[There is More to Be Done, Future Possibilities…. Will We Ever Get There?](http://www.ktdrr.org/training/webcasts/webcast10-13/13/index.html)

**Presenter:**

**Dr. Marcel Dijkers**

A webinar aired July 2, 2014, sponsored by SEDL’s Center on

Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

Funded by NIDRR, US Department of Education, PR#H133A120012

Edited transcript for audio/video file on YouTube: <http://youtu.be/zi1XutHk8Vw>

>> Good afternoon. Thank you all for joining us. I'm Joann Starks with the Center on Knowledge Translation for Disability and Rehabilitation Research, or KTDRR at SEDL in Austin, Texas. And I will be moderating today's webinar entitled There is More to Be Done, Future Possibilities Will We Ever Get There?

This is the fourth and final in a series of four webinars focusing on systematic reviews from evidence to recommendations.

Before we begin, I would like to go through some of the Adobe Connect logistics. You should be listening to this presentation through your computer speakers. If you need to turn up the volume, you can do so on your own computer in your audio settings. There's a speaker icon in the bar at the top of the screen, it should be screen and you can adjust the volume with the small arrow next to the speaker icon.

If you have any questions or comments, please type them into the chat box on the left‑hand side of the screen. Marcel or I will address these as needed.

CART captioning is available and the link to the CART is in the useful links pod on the right‑hand side of the screen and also in the chat box. This will open a link in another window.

Finally, there's a pod labeled useful documents. You can download a copy of today's presentations as a PDF or text version. Just select the file name and click on the download button.

Now, let's get started. First I want to thank my colleague Ann Williams for her logistical and technical support for today's session.

This webinar series is offered through SEDL’s Center on KTDRR, which is the funded by the National Institute on Disability, Independent Living, and Rehabilitation Research. KTDRR is sponsoring a community of practice on evidence for disability and rehabilitation research and a series of webinars addressing systematic reviews with a special focus on what is considered evidence and why, and how this evidence is qualified, synthesized and turned into recommendations for clinicians and other practitioners.

Today's presentation is a discussion of future developments and qualifying evidence that might benefit disability and rehabilitation practice. This will guide our community of practice to continue the discussion and we invite all participants to join us.

A final reminder, please use the chat box if you have any questions or comments. Also, at the end of today's session, we'll ask you to complete a brief evaluation form.

Now, it's my pleasure to introduce, Marcel Dijkers, Ph.D., FACRM, research professor in the department of rehabilitation and senior investigator in the brain injury research center at the Icahn School of Medicine at Mount Sinai. Dr. Dijkers is the director two of NIDRR funded projects, the disability and rehabilitation research project on classification and measurement of medical rehabilitation interventions, as well as the Mount Sinai advanced rehabilitation research training project. He's also senior investigator to the New York TBI model system conducted by NIDRR. Mr. Dijkers, please take it away.

>> MARCEL DIJKERS: Welcome back, everybody, for what is the last session. Although, as Joann indicated, we hope to have sparked enough questions to have this develop into a community of practice of various people who want to continue learning about evidence and especially how to look at evidence that might be appropriate for their professional activities.

So our objectives are unchanged from the last time. We want to have a closer look at how is evidence turned into recommendations and where does it come from and where does it go to?

In the first session now already over a month ago, we took a very basic look at the idea of evidence, how systematic reviewing is done and the then already becoming fairly old‑fashioned pyramid of evidence underlining conclusions, and recommendations.

In this second session, we started to look at how the American Academy of Neurology brought more sophistication into that and have spread the issues of how to synthesize information and how to make a recommendation wider than just intervention research.

Then two weeks ago, we looked at GRADE, which at the moment I think is considered the most sophisticated system, which emphasizes the values and preferences of our clients and patients, which in earlier systems not necessarily was done. And gives the decision maker systematic review of the guideline developer, quite some flexibility in grading evidence with, however, the Proviso that you should be very up front as to what your values are and how you look at a particular study and what your reasons are for upgrading and downgrading it.

Now, even GRADE not necessarily has taken care of every problem that we may have with respect to evidence and the use of evidence. Very specifically, not because there are issues quite central to disability research and rehabilitation research, and practice, that are not addressed in GRADE or we have to assume are covered by what they say, but are not directly addressed.

So I will not stop here, but if you have, excuse me, any questions about what I have said or about our objectives, please don't hesitate.

Well, we went over GRADE on July 2nd. There are four levels of evidence, high, moderate, low, very low. As of now, only set up to talk about either intervention research or diagnostic studies and very much focused on the clinical trials, sources, observational studies.

The nice thing about GRADE is that it allows you to downgrade the evidence for a particular study or cross studies based on risk of bias, inconsistency of findings, indirectness, imprecision, publication bias, but also sometimes upgraded because you find big effect size, you find dose‑response relationship, or with observational studies, you take into account confounding that may remain after you have statistically eliminated all the confoundness you have measured. The remaining unmeasured confounding variables would all work to enhance the findings of the study. It's not a basis for upgrading.

