*Systematic reviews: From evidence to recommendation*

**There is more to be done: Future possibilities….will we ever get there?**

Webinar Series - Part 4

Presented by Marcel Dijkers

July 16, 2014

Text version of PowerPoint™ presentation for webinar sponsored by SEDL’s KTDRR. More information about the webinar can be found here:

[https://www.ktdrr.org/training/webcasts/webcast10-13/13/index.html](http://www.ktdrr.org/training/webcasts/webcast10-13/13/index.html)

Title Slide 0:

Systematic reviews: From evidence to recommendation. Session 4 – July 16, 2014. There is more to be done: Future possibilities….will we ever get there?

Marcel Dijkers, PhD, FACRM. Icahn School of Medicine at Mount Sinai.

An online webinar sponsored by SEDL’s Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR). Funded by NIDRR, US Department of Education, PR# H133A120012. Copyright © 2014 by SEDL.

Slide 1: Objectives:

Discuss, within the context of systematic reviews

* what is considered evidence and why
* how evidence is qualified and synthesized
* how evidence is turned into recommendations for clinicians and other practitioners

Slide 2: Topics:

1. Overview of the process and tools of systematic reviewing, with a focus on assessment and synthesis of evidence, and the idea of a research design-based pyramid of evidence underlying conclusions and recommendations
2. How the American Academy of Neurology and others have brought in research design details and quality of research implementation in grading evidence, and have gone beyond intervention research
3. The GRADE approach, with its emphasis on the values and preferences of patients/clients, and flexibility in grading evidence: fit with disability and rehabilitation research (in red)
4. **A discussion of future developments in methods of qualifying and synthesizing evidence that might benefit disability/rehabilitation practice**

Slide 3: Questions?

Slide 4: July 2 topics:

* Addresses systematic reviews and guidelines, of Tx and Dx only
* Outcome-oriented synthesis of evidence, with strong emphasis on prioritizing outcomes based on patient/client values
* Few algorithms, focus is on making underlying values/reasoning explicit for decisions taken transparent
* Based on RCT vs observational study (and even weaker evidence??); four evidence qualities distinguished: high, moderate, low, very low

Slide 5: July 2 topics:

* Initial quality rating can be downgraded for:
	+ Risk of bias (‘small d’ issues)
	+ Inconsistency of findings of studies
	+ Indirectness (population, intervention, outcomes)
	+ Imprecision (wide pooled studies’ confidence interval)
	+ Publication bias
* Initial quality rating can be upgraded for:
	+ Large effect size
	+ Evidence of a dose-response relationship
	+ Plausible confounding: remaining confounding does not eliminate effects found, may even strengthen them
* Only weak vs strong recommendations

Slide 6:

Center: Archives of Physical Medicine and Rehabilitation
Journal Homepage: <http://www.archives-pmr.org>

Arch Phys Med Rehabil 2012:93 (8 Suppl 2):S158-99

Right: Thumbnail of journal cover
Below a line: Special Communication. Toward Improved Evidence Standards and Methods for Rehabilitation: Recommendations and Challenges. Mark V. Johnston, PhD, Marcel P. DIjkers, PhD.

From "Toward improved evidence standards and methods for rehabilitation: recommendations and challenges," by M. V. Johnston and M. P. Dijkers, 2012, *Archives of Physical Medicine and Rehabilitation, 93*(8 Suppl 2), S185-189. Retrieved from [www.archives-pmr.org/article/S0003-9993(11)01142-7/pdf](http://www.archives-pmr.org/article/S0003-9993%2811%2901145-2/pdf). Reprinted by Marcel Dijkers in compliance with Elsevier’s author rights.

