Q&A things (indiscernible - simultaneous speaking).

MR. STUART: That would probably be the most helpful.

MS. STRITE: Again, we don't know your context, but I can -- if you email me, we can set up a little time where I can spend 15 to 20 minutes with you and point you in some directions.

MR. STUART: Right.

MS. STRITE: Then you can look at things and see what has resonance and makes sense to you.

MR. STUART: Right.

MS. STRITE: We're happy to do that.

MR. STUART: Yes?

DR. THOMAS: Kim Thomas. I'm a family doc as well with the Family Medicine Residency. And I guess this is one of my first encounters with the Health Care Commission here in Alaska, and I guess I'm just curious what you all were hoping to get out of this, where you see this evidence-based medicine influencing your future path and so forth, if this is the appropriate forum to talk about that.

CHAIR HURLBURT: Maybe just real briefly because I don't want to take away from Mike and Sheri here, but our charge, as I started out this morning, has been to look at healthcare here in terms of accessibility and quality and affordability and so on. But my bias, and I've been very vocal about it, is, if there were not cost issues, we would not have a Health Care Commission, that that's -- 18% of GDP is what's driving

us. The biggest factor in state budgets across the country now is what's driving us.

So we're looking at that, and a part of that, very honestly -- and we've looked at pricing in Alaska, which is remarkably higher than Puget Sound, and we've looked at comparisons with Washington, Oregon, Idaho, Wyoming, North Dakota, to pick some comparisons and looked at some of the reasons why. And so in a sense, that's almost low-hanging fruit, although it's a sensitive issue.

But also in terms of quality -- but quality as it relates to cost -- if we do have a non-sustainable situation with the amount of our GDP that we put into healthcare, then, just from the standpoint as docs, we want to do the right thing. We want to have quality. Our patients want to have quality. But we also want to make sure that, where we spend the money, we spend it where it really does the most good. And I think most of us on the Commission, I would say, that we do not see issues of trying to control costs and trying to improve quality as being antithetical at all, but that we believe that both can be accomplished at the same time.

And the Commission then -- just to finish, one of our charges is to make recommendations to the Legislature, to the Governor each year, which we do in January, and one of those recommendations is, historically, has been for the use of evidence-based medicine, both in the clinical setting to

inform your interchange with your patient, but also in the setting of what our payers -- whether it's an insurance company or a large employer, what are they going to cover, what are they going to pay for, and to make that evidence evidence-based as much as possible.

MS. STRITE: Well, Ward, we have just a few minutes before we're closing. Do you want to lead any discussion here further or what would you like to have happen in our remaining few minutes here? And does someone have the Delfini Gram sign-up sheet, where did that end up? Thank you.

CHAIR HURLBURT: I think that, to take advantage of you being here, if there is anything you're thinking that we should have said, that would probably be the priority. Otherwise, if there are no further comments, I think we could break. Are we in -- I think we're in the same room tomorrow. Is Deb in here?

MS. STRITE: Yeah (affirmative). My understanding is we are in the same room.

MR. STUART: This half -- that far-half of the room tomorrow.

CHAIR HURLBURT: And that will be for the members of the Health Care Commission. Mike and Sheri will be here with us in the morning, as we follow-up and talk about that. But Deb, do you want to add any comments on that to wrap up?

MR. KNAPP: My question is very simple. If anybody is

trying out towards the U-Med area, I could appreciate a ride.

COMMISSIONER ERICKSON: Yeah (affirmative). Just a couple of things. The Commission, tomorrow morning, we're going to spend the first two hours of the day having just the Commission have a conversation with Mike and Sheri. After having this day of learning together and getting the opportunity to hear some of the questions and conversation from the rest of the folks in the room and having an opportunity to sleep on it, we'll have an opportunity to have a continuing conversation. And then we'll spend some time looking at the evidence-based medicine Findings and Recommendations from a couple of years ago and see if those need to be improved in any way.

And I think one other thing I would note about this is that we're going to be moving into a phase of the Commission where we're going to start taking some of these more generalized policy recommendations and coordinating with the state agencies that have -- play some role in implementing them in some way. And something that's been significant about this meeting is we've had some of those policy leaders. Our Medicaid Director, the Director for Retirement and Benefits and our State Health Plan and some of their -- the leadership from their programs, the Director of the Division of Worker's Comp have participated in this conversation today. So something else that will come out of this discussion is what

are they doing currently, what are their plans for the future in terms of implementing some of these principles about evidence-based medicine in their programs payment policies.