Okay. A lot of what I will be talking about today comes from an article that Mark Johnston and I published in 2012, in a supplement of the "Archives of Physical Medicine and Rehabilitation" that was titled "Toward Improved Evidence Standards and Methods for Rehabilitation, Recommendations and Challenges."

We Took More Or Less the Inventory and Some of You Are Ideas and Desires of Where We Ought to Be Going Or should be going. So I can recommend the article. I do not get revenue from sales of articles.

We endorse with the GRADE people that there is a need for objectivity and transparency in creating systematic reviews and guidelines. And the cookbook methods do not work. It's ‑‑ it's always possible to come up with a score of something like PEDro but the simple score and not necessarily gives you a good indication of how strong a study is or isn't. If you just consider the fact that PEDro is based on a sum score which means that two good items make up for two bad items. Well, what if those two bad items are so bad you have to throw away the whole study?

With PEDro, you won't see that, because you just calculate your sum and bingo.

We have ‑‑ that's called the small d issues that you not necessarily can nicely pour into PEDro or something like that. And in order to take a better look at how do those two PEDro items or these four interact with one another and relate to this particular study, in order to come up with decisions that every one of us would agree on or at least decisions every one of us would make the same way, we would need algorithms of hundreds of pages that none of us would like because who wants to apply 100 page algorithm and probably many of us would not necessarily even agree with the algorithm.

 We have to take into account that, yes, we all know that in randomized clinical trial and RCT almost always gives you the best possible evidence, but it's not always available, and then what do you do? .

It would be stupid to just throw it away. Throw away anything that's weak or everything that's from an observational study. There is the issue of the indirect evidence, where we borrow information from another population or another measure than the one we are interested in, and that's a fairly subjective call to make. So there are lots of issues that we want to avoid the cookbook.

But whatever we come up with should not be completely arbitrary. We need common sense and we need transparency. We probably will never get 100% agreement between everybody, but at least if there is transparency is, I understand why you do it. You set forth what your basis for the decision is. I understand that. Yes, that's what we want.

So what are some the recommendations that Mark Johnston and I made in this article? Number one, define the outcomes in terms meaningful and important to the person served. Well, GRADE already does that and I don't think it's much of an issue in disability and rehabilitation research. The outcomes that we study, function, quality of life, are important to people themselves. Nobody as far as I know, wants to sell the blood level a particular chemical and say that's an important outcome. We all want to go to function to happiness, through satisfaction with life, to what have you.

Two, update the technical basis of systematic reviewing by including modern research designs and statistical inference. And there the problem is even GRADE only looks at randomized clinical trials and observational studies, whether those are cohort studies or case control studies, but there are a number of other designs that might be very useful. And could be used to derive evidence.

There's first of all, the N‑of‑1 design and we have come across it in the Oxford rules. Oxford only aims to consider it evidence if the one person on who we have done an N‑of‑1 study is the person to whom you want to apply evidence. So that not necessarily would be very useful in all other situations.

But then there is the regression, discontinuity design and we will look at it quickly in a minute and the federal government and the Department of Education has evidence studies are what works clearinghouse and they, since about a year of knowledge of the validity of the regression‑discontinuity design, providing very hard evidence. There was the interrupted time series design, whether it's randomized or not, and replicated single subject design, yes, whether it's multiple baselines over subject or multiple baseline over outcome you may consider those to be evidence. None of these are AAN or in GRADE.

In GRADE, you might consider that they or you could put them in as very weak evidence and then see whether you can upgrade them.

So a quick look on the regression‑discontinuity design. Ann, will you help me with ‑‑ there is the green arrow.

You do an admit score testing on the XYZ test. You get a distribution of people from 15 to 60. Well, the people who score a less than 35 or weak. The people who are more than 35 are strong. Now you give a treatment to the weak people. The strong people get nothing. That's what we all would like to do as a researcher. Give the treatment to the people who need it most. You give the treatment and then you give them a discharge test, the XYZ test to both groups again.

Now, what you will find is in the no intervention group, everybody will be more or less the same as the score at baseline and you have a regression line here. And if your intervention is effective, everybody in the experimental group, here, will have shifted up and their regression line will run over here.

Well, if that's not strong evidence, I don't know what is.

Interrupted time series design. I made up this example, but it's not too far. Suppose at various times, various states, return to setting the maximum speed limit on the freeways back to 65, or higher.

I live in New York State and we are still officially at 50 miles an hour. So we have traffic deaths per 100,000 people per year for every state. We can plunk in the little Xs when the states starts to implement the higher limit, calculate a mean pre and mean post and you can see pretty much the number of that per 100,000 jump up in every state and that pretty much coincides with the point of increasing the speed limit. Single subject design with baseline over patients four patients, you clip a coin as to when you will give the treatment to them on successive days and person A gets it on day four and B on day six and C on eight, and D on 10 and, again, you can calculate the pre and post and have them here and everybody shows an increase.

Not 100% water tight, especially because a number of cases are fairly small, but it's evidence.

