(PLEASE NOTE: Transcribed automatically by Vimeo; mistakes are possible/likely. Our apologies.)

TRANSCRIPT - GR **12 20** 2**4** ***"Clinical Reasoning Practice and Education: Current State and Future Directions"*** guest speakers Jess Dreicer MD FACP and Andrew Parsons MD MPH FACP, University of Virginia

Internal Medicine Grand Rounds

- Welcome to medical grand rounds. Our last one for 2024 calendar year. We're excited to have Dr. Jess Dreiser and Dr. Andrew excuse me, Andrew Parsons talked about clinical reasoning, practice and education current state and future directions. I'll take us through our Cme slides and disclosures, and then if you are new to grand rounds in the last few months, a few new ways to collect Cme credit. Today's activity. Code 2, 2, 6, 3, 4. Our chief resident, Dr. Shaina Hassan, will come up to introduce our speakers. Thank you both.
- All right. Good afternoon, everyone. It's my pleasure to introduce our grand round speakers for today. Dr. Jessica Dreiser and Dr. Andrew Parsons. Starting with Dr. Parsons, he completed his Internal Medicine Residency, and training here at Uva, following Medical School at East Tennessee State University, and a master's of Public health at Emory University, and he's currently an associate professor of medicine in the division of hospital medicine and associate Dean for clinical competency for the School of Medicine.
- Dr. Parsons. Research is focused on assessment and remediation of clinical reasoning, specifically management reasoning. And he's currently a Phd. Candidate in the School of Health Professions education at oh, I'm going to say this incorrectly in the Netherlands.
- His work has led to impactful publications and national and international workshops on this topic. Additionally, he's an accomplished educator and mentor, winning teaching awards from the Department and School of Medicine, and fostering the growth of future Medical leaders through coaching.
- Dr. Dreiser completed Medical School at the University of Cincinnati College of Medicine, followed by Internal Medicine Residency training at Oregon, Health and Science University, and then joined us as faculty here in 2018. She's currently an associate professor of medicine in the Division of Hospital Medicine and Section of General Medicine Director of Clinical Reasoning and Coaching within the coach program and serves as associate program director for the Internal Medicine Residency program
- and leads the Residency track for leadership and quality improvement in patient safety. Her research and patient safety work focuses on clinical reasoning, specifically diagnostic excellence and enhancing our understanding of prognostication. Her work has led to publications and national workshops on this topic, and she regularly participates in an international group focusing on improving diagnosis, she finds joy in striving to deliver excellent and safe clinical care and teaching senior medical students and residents on the ward.
- winning inpatient attending of the year from the IM residents in 2023 is one of her proudest accomplishments to date. So we're very pleased to have both Dr. Dreiser and Dr. Parsons speak to us today. Please join me in welcoming them.

- Got it
- all right. Thank you for that kind introduction, and thank you for having us. We're thrilled to be here to talk to you about current state and future directions of clinical reasoning, with hopefully, with some information to improve your your own decision making, as well as your teaching related to this area.
- So just quick disclosures to get out of the way, do some work with Nejm and Nbme on clinical reasoning, teaching, and assessment tools which we will not talk about today. So our objectives more specifically are to examine historical and current trends in the practice of clinical reasoning with a large focus on the influence of cognitive psychology. And we'll talk about why and how
- we also want to define the clinical reasoning process and its subcomponents to improve our decision making in our teaching, we'll explore the evolving understanding of diagnostic management and prognostic reasoning, which is how we'll break down clinical reasoning as a larger umbrella term, and then we'll assess the current state of the literature to imagine a future era of Co reasoning with AI or large language models.
- So let's start with, why are we here talking about this topic? You know. Why do we care?
- You may be familiar with that. We're thought to be in an epidemic of diagnostic error I mentioned. We're going to break down clinical reasoning into 3 kind of sub processes of diagnosis, management and prognosis. And it turns out each of those, if you think about them in terms of the outcomes, we don't do as well as we would hope.
- Diagnostic errors. Delay or delays occur in up to 20% of encounters contribute to 10% of deaths, and up to 17% of hospital adverse events. And it's thought that up to 3 fourths of these errors have a cognitive component meaning that clinical reasoning is likely the cause or faulty clinical reasoning is likely the cause.
- Now, it's important to note that atypical presentations of common disease are thought to be the primary drivers of diagnostic errors. In other words, it's not the rare diseases that we're missing. Necessarily, it's common diseases that are presenting in an unusual fashion. And I think it's also important to note that
- if you look at the biases and inequities that have existed traditionally or historically, in in clinical research, it's likely that what we learn to be atypical is actually typical in and of itself.
- Now, on the management side, another epidemic per se, approximately one in $3 in health care is thought to be wasted in the United States. A large contributor to that is low value, care, unnecessary services. There's a number of things that go into this right misaligned incentives, etc. But at the end of the day
- unnecessary services or things that we order, and so it can be thought of as management errors in this sense.
- and then finally, prognosis, which Dr. Dreiser is going to dive into in more detail.
- I'll just say quickly that we're not great at that, either with studies. If you look at patients near the end of life, we're around 20% accurate in our estimations of their prognosis.
- So if we're going to improve in these areas, it's important to kind of understand where we've come from. And so we're going to look back at the history of clinical reasoning, and it turns out clinical reasoning has really evolved through 3 distinct historical eras. We could go back really far, of course, but we're actually just going to go back to the mid 19th century and start there