Slide 7: Need for objectivity and transparency in creating systematic reviews and guidelines

* ‘Cookbook’ methods do not work:
	+ ’small d’ issues have varying and interacting effects; a complete algorithm would be hundreds of pages and be disapproved by 99% and disliked by 100%
	+ Weak (non-RCT) evidence is not the optimal basis for recommendations, but it would be stupid not to use it
	+ Validity of indirect evidence always is a judgment call, but it would be stupid not to use it if it is needed
	+ Etc.
* We need to have methods that allow application of common sense
* As long as there is transparency: what ‘subjective’ decisions are taken, and why, by whom, should be made explicit

Slide 8: 1. Define outcomes in terms meaningful and important to the persons served

* GRADE emphasis
* Not much of an issue in Disability and Rehabilitation research: function, quality of life are our primary outcomes

Slide 9: 2. Update the technical basis of SR by including modern research designs and statistical inference

* Include among strong designs:
	+ N-of-1 design (for patient involved): Oxford CEBM
	+ Regression-discontinuity design (What Works Clearinghouse)
	+ (Randomized) interrupted time series design
	+ Replicated single subject design (esp. for AT and similar interventions) with large effect size (What Works Clearinghouse)
		- Multiple baseline over subjects
		- Multiple baseline over outcomes
	+ Not in AAN, GRADE
	+ In GRADE, not explicitly excluded either

Slide 10: Regression discontinuity design

Y axis is “Discharge score on XYZ test” with a range of 0 to 60.
X axis is “Admission score on XYZ test” with a range from 10 to 60.
There is a line in the middle of the X axis (at value 35), which an arrow designates as representing the cut point. To the left of this vertical line is the Experimental intervention group and to the right is the No intervention group. Two regression lines are plotted, separately for the Experimental intervention group and the No intervention group. The regression line for the latter coincides with the main diagonal, suggesting that admission and discharge XYZ scores are very similar for this group. The extension of the regression line of the No Intervention group suggests where the experimental intervention group would have been without the intervention. In actuality, the regression line for this group is displaced upward, such that on the Discharge XYX test the average “Experimental intervention group” member scores as high as the average “No intervention group” member.

Slide 11: Interrupted time series design

Table title: Traffic deaths per 1,000,000 miles in the year

Table with 4 rows and 23 columns. The columns are labeled state, 01 through 20, mean pre, and mean post. The states identified in the first column are Alabama, Colorado, Delaware, and Florida. There are 20 columns per state indicating year 01 to year 20. In each row an x indicates the year of increase of maximum speed limit to 65 miles/hour or higher: Alabama 05, Colorado 07, Delaware 11, Florida 15.

Alabama: 01=4, 02=6, 03=4, 04=4, 05=x, 06=6, 07=4, 08=7, 09=5, 10=7, 11=8, 12=6, 13=8, 14=9, 15=8, 16=5, 17=9,18=9, 19=7, 20-8. The mean pretest score for Alabama is 4.3 and the mean post-test is 7.1

Colorado: 01=3, 02=4, 03=3, 04=5, 05=3, 06=2, 07=x, 08=4, 09=5, 10=8, 11=6, 12=7, 13=5, 14=7, 15=8, 16=6, 17=6,18=6, 19=5, 20=7. The mean pre score for Colorado is 3.3 and the mean post score is 6.2

Delaware: 01=6, 02=7, 03=6, 04=8, 05=7, 06=6, 07=5, 08=6, 09=7, 10=8, 11=x, 12=8, 13=9, 14=10, 15=10, 16=11, 17=9,18=8, 19=9, 20=10. The mean pre score for Delaware is 6.6 and the mean post score is 9.3

Florida: 01=6, 02=5, 03=7, 04=8, 05=5, 06=6, 07=7, 08=5, 09=6, 10=7, 11=6, 12=5, 13=6, 14=6, 15=x, 16=9, 17=8,18=9, 19=10, 20=10. The mean pre score for Florida is 6.1 and the mean post score is 9.2

The mean pretest score for Alabama is 4.3, Colorado is 3.3, Delaware is 6.6, and Florida is 6.1.

The mean post test score for Alabama 7.1, Colorado is 6.2, Delaware is 9.3, and Florida is 9.2.

Slide 12: Single subject design with multiple baseline

There are two tables on this slide the first representing Over Subjects and the Second representing Over Outcomes.

Table 1, Over subjects. This table shows the score on outcome measure on successive days, from day 1 to day 11, for patients A, B, C, and D, who constitute the rows.

Patient A scores are 4, 6, 3, x, 6, 4, 7, and 8. Days 9-11 are left blank. The mean pre is 4.3 and the mean post is 6.3.