And lastly, you can do that within a certain percent over four outcomes; call them, K, L, M and N here. You give treatment focused on K that starts on day four, on day six you give the treatment for L, day eight for M and day 10 for N, calculate pre/post and you have similar level of evidence.

So in addition here to design, there is also statistical analysis, specifically issues related to statistically eliminating baseline differences between a treated group and a control group based on advanced statistics and I will just here mention here such things as multiple regression using two‑stage least square rather than the usual one stage. There is propensity scoring. There is instrumental variable analysis. Those are all very fancy advanced statistical analysis methods that can do a much better job of statistically addressing baseline and equalities to get a better grasp on does the exposure between group A and group B, between cases and controls make a difference.

Certainly GRADE does not mention whether in case these techniques are used and used well, you have a reason to grade up.

Meta analysis. This is controversial in any of the systems we have looked at. We note that it's quite difficult to apply in rehabilitation research, because generally we have such few studies addressing particular items. And if there are multiple studies, there might still be differences between them based on PICOT, difference in population functional discrepancies at baseline, interventions.

If you do an in‑person CBT and you do a telephone CBT, are you administering the same treatment or are you administering a different one? We, of course, have the problems that our outcome measures are not standardized, although you are familiar with the common data element efforts that presumably will start chipping away at that problem.

And then the time points, how long we give treatment, how early we start after injury onset, how when we do follow‑up are far from standardized. So that makes systematic review difficult if not impossible.

Recommendation number three, evidence grading and recommendations for practice should consider effect size, and direction of biases. Not controversial in GRADE. They already say rate up for effect size. Rate up for the remaining confounding. If the only confounders you can think of point, to a stronger effect, not a weaker one.

Number four, evidence of dose‑response relationships should increase your confidence in study results. Again, that's in GRADE. In disability or rehabilitation research it's hard to define a treatment let alone determine the active ingredients and then quantify its dose. And if you doubt me, I have a Brooklyn Bridge and an article I can send you that pretty much suggests that what we at the moment think as dose number of sessions of PT number of hours of psychology, length of stay in the hospital, are all very poor proxies for dose.

Number five, develop more discriminating methods of grading biases associated with imperfect masking and measurement. At the moment the various systems whether it's GRADE versus AAN, versus oh, What Works Clearinghouse all look at biases and the grading of biases in slightly different ways. For instance when it comes to blinding should you notice it and if you see it, what do you do? And PEDro, you lose three out of 10 points if your study is not blinded.

In AAN, you are downgraded two complete classes. You move from class one to class three, which is just a step above garbage. Well, in our research, blinding often is difficult, if not impossible, which leaves lots of room for biases to play the financial conflict of interest but more important probably, the expectancies of clients, patients, of clinicians and of researchers.

If you get treatment A and the comparative treatment is treatment B, which you consider much better, would you think high of your chance of having good outcomes as to your treatment A? Probably not. So there is our key problem.

Supposedly such biases have no play in the case of objective outcomes and that's why in AAN, when the outcomes you are looking at is death, you don't have to downgrade from class 1 to class 3, for death and any objective, mechanical measurement. However we should be aware that in the research we do, very often the measurement is not completely mechanical. For instance, if you are grading a patient's ADL capability on an instrument that takes time into effect at least to some degree and you do that with a stop watch, well, there's some subjectivity between saying here is where the patient starts doing this ADL, and this is where he has completed it.

So, we through the back door only let the potential for bias back in.

Now Mark and I argued that if we have a blinded assessor, the patient need not be blinded very often, and the clinician need not be blinded but if there's a blinded assessor who doesn't know whether the person he's evaluating is treatment group or comparator group and whether he's grading someone on the pre test or the post test and the test is highly reliable and the blind is not accidentally given away, why is there a need to downgrade this evidence? So we have a bone to pick with AAN.

And when you talk more about not strong measurement in general, some of our scales have not the best psychometric qualities in general, they may be flawed but, generally the bias in them, if there is any bias will go in the same direction in pre test and post test, with a net zero effect, until the there is a reason to expect that it might be different at the low end versus the high end of the scale and presumably we would move the people in the treatment group from low to high while the people in the control group, yes, start low and end low.

And, of course, we should always keep in mind that poor measurement may result in us concluding there is no effect of the treatment while it may be that in reality there is one, but a poor measure wipes it out. So reasons to work on yet a still better measurement.

Number six of the recommendation, and this is more when you get into, well, even in the systematic assessing place, but face systematic reviewing phase, but even later, consider the overall bias and conflict of interest.

Generally people at the moment only look at financial conflict of interest either for the person who does the study. Does he have stock in the drug company whose drug he is testing? Or we look for it if there is a panel put together to do guideline development.

But Mark and I argued maybe we should look at other conflicts too. What about a neuropsychologist is comparing a treatment administered by neuropsychologist with a treatment being administered by physicians? Is there a conflict of interest because if you find out that the treatment administered by the neuropsychologist is inferior it presumably will be less honor for your profession and maybe less income. Do we have a complete lack of interest there?