- where clinical reasoning was really influenced by the development of pathological anatomy of statistics, as well as John Dewey in his book here, how we think, which focused on critical thinking, and the thought was that critical thinking was a skill
- that could be applied regardless of context. So at that time clinical reasoning was thought of as a general theory of logic and problem, solving
- probably the most notable artifact of this was the Clinical Pathologic Conference or the Cpc. Which I know we're quite familiar with, and many still do to this day, where an expert discussant is asked to think aloud as they're given, you know, multiple aliquots of case information.
- So in terms of what do we take from this era, we'll think aloud is a big one. The importance of reflection is a big one, you know. We're recognizing our underlying assumptions as we think aloud in a group, providing justifications for our ideas for our actions, and evaluating these these justifications.
- Now the second era was really informed by the intellectual upheaval of World War 2.
- During this time new tools were developed and refined during the war, such as screening algorithms to select these and receiving receiver operating curves. This is actually interesting, quick story, which is, we were looking at German bombers coming in to London and realized, you know, trying to calculate the rate of false positives and false negatives, because obviously that informed the response to those bombers.
- And so a key take home from this era is that reasoning or human reasoning could really be modeled.
- This new field was really informed by something we call decision theory. And so decision theory is a formalized approach to understanding how choices are made by weighing the options, evaluating probabilities, and calculating the potential risks and benefits to different outcomes. So at its core decision theory really says that we act rationally to maximize outcomes. Now you might think that sounds.
- you know, too good to be true, and it really is. This comes more from economics than it does from medicine.
- But at that time the hope was that that's really how we could behave as human human beings. The prototypical, prototypical example of decision theory is probabilistic or Bayesian reasoning something we know today pretty much as evidence-based
- medicine. Right? So you're taking pretest probabilities of disease from the epidemiologic literature, you're continuously updating the likelihood of various outcomes as new information becomes available.
- So a big take home from this era was that diagnosis was not something that was absolute. In fact, uncertainty was a huge player, and so better described as probabilities.
- and that diagnostic tests rather than telling you if a person absolutely did or did not have a disease, it raised or lowered the probability.
- So let's move into the 3rd era, which is also the current era. This began in the 19 seventies.
- and this was largely shaped by cognitive psychology.
- Cognitive psychology acknowledges that we frequently rely on heuristics or kind of mental shortcuts.
- Biases is a way of thinking about this, I think you know, for good reason, we tend to think of biases as always negative.

- But biases are really just our cognitive disposition to respond to something, it can be positive it can be negative.
- So cognitive psychology acknowledges that we rely on these heuristics. Our intuition for our own reasoning, and this very much comes from our experience.
- So think of. I want you to think of this era and the last era as kind of 2 things in conflict. Right? Are we purely logical
- in our decision making?
- Or are we biased in our decision making and use something called pattern recognition to make decisions? Those 2 things can complement each other, but also be in conflict. So let's go into cognitive psychology a bit more.
- The biggest thing that Daniel Kahneman, which you probably recognize. This book which he really brought us was something called dual process theory. And I think if you're probably familiar with any theory that informs clinical reasoning, it's this. This is more than anything at the current moment
- informs our understanding of clinical reasoning. But we're going to go beyond that. I 1st want to start here
- on the importance of dual process theory. So it says, there's 2 systems system one and system 2 and system one is intuitive, unconscious.
- And it's pattern recognition so based on experience that you obtain through medical school through residency, through your clinical practice. You see patterns of illness.
- You also see patterns from a management or a prognostic standpoint, and that informs your decision, making
- right. If you don't recognize that pattern, or even if you do, but you try, you actively turn on a more analytical approach. That's system 2. That's much slower. It takes more mental energy
- to do that. It's thought that we spend about 85 to 90% of our lives in system one. This applies to medicine, but also just throughout our day to day life. Because we're conserving mental energy. We're using biases like the easiest example of this would be stereotyping. Right can lead us astray, for sure. But as many times correct. And so we use that type of bias to get through our day and to recognize patterns where we may not be 100% accurate. But we're largely good enough
- now. We know experts can toggle back and forth between System One and System 2
- to arrive at the diagnosis based on their experience and based on how much they recognize the pattern or how much they don't.
- Now, a key aspect of understanding dual process theory. And to inform your own practice and teaching
- is to think about knowledge. Organization. Of course, knowledge is critical to the practice of medicine right, and is critical to our application of that knowledge in the form of clinical reasoning. But it's not just the amount of knowledge. It's how that knowledge
- is organized. So over time, as we, you know, gain experience. We amass this rich cadre of illness scripts. You can see an example of illness scripts here
- an illness script is a constellation of signs and symptoms. Contextual cues related to the diagnosis of a given disease. There's also something called management script, which is a list of the potential interventions we can do specific to a given disease. So that's focused on management as opposed to diagnosis. But as we amass this cadre of illness scripts over time, we organize that so we can access that when needed.

- can think of this retrieval as kind of this neural highway that's stored in our long term memory, but it's stored there as a script or as a chunk of knowledge that's organized. And that's important, because our working memory is quite limited.
- Right? And so at any one time there's capacities thought to be
- 7 to 9, 5 to 9 objects. These can be items or chunks. I'll show you an example of this in a minute.
- and we want our working memory
- to be effective, but also to be free and opened up so that we can. You know we're not overwhelmed. We can actually think our cognitive load is not too much
- so to show you how this plays out.
- If you've done this before some of the residents have been, then please don't. Don't mention it, but I'd like you to just
- read through these letters one by one.
- and then close your eyes to yourself and see how many of these letters in order that you can recall
- to yourself.
- Let's take a few seconds to do that.
- Anyone like to share how many we're able to do.
- How many letters.
- 26. Okay, thank you. Yeah. 6. Alright.
- Now, like you do the same thing
- with these letters right now. These are the same letters, but reorganized.
- And of course, if you did this activity, you could do this with no no problem. I would think right. But I want you to apply this to the medical context
- and to the development of medical learners through medical school and residency. Right? The Cnn is a chunk of information to you because you have experience with Cnn. You know what it is same for FBI, Cbs. You can apply that to community acquired pneumonia PE, whatever it may be
- right. But to an early medical learner. These are still just 16 isolated letters, right? And so
- they aren't able to chunk like that. And so their working memory is is overloaded and they can't recognize the pattern. But they also aren't able to free up the space in their working memory to be analytical if they need to be analytical. So that going back to that toggle back and forth
- alright, so so far, we focused on content knowledge. So we've focused on biomedical knowledge.
- clinical reasoning in that aspect, what occurs in the head. And that's extremely important. But it turns out that it's not just content. It's context.
- all right. So we're learning more and more from studies of clinical reasoning. That context is critical and specifically context, something called context specificity. So this explains why performance in terms of clinical reasoning varies across clinical encounters, even when a patient's presentation and the diagnosis remain the same, so even when the content remains the same, your knowledge remains the same when contextual factors change, performance changes.
- So this acknowledges that medical knowledge and skills are not applied uniformly. If we go back to that 1st era. 1st era of clinical reasoning. Right? It's not just critical thinking, it's not a skill we can develop, and we can just apply uniformly regardless of context. The situation is very important.