Patient B scores are blank for days 1 and 2 and then are 3, 3, 2, x, 4, 6, 7 and blank again for days 10 and 11. The mean pre is 3.0 and the mean post is 6.2.

Patient C scores are blank for days 1 – 4 and then are 6, 5 ,6, x, 8, 10, 11. The mean pre is 6.1 and mean post is 9.7.

Patient D scores are blank for days 1 – 5 and then are 7, 5, 6, 5, x, 9. The mean pre is 5.8 and mean post is 9.0.

Table 2: Over outcomes. This table shows the score on outcome measure X (with X standing for four measures: K, L, M and N, which constitute the rows) on successive days, numbered in the stub from 1 to 11 for outcomes K-N.

Outcome K scores are 4, 6, 3, x, 6, 4 and days 7-11 are left blank. The mean pre is 4.3 and the mean post is 6.3.

Outcome L scores are blank for days 1 and 2 and then are 3, 3, 2, x, 4, 6 and blank again for days 9-11. The mean pre is 3.0 and the mean post is 6.2.

Outcome M scores are blank for days 1 -3 then are 7, 6, 5, 6, x, 8, 10, 11. The mean pre is 6.1 and the mean post is 9.7.

Outcome N is blank for days 1-7 and then 6, 5, x, 9. The mean pre is 5.8 and mean post is 9.0.

Slide 13: 2d. Incorporate the best of current methodological knowledge in grading observational cohort studies

* Take into account use of appropriate statistical techniques to eliminate prognostic imbalances
	+ Multiple regression using two-stage least square regression
	+ Propensity scoring
	+ Instrumental variable analysis
* Not in AAN, GRADE
* GRADE does not mention rating an observational study up if these techniques are used

Slide 14: 2e. Perform meta-analysis when there are several comparable, high quality studies

* GRADE, AAN, Oxford CEBM: not controversial
* More difficult to apply in disability and rehabilitation because:
	+ Small numbers of studies
	+ Discrepancies between studies in PICOT:
		- Population: how much difference does minor functional discrepancy make?
		- Interventions used are not like drugs: is CBT flavor 1 same as CBT flavor 2?
		- Outcome measures not standardized (CDEs?)
		- Time points, esp. of follow-up, not standardized

Slide 15: 3. Evidence grading and recommendations for practice should consider effect size and direction of biases

* GRADE
	+ Rate up for effect size
	+ Rate up for remaining confounding pointing to a stronger effect, not a weaker one

Slide 16: 4. Evidence of dose-response relationships should increase confidence in study results

* GRADE
* In rehabilitation and disability research it is hard to define the treatment, let alone determine the active ingredient and quantify its dose
* LOS, number of sessions, number of hours of therapies all are poor proxies for dose

Slide 17: 5. Develop more discriminating methods of grading biases associated with imperfect masking and measurement

* There are inconsistencies between systems whether blinding problems are noted, and if so, what the consequence is: lower quality score (PEDro: 3/10 points for not blinding; AAN: two classes less for not blinding)
* In rehabilitation and disability research, blinding is difficult, if not impossible
* Which leaves room for lots of biases to play:
	+ Financial conflict of interest
	+ Researcher, clinician, patient expectancies

Slide 18: 5. Develop more discriminating methods of grading biases associated with imperfect masking and measurement

* Supposedly, such biases have no play in case of ‘objective’ outcomes:
	+ Death
	+ Any ‘mechanical’ measurement
* However, whenever ‘mechanical measurement’ requires human judgment (e.g. when to start and stop the stopwatch for a timed ADL), there is room for bias
* On the other hand, if a blinded assessor (who doesn’t know whether the person to be assessed is in pre-test or post-test, experimental or control group) administers a highly reliable test and the blind is not broken – why is there a need to downgrade the evidence?