Or if somebody has studied the particular treatment lifelong, has developed the name of being an expert, et cetera, can we really believe, assume, that this person can be completely unbiased? That he on this next study will find the report truthfully, that what he has spent 20 years on stinks?

We may have some doubt that that is given to us humans.

So Mark and I say we have to do something about that. We cannot eliminate everybody who has conflict of interest because if we did that, probably nobody would be left. We would have no patients. We would have no clinicians, because clinicians took a course in administering this treatment, and it cost them $5,500 and two days. They are conflicted. Not the insurers. They would like to find that it doesn't work or at least equally good as something else, but at a much higher cost. No research, et cetera.

So we need to have panels who have experts who are required to declare their financial and nonfinancial conflict of interest and to the degree possible that the various panel members balance one another out in terms of the conflict.

Recommendation eight, review panels should explicate their reasons for judgment that the particular ‑‑ sorry, reasons for judgment that depart from those indicated by standard a priori criteria. And GRADE already does that, allows you that.

For instance, if you have an RCT, and that, provided evidence standards evidence hierarchy say it's the best possible, but GRADE allows you to say, but, hey, wait a minute. This wasn't blinded or, hey, wait a minute, half the people unblinded themselves. They guessed what group they were in. This evidence is not as strong as we thought. I'm going to downgrade it.

Recommendation nine. Develop and promulgate improved standards and methods for reviewing the quality of evidence for measurements. . Whether we say look at AAN it has standards for intervention studies and screening studies and diagnostic assessment studies and prognosis studies and everything under the sun, but not all are studies that do not come up with yes/no diagnosis, but assessments like very much used by people, clinicians and researchers in disability and rehabilitation.

(telephone ringing).

That's my phone ringing in. I forgot to turn it off. They will go away.

So, we claim that the issues involved in screening/diagnosis are somewhat similar to assessment, but there is still enough difference that it's worthwhile to have separate evidence grading standards for systematic reviews of the quality of class of outcome measures or of specific outcome measure. And I'm shamelessly referring here to a document that I myself was involved in putting together. It's called AQASR, which stands for assessment of the quality and applicability of systematic reviews, and apparently the reference here that I put in is wrong and Ann put a better one on one ‑‑ it's not period, AQASR, but forward slash AQASR.

And as far as we know, there's no evidence that does an evidence based practice or evidence based practice methods has addressed this issue. So if you are concerned with it, go visit AQASR.

Recommendation 10. Explicate criteria for judging generalizability of study results. You pay have noted that all of the evidence hierarchies are based on one dimension of the research only, which is internal validity. External validity, how much can we generalize this finding is completely missing. And it's not an unimportant thing to know.

Now, GRADE has put it on the table in a more or less indirect way by accepting indirect evidence.

If you want to diagnose something in people with traumatic brain injury, there is no evidence for this group but there's other evidence for, say, people with stroke. GRADE will tell you well if you think it's reasonable to generalize from stoke to TBI it's okay, but downgrade the evidence by one step.

Now, that's only so far as a panel of peer reviewers, a panel of systematic reviewers can go to make a decision, can we go from one group of patients to another entire group? But is still the issue that a clinician will have to decide, well, it's fine to say, in general, what holds true for stroke also holds true for TBI, but does it hold true for my particular TBI patient, Mr. Johnson, who is not a typical TBI person?

So generalizability issue looms large and they have not much been addressed at all.

If you look at what people say about it the Cochrane Handbook lists "Factors to Consider" in generalization, which is basically after you do your systematic review, and you want to come up with a statement as to how wide what you find in your systematic review may generalize, they say, well, consider these factors but they do not necessarily tell you how to make decisions and on what basis.

And that may be that for them it's fairly simple because Cochrane pretty much has developed their methods and applied them to drugs. And they may be simple that if you know the diagnosis, the weigh and the age of the person you pretty much can know whether you can generalize and what you should do with the drug dose.

But when you come to the other type of research that you in the audience today either do or use we come to the issue, well, this behavioral approach to resolving problems it worked in client group A. Will it work in client group B? The majority of client group B, a particular member of client group B? What is underlying the ability to benefit from a particular behavioral treatment? And what are the ingredients in a particular treatment?

Can our people respond to those? And how do they potentially affect the ability to benefit.

So we're talking about such fuzzy things in general in judging generalizability in culture and subculture, personality, the ability to learn, remember and apply new information, which that refers to new motor skills or we are learning facts, values, attitudes, et cetera.

We have issues of motivation. We have issues of comorbidities. And then we have to consider not just the physical body, but the places where treatment takes place. What are the referral patterns? What type of patients end up in our ROT outpatient clinic?

What are the resources that we have or that may say more specialized care centers may have that are relevant to administering this treatment? Aside from that, what is the expertise of our clinicians? What do we know of the all important client/patient rapport with the clinician?

With a drug, it doesn't matter if you get along with the therapist. But for cognitive behavior, if you don't like your therapist, it probably is not going to work. So we have big problems to deal with.