- So the clinical reasoning, literature and kind of this current era is moving from kind of in the head to out in the world, and how we understand our clinical reasoning practice in this way. So we think of these micro theories, things that occur inside the head as dual process, theory as illness, script, theory, things we just talked about.
- But there are these larger, more macro theories that inform how we think about clinical reasoning out in the world, and there are many of them. But I just want to hit on 2 to really drive this this point home. So we'll just talk through 2 of these, the 1st is situated cognition and situated cognition posits that reasoning is the result of dynamic context specific. Again.
- bi-directional interactions between individuals and their environment. And that's a bunch of fancy language to just say, it's not what occurs in my head. It's between me
- and the patient, right? It's between me and the environment.
- It's between the patient and the environment.
- You know this to be true. I think it's intuitive right that if you take low socio socioeconomic status things like this. We know this impacts
- the care of our patient.
- but I think we still traditionally think of clinical reasoning as just cognitive processes that occur in our head. And instead, it's this dynamic interaction back and forth and over time.
- Now, the other one I want to talk about is distributed cognition and distributed cognition, basically that moves beyond just us and the patient and the environment to
- complex systems. So this can be our interaction with the electronic health record. This can be our interaction with computers, technology with AI, as Dr. Dreiser is going to talk about.
- And so in this view, cognition is dispersed among all of these, and has a much more greater capacity for reasoning. So I think if we're going to improve our reasoning, we have to think about the process like this. Just to give you an example.
- Right? Let's say there's a motor vehicle. Accident.
- Patients rushed to the local hospital right? Some imaging is obtained. This plays out over time. They realize they need a surgery. They don't have that. The capacity to do that at that hospital they fly them
- to another facility, right? New doctors, new teams review them there.
- And eventually they go to the operating room.
- The important thing to realize here is that there's interactions between many different people, systems, artifacts, tools. And I know you're familiar with this every single day.
- Right? But I don't think we realize how clinical reasoning encompasses this complex system. The important thing to know is about this horizon of observation. So each of us
- is limited in how much we can see and view right, we try to get medical records as quickly as we can. We're trying to understand the complexities of the situation and what has occurred.
- But you can think of this as not having the electronic medical records. But you can also think of this as the nurse on the care team, having a different horizon of observation than you do right, and a different perspective, and thinking about things in a different way. So all those have to come together to work effectively to process information for effective clinical reasoning

- in terms of learning in terms of our teaching.
- How this matters is that it tells us we can't really just teach clinical reasoning
- out of context in isolation, right in a classroom.
- I know we do that, anyway, because of the logistics of how education system is set up. But we're missing the role of context and of these different factors, I think that explains, especially when you come at this from a remediation standpoint, or when a learner or a clinician struggles right? But they do well in that one-on-one coaching session.
- But then they they struggle when they're back in the real authentic environment. And it's largely because of context.
- So with that understanding that theoretical understanding, we wanted to dive down into each of these processes to try to improve your own decision making in your own teaching. So, as I said, we're going to think of clinical reasoning as diagnostic reasoning, management, reasoning and prognostic reasoning. And we're not just making this up. This is how the literature is beginning to break these down. And they're distinguished because there's different processes, different cognitive processes and different contextual factors that are in play in each of these
- processes. So we'll start with diagnosis and we'll do management, and then I'll hand it over to Dr. Dreiser to talk about prognosis.
- So what does the diagnostic reasoning process look like?
- You can break it down in multiple ways. I think, because we teach a lot of this here you'll be familiar with with some of the terminology.
- but I start by these 2 steps. One is generating hypotheses, and one is gathering information.
- And I start here because these could be flipped. And I want to make an important point about the development of clinical reasoning.
- We could start off with hypothesis generation, or we could start off with gathering data. And it depends on if you're a novice, if you're just learning, or if you're an expert per se in this, because
- the difference is whether this occurs inductively or deductively. Okay, if we know that experts based on their experience can recognize patterns.
- and that's how they reason.
- then, that incurs inductively I gather information from the patient from their surrounding context, and as I gather it, I'm bouncing that off my illness, scripts and and deciding, do I see the pattern, or do I not? Right? So with limited amount of data.
- I can quickly recognize the pattern. This is why experts are more efficient because they've honed their pattern recognition right? Just a few key features. I know exactly what questions to ask
- in an inductive way. I've made that pattern recognition. Now, novices don't have the experience as we talked about, and therefore they don't have the pattern recognition skills. What they have to do is go through an analytical process. So they're much more deductive.
- So the key thing for early learners and developing learners is to generate hypotheses early.
- So as they get a little bit of information, we want a very broad differential. And then we go through a deductive way of like hypothetical deductive process of testing against each of those differentials.
- All right, so it could go either way, inductive or deductive.

- regardless of how we do that, we develop a problem representation. That's kind of, you know, in in communicative terms. It's it's the summary statement. But a lot of times this is occurring intuitively right. And we're again. We're bouncing this off of illness scripts as we go.
- That allows us to prioritize the differential down. As we gather more information, we finally choose a working diagnosis. And hopefully, we can justify that diagnosis in a reflective type of process.
- This is just a hammer home. That knowledge, of course, is key to this as well as context as we've highlighted.
- So there's numerous ways that this could apply to our own practice and to teaching. I just want to highlight 2, because we're going to go into the other areas as well. But the 1st was this novice
- deductive versus expert inductive. We know that early hypothesis generation among novice learners improves diagnostic accuracy. So you want early hypothesis, generation based off the chief concern or chief complaint. Some behaviors have been identified in history, taking
- that suggests that a learner is engaging in hypothesis, driven data gathering. That's thorough explanation of the chief concern, early asking questions in close proximity, clarifying questions, summarizing. I think these are behaviors, you probably know intuitively as an educator, just to write them out for you, and then finally forcing prioritization is a good teaching skill. Right? You get just a little bit of information from the emergency room, let's say, asking yourself or the learner, what's my broad differential?
- What's the high yield information related to each of these diagnoses that I want to ask for. First, st what's the high yield exams that I want to conduct?
- Second thing to tell you about is structured reflection. This acknowledges the importance of knowledge organization.
- There's a number of different strategies that have been put out in the literature around structured reflection.
- These are cognitive forcing things. Where we stop and say, let's consider alternatives. Let's do the opposite. Let's use prospective hindsight. Let's think like an outsider right? These, these are structured reflection techniques. There's also conducting an assessment of fit
- which I think many of you have heard before. But this is basically saying, once I've got a working diagnosis.
- what features of the presentation are concordant with that diagnosis? What's discordant? And what would I have expected? Given that diagnosis? But is missing. And you can see for simple and even more so for complex problems that improves our diagnostic accuracy.
- So that's diagnosis in a quick nutshell. Let's jump into management reasoning.
- And to understand management reasoning.
- I think it's important to compare it to diagnostic reasoning. That kind of gives you an anchor as a way to think about this right in diagnosis. There's usually, of course, there can be more than one. But there's usually one diagnosis, one objectively correct
- answer. You either really have pneumonia or you don't right in management. There's multiple reasonable solutions. So it almost always depends. So we talked about the importance of context in management context is even more important.