Slide 19: 5. Develop more discriminating methods of grading biases associated with imperfect masking and measurement

* Flawed measurement generally will have same bias in pre-test and post-test, or in treatment and control group, with a zero net effect (unless bias is different at the low vs high end of the scale)
* Poor measurement (low reliability and validity) may result in:
	+ Not observing effects where they exist, thus concluding to ‘no difference’ between Tx and comparator when in reality there is a difference

Slide 20: 6. Consider overall bias and conflict of interest

* Financial conflict of interest typically is the only one reported in the primary paper, and (we hope) is considered in putting together a guideline panel (IOM standards)
* But should other conflicts be explored?
	+ Comparison of a treatment administered by one’s own profession (neuropsychology?) with that administered by another profession (medicine?)
	+ A lifelong investment in studying a particular treatment, clearly expressed in a few non-systematic reviews that hardly acknowledged, let alone appreciated, alternative treatments

Slide 21: 7. Establish requirements to ensure expertise and minimize bias of review panels

* If we eliminated all persons with any COI (financial, intellectual, other) expertise from a review panel, the panel would be empty:
	+ No patients
	+ No providers
	+ No insurers
	+ No researchers
	+ Etc.
* What we need to do is have panels with experts who
	+ Are required to declare their financial and non-financial COIs
	+ Are balanced in terms of the conflicts that exist

Slide 22: 8. Review panels should explicate their reasons for judgments that depart from those indicated by standard a priori criteria

* + GRADE

Slide 23: 9. Develop and promulgate improved standards and methods for reviewing quality of evidence for measurement

* While the issues involved in screening/diagnosis are somewhat similar, assessment is different enough that it is worthwhile to have separate evidence grading standards (cf. [shameless commerce division]
* AQASR: *Assessment of the Quality and Applicability of Systematic Reviews* [ <https://www.ktdrr.org/ktlibrary/aqasr> ]
* Disability and rehabilitation researchers should be especially interested
* No EBP organization has focused on this – not even the Campbell Collaboration

Slide 24: 10. Explicate criteria for judging generalizability of study results

* EBP evidence hierarchies are based on one dimension only: internal validity
* External validity is missing in action
* GRADE has put it on the table with accepting ‘indirect’ evidence
* A panel can only go so far – deciding whether a treatment that has been shown in several studies to have benefit for ‘the average person’ in population A (NNT 4.1), is also expected to benefit the average member of population B
* The clinician still has to decide whether his/her next patient/client is close enough to that ‘population A average’ to be likely to have benefit (more on this later)

Slide 25: 10. Explicate criteria for judging generalizability of study results

* Cochrane handbook lists ‘factors to consider’ in generalization, but does not spell out how and on what bases to make a decision
* For pharmaceutical treatment decisions, diagnosis, comorbidities, weight and age may be all that is needed to decide
* What is the basis in disability and rehabilitation treatment to decide that a behavioral approach that works with client group A will work with the majority of/a particular member of client group B?
* What are the ingredients in a D&R treatment, and what patient/client characteristics make deployment of or effect of these ingredients impossible?

Slide 26: 10. Explicate criteria for judging generalizability of study results

* Patient issues to consider
	+ Culture and subculture
	+ Personality
	+ Ability to learn, remember and apply new information
		- Motor skills
		- Facts, values and attitudes
	+ Motivation
	+ (Co-)morbidities
	+ Health system issues to consider
	+ Referral patterns
	+ Resources at 1⁰, 2⁰, 3⁰ care centers
	+ Expertise of clinicians
	+ Patient/client-clinician rapport

Slide 27: FORM approach: generalizability to patient population and health care/other context

Table with 2 rows and 4 columns. The columns are A (Excellent), B (Good), C (Satisfactory), and D (Poor).

Row 1 describes the component of generalizability to target audience. A (Excellent) Population studied is the same as the target population, B (Good) Population studied is similar to the target population, C (Satisfactory) Population studied is different but it is clinically sensible to apply this evidence to the target population, and D (Poor) Population studied is different and it is hard to judge whether it is sensible to generalize.

Row 2 describes the component of applicability to target context A (Excellent) Evidence is directly applicable to the context of the target population, B (Good) Evidence is applicable to the local context with few caveats, C (Satisfactory) Evidence is probably applicable . . . with some caveats, and D (Poor) Not applicable to local context.