On this page is a reference to a FORM approach that was developed. It's called FORM. It was developed in Australia. And it's in the paper we have or you can find it in the forum. I think it's a database ‑‑ on the website now, where they essentially look at generalizability to the target population and applicability to the target context and look at or distinguish four levels of generalizability based on who are the patients and what are their circumstances? So if I go back to here, first row is patient issues and the second row is more or less health system issues and they say the population is the same as the target population, similar to, different, but clinically sensible to applying, difference and hard to judge.

And then, over here, directly applicable in the context of a target, applicable to the local context with a few caveats, with some caveats and not applicable at all.

And, of course, they say more about this. I'm not convinced that this is, all we can say, but it's a start and they have looked at it.

Well, in the interest of time, I will, skip the next two slides that, go more into issues of generalizability from another perspective, the effectiveness trials, the pragmatic trials, the paradox, et cetera.

So I go to recommendation number 11. Choose and develop methods for translating evidence into practice recommendation. It's not so that you ‑‑ at the end of a systematic review, you can, hand over an evidence table and say, here you go, clinician. You work with it. We, need to put it's more or less together in a narrative that takes into account various issues.

The strength of the evidence, the alternative interventions that system might underlie, the findings. The benefits summarized and the common and rare risks, the net benefit of clients ‑‑ to clients, and then the issue of client preference and/or needs. And, all of these need to be considered before clinician, can make a decision on whether to use the treatment or not, and, we say and GRADE says, well, you as the systematic reviewer or you as the guideline developer have to rule this out.

Number 12 recommendation; develop evidence standards and methods for assistive technology devices and services.

We observe that very often AT has an on/off quality. You give some on/off quality. You take it away and it disappears completely. Drugs have that too, in the sense that some drugs, wash out of the body and are gone. Other drugs, start in the healing process and, the effect is remaining.

But take giving a patient with a spinal cord injury a wheelchair. Before you ever give them a wheelchair, he has no mobility, unless he's interested to crawl forward on the floor on his arms. You give him a wheelchair without any training, he can however limited push that wheelchair forward. He's got mobility.

Take the wheelchair away again, mobility disappears. We have an enormous effect size. In how many patients do you need to see that effect size? Do you need to see that in 200 people with spinal cord injury until you believe it or are it enough for you to that it's shown in three patients, and you say, I believe it? This thing, a wheelchair, it works.

So in these circumstances, a large scale RCT seems overkill and we may go with smaller studies, a single subject design with baseline over three or four patients or something like that.

So what are the designs we would accept and for what outcomes would we accept them? My example of giving a wheelchair to a person with spinal cord injury was, of course, an easy one. It is not so easy when you give a communication boards to somebody with a stroke.

You will have to explain to them how the communication board works and all kinds of other stuff, and it's not on and off. It's much more growing into it. So where do we draw the lines in terms of outcomes?

Recommendation number 13, develop a process to synthesize and grade the evidence inherent in clinical experience. Oh, what's going on here? Didn't we start evidence‑based practice to get rid of authority, to get rid of people claiming I have always done it this way. It always worked for me. Didn't we say that we wanted to the facts, and nothing but the facts, ma'am, like Sergeant Friday, I guess?

Yes, it's true but do we want to throw out clinical experience completely? Is there some way we can harvest it, especially if we don't have strong evidence, RCT evidence. If we have only weak evidence from a few single subject design studies that weren't done too well would we make stronger recommendations for the intervention if we tapped the brains of a large group of well‑trained well‑experienced clinicians and got their feedback as to what they believed the value of the intervention was?

Evidence first throughout here, anything here that wasn't research‑based evidence, then in its second stage it acknowledged that it still needs to be applied this concert with patient values and with the expertise of the clinician that needs to apply it yes, they said, no, we don't make you do cookbook medicine, or cookbook practice. But they have not necessarily laid out how individual experience of the clinician interacts with the evidence, yes.

And they have built into or at least GRADE has built into the taking care of the values of average patients, but not necessarily of your next patient.

So the question that Mark and I pose is, is there's a role of the experience of the average or maybe only the expert clinician in creating the body of evidence? Especially if you look to disability and rehabilitation interventions. Generally, individualized, finally type rated applied to patients.

Evaluating them using randomized clinical trials will take.

 first half of eternity. Nobody has time to wait for that.

We have all the designs that are somewhat faster, for instance, practice‑based evidence designs, but they still take quite sometime and quite some money. We have to still do prospective data collection on large numbers of patients, yes, doing measurement of the treatment and the outcomes. Can we speed this up?

We know that we set that ‑‑ what the clinicians report, it was unsystematic subject to primacy, and latency is effect, it's nonquantitative, it's always poor measurement in quotation marks of the outcome, the causal effect that what the clinician did caused outcome is never given to scientific standard and you can probably fill in more here. But can we at least tap all of this knowledge in one way or can we have a serious look at going a little bit further than quantified reporting of experience and do people ‑‑ have people do either retrospective pre‑post studies with standardized outcome measures, which you, for instance, can do if you have FIM data and you have OT relevant outcomes like the first five FIM items here for three years before and four years after you made a major change in how you treated your patients in occupational therapy.