- because, regardless of whether you have pneumonia or once you have pneumonia, there's many different ways that we may treat that
- inpatient, outpatient follow-up, etc, based on contextual factors. Right? So patient provider system
- constraints think about during the the pandemic. These things really influence our management.
- It requires communication and shared decision making with the patient. It requires ongoing, monitoring and adjustment, because it's kind of a moving target, and so overall, you can think of it as a prioritization task. That's where it's a different task that we're teaching. It's not a classification. It's not recognizing a pattern and fitting an illness script. It's prioritization among different options weighing the benefits and risk of different options. And this is why, overall, I would say it's a more complex skill.
- And so we naturally focus this later in medical school or largely in residency.
- If we were to break this down into a process. And again, this is overly simplified. But for purposes of our own decision making and teaching, we would start with a working diagnosis. We would activate a management script. And what I mean by. That is, let's say, we, we think our diagnosis is community acquired pneumonia
- right immediately. We should activate kind of a menu of options in our head. A menu of interventions related to that diagnosis, irregardless of the patient or the contextual factors. It's a full menu. It's not what we're going to do, but it's everything we could do.
- Okay? And so this is kind of the key step here in management reasoning, moving from that menu of options which is an early skill. That's knowledge based. You have to realize that as a Med student, right? And moving into residency, what's my menu of options? And then it's selecting off that menu based on a number of factors. And you select off that menu. I'll show you what those influencing factors are thought to be in a minute. And then, of course, once you implement that plan, you have to monitor and adjust over time.
- Now, I'm not going to go through all this. Just want you to see the complexity here. This is some work that Charlie Morris, one of the residents here, is doing with us in terms of identifying specifically residents in this case, what are the influencing factors that
- that are at play as we select from that menu
- and choose what to do to our patient? So there's factors related to the actual intervention. There's patient, specific factors, case factors, a lot of provider specific factors.
- then setting legal risk feasibility, etc. So how do we utilize this as a teaching
- tool? Well, let's say we have a patient coming in.
- You know they're 50, and they're coming in with fever cough, and some shortness of breath.
- On the 1st column you have basically a management script. These are all the categories of interventions that you could do. And you want to activate that script. So if you think of if you're thinking about community acquired pneumonia or just that syndrome, I described right. What are all the possible menu of options. You can use this to, you know, for an early learner for teaching, because you you want to make sure this is all inclusive.

- Then it's about pulling out some of these influencing factors, some very common ones, diagnostic uncertainty, risk value considerations such as cost, and then shared decision making
- right. And this is giving you a language. This is giving you a way to think aloud, and a way to ask your learner to reflect, to say, why am I choosing something off of this list for this given patient? But maybe another patient with the same condition. I'm choosing something else. So, management reasoning again, it depends. It's prioritization. And it's all about these influencing factors. And we can use this to help us teach management so that I'm going to hand it over to Dr. Dreiser to talk about prognosis, and then the next era
- of uncle reasoning.
- Alright, y'all hear me
- all right. So now, kind of moving to that 3rd category here of prognostic reasoning. So, historically speaking, it's been understudied compared to the other subdomains that we've talked about in terms of diagnosis and management. This is kind of older data demonstrating that, and then also replicated in some survey data that myself and some colleagues have collected from experts
- in clinical reasoning around the country, sort of asking about the frequency with which they do prognostic assessments themselves, and the frequency with which they assess that and teach that. And in learners. And again, we see sort of a gap here between those 2 things
- to kind of anchor us in, you know. Sort of this idea of prognostic reasoning, which is, is kind of a newer idea in clinical reasoning that we think is valid to sort of pull out because of this gap in, how we do it, and how we sort of teach it and assess it wanted to kind of offer this definition, which is that it's really the process of judging the severity of a patient's illness and injury.
- That's kind of the 1st step, and then based on that thinking, forecasting, potential clinical outcomes in the future. So step one, you know, we often shorthand. This, as sort of, you know, is this patient sick or not sick?
- You know it's not actually a binary. We're sort of, you know, assessing this in a more nuanced way, and then sort of after that assessment, looking to the future and making forecasting outcomes. So this could be a hyper, acute sort of forecast like, does this patient need to be intubated in the next 60 seconds? You know, all the way out to what's the 5 year survival predicted for this patient. So
- that's kind of the kind of to to the essence sort of of prognostic reasoning.
- And then, similarly, to how we discussed diagnostic reasoning. There.
- there's sort of a system one component, and I think that this is one that we often now Hone, maybe not explicitly, often through teaching, but sort of through experience, where we sort of say, Okay, you know, what's your what's that? Does this patient pass the eyeball test what's going on with them, and we sort of quickly make a assessment of their sort of illness, severity. And there is a bit of empiric data to support sort of eyeballing patients, mostly from emergency room experts, physicians
- where, when they were sort of given even just 15 seconds to look at a patient. There was sort of fair agreement in their assessment of how sick that patient was, and then, when they were in a separate study. When they were given a little bit longer up to kind of 5 min of clinical information. They could actually predict sort of that forecasting. They could predict about 80% of the time the disposition of the patient. Were they going to get go home? Were they?