From Table 2 in "Toward improved evidence standards and methods for rehabilitation: recommendations and challenges," by M. V. Johnston and M. P. Dijkers, 2012, *Archives of Physical Medicine and Rehabilitation, 93*(8 Suppl 2), S185-189. Retrieved from [www.archives-pmr.org/article/S0003-9993(11)01142-7/pdf](http://www.archives-pmr.org/article/S0003-9993%2811%2901145-2/pdf). Reprinted by Marcel Dijkers in compliance with Elsevier’s author rights.

Slide 28: The call for pragmatic trials (effectiveness trials)

Proposed criteria to distinguish effectiveness from efficacy trials (Gartlehner et al., 2006)

* 1. Populations in primary care (*rather than tertiary care*)
	2. Less stringent eligibility criteria (*rather than the usual very restricting RCT criteria*)
	3. Health outcomes (*rather than proxies such a serum uptake or impairment level outcomes*)
	4. Long study duration; clinically relevant treatment modalities (*rather than a pre-post study with academia-only treatments*)
	5. Assessment of adverse events
	6. Adequate sample size to assess a minimally important difference from a patient perspective
	7. Intent-to-treat analysis

From Table 1, p. 5, *Criteria for distinguishing effectiveness from efficacy trials in systematic reviews* by G. Gartlehner et al., 2006. Technical Review 12 AHRQ Publication No. 06-0046. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK44029/pdf/TOC.pdf>. This document is in the public domain and may be used and reprinted without permission.

Slide 29: The paradox of generalizability

The figure has two square panels. Panel 1 is on the left and Panel 2 is in the center. Each panel has Dimension Y on the left hand side and Dimension X on the bottom of the square. A key on the right identifies the nature of five circles within each panel: Patient population, Efficacy sample, Effectiveness sample, Clinician A patients, and Clinician B patients.

Within Panel 1, the patient population is represented by a large circle. A much smaller circle in the upper left part of the large circle represents an efficacy study sample. An even smaller circle partially overlapping that smaller circle shows some of Clinician B's patients represented within the efficacy sample. (But some of B's patients are outside the efficacy sample). In the bottom right of the large circle, a smaller circle shows Clinician A's patients as part of the patient population; none of A's patients are within the efficacy study sample.

Within Panel 2, the patient population is also represented by a large circle. A fairly large circle in the upper left part of this large circle represents a large effectiveness study sample. A much smaller circle that falls entirely within it shows Clinician B's patients, all completely within the effectiveness sample circle. In the bottom right of the largest circle, a smaller circle partially overlaps the effectiveness sample and shows most of Clinician A's patients represented within the effectiveness sample, but a few falling outside it.

From: Figure 2 in Dijkers, M. P. J. M. (2011). External validity in research on rehabilitative interventions: Issues for knowledge translation. *FOCUS Technical Brief (33).* Austin, TX: SEDL, National Center for the Dissemination of Disability Research.

Slide 30: 11. Choose and develop methods for translating evidence into practice recommendations

* Practice recommendations should consider, at a minimum:
	+ Strength of evidence
	+ Alternative interventions (comparator[s])
	+ Benefits and (common and rare) risks
	+ Net benefit to clients
	+ Client preferences or needs
* GRADE has put these on the table

Slide 31: 12. Develop evidence standards and methods for Assistive Technology devices and services

* AT very often different from pharmaceutical, behavioral and other rehabilitation and disability interventions:
	+ on/off quality: effect is immediate (even if it may further increase with practice)
	+ Effect size is often very large
	+ Eliminating AT undoes the effect
* In these circumstances, a large-scale (pragmatic) RCT seems overkill
* What are the designs we would accept instead, and to which ATs and AT outcomes would these standards apply and not apply?

Slide 32: 13. Develop a process to synthesize and grade the evidence inherent in clinical experience

* EBM (and EBP) was developed to reduce/eliminate the role of lore, tradition, authority, etc. in health care decision making, and instead base it on hard facts: the evidence of clinical research: well-performed, varied, large samples, well integrated
* Only in its second stage did EBM acknowledge that evidence needs to be applied in consort with patient values and clinical experience (no ‘cookbook’ medicine/practice)
* HOW this is to be done in a systematic fashion has not yet been worked out to any large degree (Dijkers et al., 2012)
* Values of average patient can be built into evidence qualifying process (e.g. GRADE)

Slide 33: 13. Develop a process to synthesize and grade the evidence inherent in clinical experience