Well, you have a fairly good outcome, you have a standardized outcome and you have routinely collected data. And, of course, we have the potential of doing single subject designs with our patients and replicate those. Both of those possible, but already much more difficult in terms of doing the research and getting the permissions than tapping into the valuable brains of clinicians.

We ‑‑ some guideline development projects, we use the expertise of clinicians and researchers, especially if there is no hard evidence available. We still want to make some recommendation. So very often it goes with a kind of a Delphi process. Draft recommendations are made, people discuss them electronically. Very often it's Delphi, without identifying who says what. They vote. They look at the vote. They see that there is no perfect agreement on the votes. They do more discussion. They vote again, see whether there is more agreement, and at some point in time, when the agreement is ‑‑ 80% of people agree, they say, okay, this is good enough for us. There we use the expertise. So we might ‑‑ you do something, yes, similar to break recommendations and their components. Eliminate the role of authority by doing a blinding or masking of who was saying what of the panel so that the super specialist who has been doing for 50 years does not get a greater voice than a newbie clinician. So that might be a way to go. Can we go further?

Could we think of something with crowd sourcing, go further than the 15 people who serve on a Delphi panel? What would be our criteria? How would we eliminate the hard conflicts of interest if we did crowdsourcing? So there are a number of issues to be discussed, but they may be worthwhile discussing.

Okay. And this far, the recommendations.

And if you have any questions, please write them down. I will take a short break to drink some water and pick it up again.

Let's move on. We already have come upon with this issue that lots of research that we do in disability and rehabilitation is not strong because sometimes we for ethical and other reasons cannot do a randomized and clinical trial. Or if we can, it's very hard to blind the therapist very often, very hard to blind the patients.

Sometimes we can use a blind assessor, but for many of the outcomes we want to do, satisfaction with life, it's no good to have an assessor. We need to get it from the horse's mouth which is the patient, and the patient knows whether he got treatment A or treatment B because we had to tell him in the informed consent document, more or less what would be compared and he could figure out which group he ends up in.

As a consequence, we, our type of research gets dinged in terms of the level of its evidence. I indicated in PEDro, we lose here 10 points for not blinding and AAN, we are bumped down from level one to level three. GRADE we probably will get downgraded one step because the increased bias, et cetera, which makes it very difficult for our evidence to be properly appreciated and used. It's either disregarded or the recommendations for our interventions are weak recommendations and sometimes payers may even refuse to pay for our services because there is no class evidence that your treatment worked, Mr. Rehabilitationist! And if there is no class evidence that it works we are not going to pay for it.

So there is a committee in the ACRM, now called the evidence and practice committee, that discussed, okay, under what circumstances would we consider the type of RCT evidence that comes out of disability and rehabilitation research, with subjective outcomes as, AAN level II or even level I.

 We very much has worked on a number of our systematic reviews have used the AAN system. But we always bump up to, well, the strongest evidence is pretty much level III. What would we expect to bump it up one steps or two steps?

So we discussed this and came up with a fairly lengthy list. Number one, we would ‑‑ the study would need to compare two concurrent treatments or an face valid treatment. So professional may know here that it's not really a treatment. It's placebo, but to a patient, it looks like a treatment. Or you compare two active treatments.

B, the only reason you are not blinding is because the outcome cannot be blinded. Don't give me blinding because it's so much easier to measure outcomes non‑blinded. Only when there's absolutely to way you can blind the other therapist or you can blind the patient, client I might accept upgrade your evidence.

Number C, the study contains rigorous methods to create equal expectations of outcomes across the non‑blinded conditions.

Number one, nothing does the therapist, the researcher said would give the patient to believe that one of these treatments is better than the other. And we queried the patients. The independent person queries the patients as to what they believe before treatments start to treatments are and how one will be better than another ‑‑ than the other.

So not only should nobody say anything to bias the patients in a particular direction, but the patients even out of their own reasoning should not here, in general or in majority give the preference to one treatment over the other.

And, of course, then the treatment should not be one that trains you to say particular things. In cognitive behavioral treatments or to think particular things. Think positive thoughts. Well, the outcome, of course, will be positive thoughts. So that's not necessarily a good study to consider for upgrading.

D. Because of the inherent subjective nature, the subjective patient reported outcome PRO, is the best measure of the construct of interest. Satisfaction of life, satisfaction with life is properly measured with PRO.

Ability to put one's pants on can be served by a therapist and graded and timed, et cetera, and if you as a researcher take a short cut and ask the patient to report whether putting on his pants is easy, moderately difficult, or very difficult, that's a short cut. It could be blinded if you desire not to do it, we don't need you.

Criterion E, administering the subjective PRO uses rigorous methods to minimize bias. Same measurement procedure for both conditions, . Standardized administration and scoring procedures that contain no biasing instructions. And the assessors should be trained in the procedure, blind to treatment group assignments and have no conflictive interest. They should not be part of the routine study team.