- Would they be admitted to the floor to the Icu. So sort of given some validation to our sort of experience doing that kind of assessment.
- And then also, there's there's many sort of system 2, you know, risk scores, large data sets about particular conditions that allow us to to sort of do the forecasting as well. So this is just one example, sort of based on a patient's Gfr sort of their level of kidney disease. Some other patient, specific factors. And you can see this will sort of
- spit out for you both a kind of 2 and 5 year sort of forecast of their risk of transplant or dialysis needs. And then this sort of nicely ties to the management decision. Sort of you know. What do you do about that and give specific recommendations here in this example of kind of you know when to refer to a nephrologist, for example.
- Similarly, also that we've talked about the the sort of impact of context is is very important in prognostic reasoning. And so this is from a sort of scoping review of our prognostic assessments and sort of finds that in aggregate we tend to be pessimistic
- on our prognostic assessments versus sort of the gold standard or actual. But when you sort of look further into more specific contextual factors, you can see other patterns, so one that I think I had sort of heard on but heard about before. Digging into the data was sort of this, you know, specialty impact. And so sort of the optimistic oncologist, you know, does bore bore out in the data
- along with the generalists. And then you sort of see that the the specialists that spend their time with the very ill patients have more of a pessimistic view of
- prognosis.
- We also see this tracking with other factors. We see this improving in terms of accuracy with experience. And it's also impacted by things like framing. So if you ask about sort of the assessment of mortality of 6 months versus survival at 6 months. That will impact prognostic assessments. And also where you're seeing the patient, are you in clinic? Are you in the intensive care unit
- as well.
- That's a kind of bit of a primer on prognostic reasoning. And then I'm going to turn to sort of what we're kind of naming era for this kind of code reasoning, where we're going to pull in generative AI or sort of large language models into how we do clinical reasoning. So I'm going to tell you a little bit about the current state of the literature, which is literally changing day by day.
- and then sort of what kind of looking through that sort of what can we do in terms of best practices now to sort of utilize that in our clinical reasoning?
- But first, st I'm going to have a little aside and tell you a story about an ox, and so at the end of this you will at least have a piece of trivia or a new skill. If you play group trivia to help you answer random numerical questions better.
- So the story is that in 19 0 7. There was a statistician in the Uk. Named Sir Galton, and he attended a county fair, and noticed that there were about 800 people who had guessed the weight of an ox, and he noticed that if you looked at each individual guest that the average sort of the individuals were off was by about 3%.
- But if you took that data sort of in aggregate and looked at the Median, that the guess was actually within 1% of the actual weight of the ox, which was 1,198 pounds, and the key points here. These were independent guesses from a large group of people.

- and that they were sort of taken together. So we will. We will come back to how this is helping me think about sort of the strengths and weaknesses of large language models and sort of their ability to reason or sort of, you know.
- Sort of reason, I would, I should say.
- okay, so just to kind of get on the same page of you know what is large language models? Generative? AI. I'll probably use those terms interchangeably, and as I am also sort of, you know, very recently trying to wrap my mind around what this is
- pulled out. This definition from Openai's website. So really, Llms are a model that learns how words tend to appear in context with other words and large language pulling from a, you know, a big swath of text. So really pattern recognition.
- And then based on that pattern recognition in response to a prompt or sort of query from a user, predicts what would be the next best word to come in a sequence and sort of does that iteratively, so hopefully.
- with that background we can kind of look at what has been kind of published so far in the literature and reasoning.
- So still wanted to stay pretty broad here, and just sort of recognize that this is really a, you know, going to be a hugely disruptive technology. We're going to kind of zoom in on sort of the critical thinking or clinical reasoning piece. And you know, also mention prompt engineering and a bit about sort of the bias that's been found so far. And what are the strengths and weaknesses as we see them now.
- So, as Dr. Parson sort of alluded to, you know, medical knowledge is necessary, but insufficient, you know, to sort of fully engage in optimal, clinical reasoning. But question one is, you know, medical knowledge. And so this, this information is probably familiar to most of you. It's made a lot of headlines that these Llms have done quite well in answering Usmle or Usmle like
- medical knowledge questions. And you see that sort of over time. There's increasing performance as these models get better and better. So sort of medical knowledge check, we have kind of reached. They have reached that.
- And so the next question is, how about diagnostic reasoning? And so this is a study from Dr. Go. And Dr. Parsons was was part of this group as well.
- That took complex medical cases and compared physicians plus usual resources like up to date physicians plus Llm. Chatgt before in this case, and then an Llm. Alone, and you may have also heard about this, because this has been kind of all over the media.
- and I think much. Probably I assume, to the chagrin of the kind of like hypotheses. The highest performer was actually Gtv. 4 alone and be out kind of the clinician with both resources.
- And so the question is, sort of you know this is maybe a bit surprising. You would have maybe expected, at least, if Gtp. 4 was going to outperform us, that we should, with its help, be able to match it, and I think there'll be sort of forthcoming qualitative analysis that will speak to this a bit further. But I think there's a couple of sort of theories about why this might be
- so. One is anchoring bias so clearly in the cases that the Gtp. 4 was giving the correct answer. The physician did not change. They sort of stuck to their their original thought. It didn't change over to the correct answer.
- and then the second is again going back to me, giving you a definition here. Right? This is a very new technology. This is not something that we learned how to use in Medical school or Residency. We're sort of all learning in real time. And so was

there optimal use by the clinicians in this case to sort of get to the diagnosis. And so that brings us to this idea of prompt engineering.