* Is there a role for the experience of the average or expert clinician in the body of evidence?
* D&R interventions consist of finely titrated combinations of large numbers of therapeutic activities individualized for delivery to patients/clients
* Using RCT designs to evaluate every permutation and combination is impossible
* Other designs (e.g. Practice-Based Evidence – PBE) have been proposed
* PBE is still an expensive and slow process: prospective collection of mass quantities of broad patient, treatment and outcome data

Slide 34: 13. Develop a process to synthesize and grade the evidence inherent in clinical experience

* Clinician-reported treatment effects were and are distrusted:
	+ Unsystematic
	+ Primacy and latency effects
	+ Non-quantitative
	+ Poor outcome ‘measurement’
	+ Causal effect not proven to scientific standard
	+ Etc.
* Alternative (all involving much larger clinician effort):
	+ Retrospective pre-post studies with standardized outcome measures (e.g. analysis of FIM data in eRehabData/UDS - https://web2.erehabdata.com/erehabdata/index.jsp)
	+ Prospective single-subject designs, replicated

For additional information, see the Special Issue: Single-case Experimental Design Methodology, *Neuropsychological Rehabilitation*, 2014, Vol 24, Issue 3-4

Slide 35: 13. Develop a process to synthesize and grade the evidence inherent in clinical experience

* Wherever a guideline development process uses ‘expert consensus’ to replace or supplement research evidence, the hundreds of unrecorded unsystematic N-of-1 studies these experts have done is ‘summarized’
* Delphi process may be used to
	+ Break recommendations into components (effects on various outcomes; adverse effects; role of various comparators, etc.)
	+ Eliminate the role of ‘authority’, with everyone on the panel deferring to the most expert person (or the loudest screamer)
	+ Use various rounds of voting with electronic discussion to allow consensus to develop

Slide 36: 13. Develop a process to synthesize and grade the evidence inherent in clinical experience

* Can we go further? Can we give additional novice and experienced clinicians than the 8-15 on a guidelines panel a chance to contribute their experiences?
* Possibilities with crowdsourcing?
* What would be our criteria for judging that no biases crept into this: COIs; pre-existing preferences for the treatment (or against a comparator); etc.?

Slide 37: Questions?

Slide 38: When would D&R research evidence be strong?

* Typical of D&R research, whether RCT design or not:
	+ Cannot blind therapist
	+ In most cases, cannot blind subject/patient
	+ Blind assessor works for some outcomes (e. g. observer rating of ADL ability), but not others (e.g. satisfaction with life)
	+ Subjects and therapists know what the comparator is, and potentially have reactions that bias outcome reporting
* As a consequence, D&R research gets downgraded in most evidence schemes
	+ PEDro: 3 out of 10 points
	+ AAN: unless an ‘objective’ outcome is involved, downgraded from level I to level III
	+ GRADE: likely 1-step downgrade because of increased risk of bias
* Consequence: evidence disregarded, or recommendations for D&R interventions never being strongest; payers refuse to pay for services

Slide 39: When would D&R research evidence be strong?

* ACRM Evidence and Practice Committee (EPC; formerly Clinical Practice Committee) discussed “under what circumstances would we consider D&R RCT evidence with subjective outcomes as AAN level II or even level I”?
* Checklist created with very stringent criteria for ‘upgrade’ to level I, somewhat less stringent ones for level II upgrade.

Thank you to the members of ACRM’s Evidence and Practice Committee (EPC) for sharing “Criteria for Patient Reported Outcomes,” an unpublished working draft developed during 2012-2013. The following slides (40-51) describe and summarize these criteria.