Criterion F, the subjective PRO demonstrates evidence of reliability and evidence of high reliability either high test, retest reliability within a very short period of time or confirmative evidence high correlation with an outcome that is expected to correlate strongly with it.

So strong measurements. Criterion G, the investigators who have conflict of interest should stay far away. They cannot personally deliver treatment. They cannot collect any of the PRO outcomes. They cannot be involved in data analysis. There should be an independent statistician here. As long as the final results have not yet been determined they should be remaining blinded to treatment conditions, had which is a steep rule.

Now, we note that the methods used to address criteria C to G are reported in the manuscript or in the related documents and they are reported in a positive way. We do not want to be in a position where the people who are going to do the potentially grading up say, well, it didn't say here whether the person with conflict of interest, had the investigator was blinded to analyses or not, but let's assume so. We think that's how they always do it in Philadelphia.

Nope. It either is said so or it doesn't exist.

So, to summarize, we came up with rules for what you should satisfy for grading up to a class II in AAN and what you needed to do to upgrade to a class I and basically everything I have gone over this far is needed for a class I.

For a class II, the only thing that can be missing is here ‑‑ Ann, help me.

Okay. This here, C2, the independent assessment by people without conflict of interest here that ‑‑ here, that the individuals were not biased in any ways by the ‑‑ what was going on in the study.

Otherwise, everything needs a check mark.

The same for criteria E through F, this over here, the first one looks like it means no, but this is just the active ingredients. Here are the things you have to talk about, the yellow is just an introduction. And the same over here.

So these are very high standards. Have you ever seen a study that satisfied all these standards are? I haven't. Just because I could, I was working with some post‑docs on a systematic review, and we looked at, for the hell of it, at 31 studies that dealt with treatment for anxiety after TBI. And we looked at here which one of these studies satisfied criterion A, B, C, C1, C2a, C2b, through G4. Out of the 31 here, 22 times we decided it clearly wasn't done. No upgrading for you.

Nine times, we said it clearly was done. They reported it, and it was satisfactory. We normally or you heard me say, if it's not reported, it's not done, but for this particular study, we separated out clearly not done, clearly not the case in the yellow column versus separately, they didn't even say anything about this. And then we had a few cases where it was clearly not applicable. So you see that even if you take into account some of the nonapplicable here, C2b was 0 times satisfied, . We are far from satisfying any or all of these criteria that according to this sample will allow us to upgrade from AAN class III to either class II or class I. And we didn't expect to find any. We wanted more to get an initial idea how far are we from this level.

And we ‑‑ when we are talking about the members of the evidence and practice committee, we are talking with the people at AAN to discuss these issues, and potentially fold this into their manual so that more studies that are being evaluated using AAN criteria might be upgraded. So that would be a nice success for this committee.

Any questions?

Okay. We are almost at the finish.

All of these issues, whether I refer to the recommendations that Mark and I did, the recommendations that the evidence and practice committee did, not necessarily have been adopted by anybody let alone by everybody. Or even discussed by everybody, but we need to have to have a serious discussion about, well, originally, we had the second evidence pyramid which was simplistic and just for evidence on treatments. We have become more sophisticated. We are addressing now additional questions, additional to treatment. We have with great more flexibility in terms of looking at stuff, but we're not there yet. For instance, what about qualitative research, and the systematic review of qualitative research? Does anybody accept that as evidence?

In the first session, somebody had a question about mixed methods research. Whether the mixing is triangulation of quantitative and qualitative or the qualitative is better or it's what they call sequential, exploratory or sequential explanatory. I'm not aware, in answer to your question, of anybody who has developed evidence for mixed methods research.

Nor are there criteria for how to do systematic reviews of mixed method research or how to address systematic review that God forbid, incorporates both qualitative and quantitative primary study, in addition to mixed methods. If we do that what do we mix, apples and oranges? Do we get the fruit salad? Very nice. Very nice or is it an undifferentiated flavorless mush and nobody knows what to do with this?

As I indicated before, evidence‑based practice, evidence‑based medicine is always focused on the individual patient. And always assume that delivering the evidence and into the hands of clinician is enough. We are becoming more and more aware that the researchers cannot just put evidence in a nice package, put it in the inbox of the clinician, and presume that now it will be implemented, period.

We are becoming more and more aware that we have to have evidence translation or, sorry, knowledge translation that we researchers and maybe specialists in KT need to help clinicians to understand the evidence, evaluate the evidence, implement ‑‑ evaluate whether they and their organization are ready to start using it, and then prepare themselves and their organization for actual start of use.

And this whole issue of KT is developing into a science, which, of course, has a need for its own systematic reviews that integrates the various primary studies. .

So that we have systematic reviews of what's the best way to deliver new knowledge and new technology? And what's the best way to make sure that it gets used systematically. So we do not really have a KT ‑‑ sorry, an evidence‑based practice in this area, and issues of guidelines developments.