- Which I sort of alluded to earlier. And so this is
- basically a sort of a new version of the idea of sort of garbage in garbage out. And so what it boils down to is, it matters
- how you prompt these, these and the prompt is sort of the question, or the you know, the text that you put into to the Llm. And so
- it matters, the the framing that you give to the Gbt. Even something like telling it. You know you are the expert on this, and sort of giving that framing will give you a different response and output than if you ask the question or prompt in a different way.
- And so I think that this is going to be important in both our clinical use and our teaching, and also learning from our learners who, you know, are going to be. Some of them are going to be more adept at using this technology, but is an important principle in sort of utilizing this for our reasoning.
- The next question is sort of about management. So this is the same group. And they in this case looked at management questions. So, as Dr. Parsons mentioned, these are complex cases, and often don't have one right answer. And that was true in these cases that they chose. And interestingly, they found a different result this time. So they actually found. It's not pictured here, but equal performance between
- the the Llm. And the physician plus Llm. And both of those groups outperformed the physician with the conventional resources, you know, up to date Google, etc.
- And so, you know again, sort of hypotheses, you know. Why did we see a different result? Perhaps, in these cases of management, where, unlike diagnosis, there wasn't like one answer. The physicians were more open to the suggestions of the Llm. That you know I'm not being. I'm not wrong because there's no right answer here. So I'm more open to sort of accepting suggestion.
- So yeah, I think we'll see more. This is, was just accepted for publication, and then, I'm sure, more qualitative analysis to follow. But perhaps why, we're seeing kind of the differences here.
- and then, not surprisingly, I haven't come across any sort of similar studies in prognosis, but I did think that this was an interesting study to kind of speak to the prognostic reasoning abilities, so to speak. So this was a study where they used real data from a 911 operating system.
- and they grouped the calls into groups of 4, and they had a panel of paramedics, and they gave them these groups of 4. And they said, Which of these cases, do you think is the most urgent. And then they posed the same thing to chat. Gtp. 4. And so in kind. Of all of those cases there was about 3 quarters of the time there was agreement between the paramedic panel and the Chat Gtp.
- And but interestingly, when you excluded, when you only looked at the cases where there was unanimous agreements among the paramedic panel, that agreement increased to about 94%.
- And so I've kind of presented broadly, you know, sort of some early studies and evidence in sort of those 3 broad domains of diagnosis, management and prognosis. But I think there's also some interesting things that can be gleaned by looking at sort of more specific use cases and sort of highlights, the strengths and weaknesses of Llms. And so this is a study where they looked at sort of 4 different more specific cases and sort of gave us performance data in relation to them. So a case of pneumonia, breast cancer sort of a question of does this patient need a

stress test? And then a case which they sort of labeled as uti, but really was actually bacterurea with a urine culture and I think there's a couple of of takeaways, you know. We'll have to sort of have more studies to see if this does generalize but but basically, the way it worked is they had both the Llms and the humans sort of say, before any just given a little clinical information sort of what's your pretest probability of this disease? And you can see in blue, you see, sort of the true pretest probability. And then, in the grayish color. You can see kind of the human distribution, and then in the yellow, the Llm. And so you can see the the kind of before. And then they were presented with a positive test in some cases, and then asked to say their post test probability and a negative test in some cases, and asked to say their post-test probability. And so the green boxes show the negative test result. And so you can see that in those cases, after a negative test result.

- The Llm. Was superior into moving very close to sort of the true range and outperformed the the humans but the sort of case D, which, again, was sort of a case of bacteria, you know a slightly more complicated diagnostic case. You see that the Llm. Actually moved in the wrong direction after presented with a positive urine culture in the case where the patient was asymptomatic.
- And so I like this because I think it complicates us a little bit more, you know my takeaway is that it is. The Llms. Were were better most of the time and certainly that they were more precise, but not always accurate. And so I think that sort of what I'm taking from this is that that we really need to think about the context and the case specificity when we're kind of applying this to aid our reasoning. And even though it may be true that most of the time it may outperform us, we'll sort of see if that holds up with future studies, that it's not going to be something that we can sort of universally adopt, because that's not always going to be true. And then, lastly, in terms of studies, I'll just share this other one that I found really fascinating on biases. So we talked about kind of cognitive biases earlier, and in this study they presented the Llm. With cases to sort of evoke to see if it would sort of fall for these same cognitive biases as humans do. And you can see kind of on the right hand side here that in 8 out of 10 of the biases that the Llm. Was actually more prone to these biases than sort of the rates that we have on physicians and I think that the the 2 that that base rate neglect is is interesting, you know. That sort of thinking about. Is, was it susceptible to? Or what? Yeah, was it susceptible to sort of being let away from kind of the actual you know, frequent rate of disease. And so I think it makes sense that the the model was not swayed against that, that it was able to kind of know the the true rate of disease, that it would be better that than sort of humans. As we discussed. You know, we don't think numerically in in the same way as actual mathematical models and then the capitulating to pressure is also interesting, because this is sort of having a request from a patient that's maybe not fully appropriate. And so it makes sense to me. I got a reminder. These are not humans, and it was not as susceptible to maybe the tug that we might feel to sort of make a decision to kind of maybe serve the human relationship that we have.
- And so here is where we sort of go back to. Why I told you about the ox. So my theory about kind of why was it more susceptible to all of these other cognitive biases? So again, if we go to back to thinking about the story, and why did sort of aggregating. You know those many different responses get us closer to the truth. And again it goes back to the fact that those were all independent guesses that were from people that had no special knowledge about. You know how much an ox weighed but if we go back to the understanding of how the Llm works. It's pulling

together these, these associations from many human sources that have biases. And so we're sort of seeing a concentration of human biases, because it's pulling together those pulling together, all of those those human biases. And in the text, and so kind of takeaways from from all of that together. I think that using