More about the EPC: <http://www.acrm.org/resources/evidence-and-practice/>

Slide 40: EPC criterion A:

* The study compares two active concurrent treatments **or** an active treatment and a concurrent control treatment condition that, to the participants, is face valid as a treatment

Slide 41: EPC Criterion B:

* Treatment is ‘not blinded” due to the inherent nature of the treatment (e.g. behavioral intervention, pet therapy), **not** for any other reason

Slide 42: EPC Criterion C:

* The study design contains rigorous methods to create equal expectations of outcomes across non-blinded conditions, including:
	1. No statements are made in the intervention protocol and no steps taken in the experimental procedure that would bias participant expectancies differentially.
	2. Participants are queried (by **independent** individuals without COI) about their awareness, feelings or expectancy biases associated with communications with
		+ any and all research staff
		+ resulting from their assigned treatment group (ONLY required if a face-valid comparator is used)
	3. The treatment itself does not involve directly training the participant in changing verbal behavior related to describing outcomes or symptoms (e.g. Cognitive Behavioral Treatment)

Slide 43: EPC Criterion D:

* Because of the inherent subjective nature of the outcome (e.g. pain, depression) a subjective PRO is the **best** measure of the construct of interest

Slide 44: EPC Criterion E:

* Administration of the subjective PRO uses rigorous methods to minimize bias, including:
	+ The same measurement procedure is used for both treatment conditions
	+ There are standardized administration and scoring procedures that contain no biasing instructions
	+ Individuals (who are **independent of the research team** and are **without COI)** administering the PRO must be
		- Trained in the administration and scoring procedures
		- Blind to treatment group assignment

Slide 45: EPC Criterion F:

* The subjective PRO measure demonstrates evidence of reliability based on either
	+ High rates of test-retest reliability within a very short period of time (e.g., same day, within days) reported in the research literature **OR**
	+ Confirmatory evidence obtained using a secondary measure (e.g. observation) that is expected to be associated with the subjective outcome

Slide 46: EPC Criterion G:

* Investigators with COI (e.g. financial, intellectual):
	+ Must not personally deliver treatment.
	+ Must not collect PRO from participants.
	+ Must not analyze the data; an independent statistician must do the analysis.
	+ Should remain blinded to treatment condition until the final results are determined.

Slide 47: EPC criteria general note:

* The methods used to address criteria C-G and the results of these independent queries to provide evidence that there is no systematic bias in the intervention protocol must **be reported in the manuscript** (or supplemental materials) and demonstrate the equivalence of expectations between treatment groups in order to meet the requirement for upgrading to Level I or II
* i.e. there can be no assuming that something was done or avoided: there must be clear positive statements specifying how the research was conducted

Slide 48: EPC Summary

The slide has a table with three columns. The first contains as column heading the word ‘CRITERION:’; the second has as column heading ‘Class I’ and the third has as column heading ‘Class II’. Below that the rows list each of the criteria, and whether or not they are required to upgrade to Class I or Class II; entries are as follows:

A. The study compares two active treatments OR an active treatment and a face valid control treatment (Class I check, Class II check)

B. Treatment was not blinded due to its inherent nature (Class I check, Class II check)

C. The study contains rigorous methods to assure equal outcome expectancies across non-blinded treatments, including (Class I --, Class II --):

1. No statements are in the treatment protocol and no steps taken in the procedures that would bias participant expectancies differentially (Class I check, Class II check).

2. Individuals independent of the study team and without COI query participants

 with regard to potential differential treatment expectancies resulting from:

a. Communication with investigators, research coordinators and treating

 clinicians AND

b. Their assigned treatment group if a face valid, control treatment is used

Merged cell for 2a. and 2b: (Class I check, Class II no check)

3. The treatment itself does not directly train participants to change their verbal

 behavior related to describing their outcomes (Class I check, Class II check).

D. The subjective PRO was the best measure of the construct of interest due to the inherent subjective nature of the outcome (e.g. pain, depression) (Class I check, Class II check).

Slide 49: EPC Summary

This slide continues the table with listing of criteria in column one, with checkmarks in the next two columns.

CRITERION:

E. Administration of the subjective PRO uses rigorous methods to minimizes

 bias including (Class I --, Class II --)

1. The same measurement procedure for both treatment conditions. (Class I check, Class II check)

2. Standardized administration and scoring procedures that contain no

 biasing instructions (Class I check, Class II check)

3. If an independent person administers the PRO, he/she must be:

a. Trained on the administration and scoring procedures AND

b. Blinded to treatment group assignment.

Merged cell for 3a. and 3b: (Class I check, Class II check)

F. The subjective PRO measure demonstrates evidence of reliability based on

 either:

1. High rates of test-retest reliability within a very short period of time (e.g., same day, within days) reported in the research literature OR

2. Collecting confirmatory, second person observations of study participants on a measure expected to be associated with the subjective PRO.