And as we continue this work, our practice has been to kind of float our preliminary ideas out in draft form for public review and response. So I would encourage folks, if you're interested in seeing the sorts of things that are being discussed in terms of policy implementation at the state level, that you get on the Commission's LISTSERV and track our future meetings and follow the release of reports.

MR. STUART: So our goal here has been, really, to give you some basics, some principles, some concepts rather than to teach you the how's in detail because we can't do it in one day.

The basics though about being -- doing due diligence with the evidence, I think we can make a very good case that that is the way you do quality care. That's the way you do quality improvement is you look at the evidence first. You don't ignore other considerations. You don't ignore pressures that have nothing to do with evidence. It's just that you don't forget to do the evidence and that's our bottom line.

CHAIR HURLBURT: Thank you. I'd like to thank everybody who came today. A lot of you came from out of town and spending the day, for those who won't be here tomorrow, and I'll also express our appreciation to Mike and Sheri for

sharing so well. Bob?

COMMISSIONER URATA: Bob Urata. I'm a member of the Commission, and I have something to ask that maybe you could answer tomorrow. So I guess one might consider it homework, but I was wondering if there is an easier, practical way that one -- perhaps the State could sponsor it or something -- could look on the website and find evidence-based information that has already gone through the processes that you discussed today so that, when I look at the *New England Journal of Medicine*, I don't have to go through all the stuff that you taught me today because it takes time and it takes knowledge. And I'm getting older and somewhat forgetful.

MR. STUART: Here's what I do. To answer your question for you, I would look at UpToDate for what's currently going on for background reading about an issue, but I would not use it for evidence. I would then go to DynaMed because they do critically appraise.

MS. STRITE: And be aware there is variation in any source.

MR. STUART: But that's still -- the best source I know of is DynaMed for many, many questions because they work so hard at keeping up-to-date, more than UpToDate does maybe, and they do it in a different way. They have a very large critical appraisal staff; UpToDate does not. UpToDate gives you background reading of all the things you can do to people.

In other words, every overuse item you could dream up will be in UpToDate. On the other hand, some of it's really good, and you have to have these principles that we did today to know whether you're reading something that's really good or something that really is just a summary of the results section of a study and that's where DynaMed comes in, in my opinion.

MS. STRITE: And I want to give you a bit of a different answer, too. That is that I can tell that you are really young-at-heart, and I can tell you're really smart and you care about patients. And learning this stuff -- critical appraisal can be labor-intensive and onerous work, it's true, but understanding those basic concepts on that page five -- and again, page 25 is just a different flavor -- I mean, it has made my life so different knowing this, even before I really had much basic understanding.

I had a dentist that wanted to move all my money out of my bank account and into his, and what I did was I went to the dental literature and saved myself a whole lot of grief and a lot of money and fired him. So -- and you're going to be a patient, too, at some point, potentially, as we're all getting older.

So I want to encourage people to feel that this is doable to, at least, have the basics so that, if you are, you know, facing a drug rep or a newspaper article or even your colleague, you're aware that there may be issues with the

evidence and that there are even just a few things that you can look at, even quickly, to try to get a sense of if you should even be bothering with that article or not. And that top part of page five gives you some quick things to look at. So if somebody is reporting a subgroup analysis and it's not pre-specified, hurray for further research, and in terms of applying it, I'm walking out of that room.

So honestly, this work is not that hard, and I want to encourage anyone who is involved in clinical decision making -- and that includes patients -- that I think it's a valuable, useful, interesting skill to know, at least, some of these fundamentals, and I think that that will save you a lot. And I know that you're young-at-heart.

CHAIR HURLBURT: You know, I'll my ignorance. I don't know DynaMed; how do you spell it?

MS. STRITE: DynaMed, D -- oh, I can never spell it.

MR. STUART: D-y-n-a-M-e-d. And.....

MS. STRITE: And it's on that short list on page 27.

MR. STUART: Just Google DynaMed.

MS. STRITE: Yeah (affirmative). And we would encourage -- you know, if there is a way to purchase that -- and again, Mike and I are on the Editorial Board, unpaid. If there is a way to make that available for your providers, that would be doing a great service for them.

MR. STUART: I told Deb that we would quit at five, and

it’s five. And so thank you. You were great today. I was amazed at how much critical appraisal knowledge and skills and competency I saw today, and you were a wonderful group. Thank you for being here.

4:59:49

(Off record)

SESSION RECESSED