- I'm really excited about the possibility of this really could improve, you know, diagnostic, prognostic management errors that's super exciting, you know we will, I think, kind of see hopefully soon, sort of what these true effects are. But I think now thinking about it really as a readily available. Second opinion, sort of independently, you know, this is my idea of what's going on.
- And then, just as you would, you know, before opening up the reed of a chest X-ray, and then kind of calling on one of the Llms. For a second, consult using it, using that that prompt engineering knowledge particularly to challenge you when you maybe feel like you're missing something, or you're you're anchoring, you know. You can ask it to like. Tell me why I what I might be missing, or what doesn't fit here, I think that's that's a really exciting application.
- And and then sort of from our our final examples of just really thinking about the use applications and the specific cases and context when we're applying it.
- All right. Now, since I'm going to skip over that if this was super fascinating to you, we would love you to join us in our clinical, reasoning, reasoning, research, collaborative, and kind of study. More of this and thank you for your attention. We'd be happy for any comments or questions.
- Raise my fingers. So my question is, oh so my question is about evaluation feedback. So like as you're teaching these concepts, especially to early learners you know, you guys did a great job highlighting how complex it is, and I think evaluating all those domains can be equally as complex and context, dependent stage. What institution, what team they're on so like, how how do you think about evaluating clinical reasoning oops?
- So I think it's kind of acknowledging the what we said like content and context, right? So I think there? Well, the short answer is that there is no one kind of gold standard way to evaluate this. I think it's because of the complexity as well as the content context dependent or in all different contexts. Right? And there's it always depends every single patient where we're doing that. So there was a scoping review of this clinical reasoning assessment done this a couple of years ago, basically what it says. You have to do it all, and then you have to look at all of it and make kind of a holistic decision about a learner. And so you want to be able to assess them in as many different contexts as possible, and see how they're performing. And so to me, in terms of like, how does that inform our teaching? Well, I think you have to pick a few of these things that are probably most critical. Right? Like, I focus on early generation of hypotheses, or like stopping, reflecting, thinking loud, right, like these key things that you can teach us skills.
- And then assessing the learner on that as well as you know, obvious content knowledge based things in many different contexts.
- And I'll just add, from kind of a remediation approach, that what we often do is we try to to identify the like most severe, you know, issue basically and start there because it can be, you know, overwhelming, as you said, to kind of try to focus on everything. And so what do. I feel like they most need to develop and start there and then build one other thing about part of why we try to break down each of these into process is because when we started right, if you just think of this in terms of outcomes, diagnostic error management, we don't do very well. And so. But a lot of

that's based on outcomes or later chart review or different things, right? So it's hard in the moment to adjust the outcome. And so I think what we're left with is thinking about each of these in terms of a process and saying, Okay, well, let's work. I mean, that's what we've done.

- Remediation board is like. Here's the process. Let's give you the process. So you have a mental model. And then we were struggling. Each one of these is my 1st skill.
- Making this entirely personal, I, over the last 6 months have moved from being like an upper level resident, and a general attending to a Cu fellow, and I feel like my entire job now is, residents can without me make a diagnosis, make a treatment plan, and my job is, do I think that we can turn this around in an hour on the floor? Or is this someone needs to come to the Icu? And then, once they're in the Icu talking with family, and I'd like with plenty of times, and that's my own problem. But it's also really hard to communicate that reasoning to someone else. This is the data that I'm using to make that decision. Or this is why I feel that way. Can you talk a little bit about how to how to communicate or justify prognostic reasoning when a lot of it is kind of the spidey sense I'm nuts.
- I mean, I you know I would welcome. I would maybe open that up to people who are more experienced than me in the room, so happy to hear other comments. I mean? That's not something I have done a lot of, you know, sort of research on. I think that. From what I have read in the literature, I think, being to be honest and then just looking about language. And I was actually just talking to a Med student who was interviewing here the other day about this about just how you know. Sometimes we think that we're speaking plainly, and we forget how much jargon we have developed over time, but interested in your thoughts on that question, communicating pregnants as well.
- Let me let me just add to that, because I think Dr. Dreiter broke it down nicely right system one system, 2 for prognostic reasoning. But I think and she doing the research on this. But what they're finding from my understanding is it's much more system one than system 2. So that's what I think is, that's why you're speaking to how challenging this is. You can come up with all these different scoring systems for prognostic reasons. But it's never going to be individualized to that patient which prognosis is all about that individual patient and their contextual factors and their situation in front of you.
- So I think there's a huge experience component there like it's huge pattern recognition from dealing with this over and over.
- So I mean, that's why we need to study this more. I think it can help us out. But I think that's that's why it's hard and just to to normalize. You know.
- I know that you're an excellent coach, and you're not alone in that in that struggle. But it reminds me of a study. I have seen actually where they they took like 6 months in an Icu, and they gave the physicians this like they gave them data on prognosis based on some patient factors, and they had a nurse that was facilitating family meetings, and they still didn't make an impact on the end of life care that was done. And so it's really hard thank you to it because of that system. One nature of it, too. It's hard to justify, much easier to point to an objective system and say, 13% chance of winning.
- And as clinicians adapt these language learning models into their clinical practice what is your sense of how this is gonna okay their intrinsic capabilities in terms of usage, how we want.

- I think I mean, that's 1 thing I'm worried about one of my points of like do this independently and then consult which I mean, I think that's optimistic. But I think that yeah, I do worry about. And actually, we didn't have time. We ran out of time, but I was going to play a clip from the office, so I don't know if anyone remembers this, but Michael is driving a car, and the GPS. Is telling him to turn right, and he starts turning, and Dwight's like, what are you doing, Michael?
- There's a lake right there, and he's like, No, no, the GPS knows, and he drives right into the lake. So
- I'm gonna be double that. So that question is always the question I think it's the right question to ask, but it assumes that we do this really well, currently right? And then I could like mess this up which is definitely good. But I also think there's huge room to grow. And so I think it's just gonna be different. Skills are gonna become important more important, the shared decision-making. So you'll be translating hopefully more evidence-backed, informed decisions of you and your co-reasoning with AI to a patient and ready to go from there, probably a greater focus on management, probably a greater focus on prognosis.
- I think AI is gonna be much better at I have a theory on the face rape, neglect, issue that present bounce it off. You can see what you think.
- Training, diagnostic, creating a diagnostic hasn't changed since I was an intern 1,000.
- We. So we still train people to Zebra hunt we mentioned. We give them a little bit of information and create an enormous differential suits 2 weeks ago. It's same as when I was super.
- Everyone in this room could list Wilson's disease and the mushroom poison. That's cause acute liver failure. But none of us has seen that one of those things I think that we shouldn't see. It's too broad and therefore base rate neglectis. Why, we're terrible at that is, we're training people to find zebras when almost everything out there is a hoarse what do you guys say about that?
- I totally agree. You're preaching to the choir. I mean, I think part of the you know. Part of the issue is our education system right? We don't know what the students going to go into. So we have to teach you to be able to be a specialist in the future, even if you end up being a journalist.
- So I think that's that's a pardon.
- Did you want to share what was sent out to the scm coaches? The article about sort of the pitfalls of using AI unless you do it in an iterative fashion with that neurologic case that went bad. I don't know if you're oh, I'm not familiar. Actually, I'm sorry. So there's a nice article they got sent out to the sem coaches that made it. The AI was going down the wrong path in terms of misdiagnosing a neurologic stroke case. Sorry it was caught by iterative questions put forth by the person who was prompting, and that you needed to use enough prompts to make sure you weren't going this the second way. And the second question I had is.
- I do some medical legal consulting you, too. Do you think that that's going to change what happens in terms of misdiagnosis that people are using AI. And can I just blame my Lom colleague for missing that diagnosis here.
- I don't know the answer to that. I think it's going to be complicated. I you know it used to be joke about Googling right your symptoms, and we'd have patients come in that have googled it. And now you're hearing, I'm sure, more than me in the outpatient world that patients may be using AI or Chat gpt, but they may be right when they come in with what? And many times that could be. Well, I've gone to 6

doctors now, and now I'm coming in. And finally, Chatgpt was right. And I mean, that might be an extreme circumstance. But yeah, I think we're going to have to have to do that.