And in doing all of this, we are widening out, broadening out from even what GRADE has starting here to do. Some of these issues are being considered elsewhere. In London, there's what's called an EPPI‑Centre which deals with evidence for social interventions, social work interventions, that kind of stuff. They have a set of principles that I cite here from Hansen, and Hansen was one of the earlier slides two weeks ago, or four weeks ago.

Where they have what I would call, and both primary research and reviews of evidence ‑‑ reviews of research essential to progress. There's a range of primary research methods. There's a wide range of review methods qualitative and quantitative research, et cetera. Reviews should follow the research principles of quality, rigor and accountability. Similar to how these are used in primary research. GRADE does that for are particular research.

Review methods often reflect the assumptions, the method logical assumptions and reviews should be driven by questions which vary in many ways, including theoretical and ideological perspective.

Users of research have particular perspective of priorities that can usefully inform primary research and reviews of research. More or less, the last one is the end of the K2 loop or the beginning we should be doing research focused on questions that clinicians are answering.

So all in all, this broadens here, the map of more evidence‑based practice needs to be or the recommendations need to be. And it goes into such things as, well, Mr. Clinician, what evidence do you need, consider, practitioner? Do you need the traditional application of systematic reviews? Do you need evidence on the implementation of the evidence, what I refer to with evidence on KT methods? Do you need attitudinal evidence on what we know that clinicians worldwide think about this?

It's nice to want to implement something in your organization, but if there's evidence that clinicians don't like it, it probably is going to be a failure.

Do you neat economic evidence and some of that is being collected in reviews?

Do you need ethical evidence? What are the values that patients have or maybe clinicians have and what does that say?

Well, I told you that nobody gets there and then, well, not too long before I started preparing my slides for this series, what do I get? News bulletin from Cochrane that they have publish a review of qualitative evidence here. It's not about the effectiveness of health interventions or the accuracy of diagnostic tests traditional areas of Cochrane. It's about barriers to and facilitators of implementation of lay healthcare workers.

It talks about the use of lay health care workers accepted around the world. It was in combination with the outcome of the traditional Cochrane review on the effectiveness of lay healthcare workers in community and maternal child health.

And by combining the existing systematic review, with the brand new qualitative review on barriers of facilitators, they could put that together and come up with a set of recommendations, which ‑‑ sorry, Cochrane didn't do itself but it was done by a branch working for the World Health Organization.

So new things are happening here. Whoa! Cochrane has gone to systematic review of qualitative research. Who would have thought? .

And they are proud on the fact that the qualitative research was used in combination with the traditional quantitative intervention study. .

So we're already moving into some of the claims that people have made, evidence of effectiveness is not enough. We need evidence on the process that serves as delivery, the appropriateness of care, the acceptability of what we proposed to do, the satisfaction with care as the last four, five, six slides have indicated we have to start developing evidence for or methods for qualifying evidence for much broader questions than we traditionally have looked at.

So one could look at the Johnston/Dijkers recommendations, that I brought up, the ACRM, and the improvement of AAN, the grading and of qualitative of evidence and qualitative and mixed‑methods studies. Reviews of or incorporating into phrasing of recommendations, qualitative, quantitative and mixed‑methods evidence. And one could start looking at the role of evidence in the larger context and develop that evidence for KT.

There is lots of work to be done, and my question is, well, would is ready to start the discussion? Who is available to start in participating in your community of practice, that SEDL very much would like to start. So this is what I have to present for today. I'm ‑‑ even though we are at the end of our time, I am staying around for a few more minutes in case there is any questions from any of you. If not, you will be invited ‑‑ you have been invited to evaluate the ‑‑ sorry, to complete the evaluation form of today's session and probably it goes back to earlier sessions.

This is the end at least for today. If you would think that you are not necessarily ready for participating in a community of practice, but would more of the type of presentations I have had over the four last sessions, we can develop those SEDL and the grant that NIDRR is funding, consider this worth while information and work file methods to develop and they will support it.

So thank you from me, for all of you, and I hope we will rub shoulders in evidence land. Thank you all.

>> JOANN STARKS: Well, thank you, very much, Marcel, for sharing all of this information. The sessions have all been very interesting, and before we conclude, I also want to emphasize what you said, if anyone has any questions, please feel free to type them in there, in the chat box, and we would be happy to have Marcel respond.

I want to thank Marcel and also everyone that participated in today's webinar. As he mentioned, we do have the brief evaluation form. The link is there at the bottom of the last PowerPoint slide and it's also in the chat box and we will be sending it out in an email to everyone.

And a reminder that we do have the archives posted, if you missed any of the previous sessions and today's session will be archived probably in about a week or so. We do appreciate the support of NIDRR to carry out these webinars and our other KTDRR activities. We will be following up soon with more from the community of practice on evidence in disability and rehabilitation research, with discussion via email and also more teleconferences if needed. We hope through these discussions to be able to identify some future topics for future webcasts.

So thank you again for joining us today and good afternoon.