- I don't know how that would play out.
- I just want to say one other thing I was thinking about George's question. That's I find, when I mentioned decision theory early on, I find that fascinating because I think as internal medicine doctors, or at least myself, you kind of pride yourself on the ability to be logical and rational, and not that I think of myself as a doctor machine. I'm as a computer or a machine. But like that I can utilize likelihood, ratios, etc, etc, and we just can't. I mean, like we can try. But I don't think our minds work that way. And so I think that's part of why we always struggle with probabilistic reasoning. Even apart from so we really need a machine that can say, no, this is 90%. Still, the likelihood of this diagnosis, even with these atypical features, that the patients.
- So I know you said that you think Llms are gonna be better at diagnosis than we are going forward.
- There's also been some studies looking at Llms and communication, and the patients perceive their empathy as better than when the humans on the other side. So how pessimistic they're trying to be in which this is evolving.
- There's the phone.
- I mean, I I think that the it's it's like, I said, really intriguing. I'm really intrigued, and I'm hopeful that make, you know, impact on outcome errors that for some time. I don't know. I'm really interested. I don't know if anyone's done this study but the empathy question. I really want someone to do a study where you give the same 2 responses, and they know that once from a human is once from an Llm. And do they still feel the same empathy, because that study that you referenced the major difference was that the answers were a lot longer.
- And you know, I sort of wonder if that was the empathy.
- So I think that I don't know what will happen. But I don't think that we will be put out of a job. I think that sort of as exemplified by the bacteria case. I think that there's so much complexity to this that we're going to remain in the game.
- Part of why, I said, diagnosis over the others is my take home from the differences between the diagnosis, study and management is that we can really interpret the contextual factors so much better. And I can. We can be in the room. We can read the room patients, emotions things. So I think that's why we said this kind of co-reasoning. I don't think it's going to be that we lose our jobs. I think it's going to be figuring out. How do we best work together to kind of maximize these situations. And maybe it, you know, maybe it does. Just listen to our interview with the patient and walks out, and it has the diagnosis right? But I think after that the management and the prognosis and all that, the communication with the patient is going to be more complex.
- Hopefully, they're a couple of questions to that performance.
- Ethical reasoning.
- Thank you. I'm not sure I'm know the definition also.
- And then similar to in our previous discussion. This is from Dr. Wolf. I noted recently that AI for a case that involved multiple somatic complaints that did not quote make sense from a biomedical standpoint, but did from a biopsychosocial standpoint. AI provided a list of possible biomedical diagnoses, but no psychiatric diagnoses and examples of behavior. Somatic symptom disorder, depression, panic disorder any

thoughts on the ability of AI to help us in the biopsychosocial realm. For me I guess that it makes me just sort of think about the sort of model that it's trained on, and it's sort of the bias that it's going to have. And so I wonder if that's bias that we're seeing and you know. I think it's sort of exemplified by that by that paper, you know. Bias, in some cases, you know, even concentrated human bias is sort of how I'm thinking about it. So

- I think that's a perfect example of how I think we could work together. Because, right, if you put those symptoms that patient was having into AI. It's gonna come out with purely probabilistic reasoned answer to that right like this is the most likely thing just purely based on these symptoms. But perhaps the doctor, like Dr. Wolf, has known this patient for 30 years, seen them in clinic over 30 years you know, knows their whole family like that's going to put a totally different slant on the interpretation of that information.
- Oh, my God!
- But it still might be better than just him alone, because right? That's forcing him to be to like, acknowledge the base rate of disease. He may be biased in a negative way, not Dr. Wolf interaction, and knowing that patient. So I think there's doing that and then I'll let this be your last question. What are your favorite things you've incorporated from your research and education to your own clinical practice, particularly when practicing alone, without explaining.
- I ready so oh, so mine is kind of. I don't know if you call it. It's like it's 1 of these structured reflection tools where you're kind of thinking the opposite or forcing the opposite. And so I like to do this when I'm on teaching service to. Let's say you're 24 h into an admission, and you've diagnosed you know, Dbt and PE or something, and you've started the management plan. And just to say, What if we found out right now, this is not dbt or PE like? What if we just found. This is not the answer you know, and just like, reflect on that, and and then move forward with like, what would we do differently?
- And I think, thinking that way, I started doing that myself, because most of the time you are right, and it's not that. But when you get into a pattern of doing that all the time. I found that very helpful, because you always have kind of a plan. B. Ready to go. And you even considering plan B, hopefully, sometimes, or sometimes it probably is Plan B, right. But like, I think, that just makes you kind of think in a different way about how it would change, not just the diagnosis, but the management. How would you talk to the patient differently?
- Mine is the way that I try to evaluate new patients, which is to try to do it as independently as possible, and then pull in sort of the ideas of the other team members and doctors so briefly. How I do that is, I just I start with the vitals, and I don't look at any. You know. I sort of go with the objective data. Try to make my own thoughts, and then sort of see what other people are thinking. And I'm excited to kind of start to do that with with Llms as well as kind of another consultant, and then I'll just briefly say, because it's going to be my new favorite and teaching service. Coming in February. I didn't have time to talk about the study, but they did a study of inserting problem representations into Llms. And they found that basically higher quality prop representations where there wasn't a litany of irrelevant past medical history gave you a better differential diagnosis. So now I have data to get on my so.