Merged cell for F1. and F2: (Class I check, Class II check)

Slide 50: EPC Summary

This slide completes the table with listing of criteria in column one, with checkmarks in the next two columns.

CRITERION:

G. Investigators with COI (e.g. financial, intellectual): (Class I --, Class II --)

1. Must not personally deliver treatment (Class I check, Class II check).

2. Must not collect PRO from participants (Class I check, Class II check).

3. Must not conduct the data analysis; an independent statistician must conduct the analysis (Class I check, Class II check).

4. Must remain blinded to treatment condition until the final results are determined (Class I check, Class II check).

Slide 51: Application to 31 treatment studies dealing with anxiety treatment after TBI

Table with 5 columns (Criterion; Clearly not done/ not the case; Clearly done/ the case; No report; Not applicable)

Criterion A: Clearly not=22, Clearly=9, No report=0, N/A=0

Criterion B: Clearly not=11, Clearly=17, No report=1, N/A=2

Criterion C1: Clearly not=9, Clearly=3, No report=13, N/A=6

Criterion C2a: Clearly not=16, Clearly=0, No report=11, N/A=4

Criterion C2b: Clearly not=9, Clearly=0, No report=6, N/A=16

Criterion C3: Clearly not=17, Clearly=6, No report=8, N/A=0

Criterion D: Clearly not=7, Clearly=21, No report=3, N/A=0

Criterion E1: Clearly not=3, Clearly=12, No report=7, N/A=9

Criterion E2: Clearly not=1, Clearly=12, No report=17, N/A=1

Criterion E3a: Clearly not=1, Clearly=1, No report=16, N/A=13

Criterion E3b: Clearly not=3, Clearly=5, No report=10, N/A=13

Criterion F1: Clearly not=5, Clearly=4, No report=21, N/A=0

Criterion F2: Clearly not=20, Clearly=0, No report=11, N/A=0

Criterion G1: Clearly not=4, Clearly=6, No report=21, N/A=0

Criterion G2: Clearly not=7, Clearly=1, No report=23, N/A=0

Criterion G3: Clearly not=3, Clearly=1, No report=25, N/A=2

Criterion G4: Clearly not=4, Clearly=2, No report=24, N/A=1

Slide 52: Questions?

Slide 53: And I have not said anything yet about:

* Qualitative research and the systematic review of qualitative research (study synthesis)
* Mixed methods research
	+ - Triangulation
		- Embedded
		- Sequential exploratory
		- Sequential explanatory
* Mixed-methods systematic reviews
	+ Systematic review of mixed-method studies
	+ Systematic review that incorporates quantitative and qualitative (and mixed-methods) primary studies
		- Apples and oranges? Fruit salad or undifferentiated flavorless mush?

Slide 54: One step back, one step forward:

* EBP/EBM has always
	+ focused on the individual patient
	+ assumed that delivering the evidence in the hands of the clinician is enough
* Increasing emphasis on fact that evidence, however strong and nicely packaged, is not enough
* Evidence, and even a guideline, is only a step in process of knowledge translation (KT)
* KT is developing into a science, with a need to have systematic reviews of the evidence on what is the best way to deliver new knowledge, technology, etc. into the hands of clinicians, and see it used systematically

Slide 55: Principles of the London-based EPPI-Centre (Evidence for Policy and Practice Information and Co-ordinating Centre):

1. Both primary research and reviews of research are essential to the progress of knowledge
2. There is a range of primary research methods. There is a wide range of review methods
3. Reviews should follow the research principles of quality, rigor, and accountability, similar to how these are used in primary research
4. Review methods often reflect the methods, epistemological assumptions, and methodological challenges found in primary research
5. Reviews should be driven by questions which vary in many ways, including theoretical and ideological perspective
6. Users of research have particular perspectives and priorities that can usefully inform primary research and reviews of research

(Cited in Hansen, 2014, p. 13)

Slide 56: What evidence do you need?

* Impact evidence (effects of interventions)\*
* Implementation evidence (knowledge about the process of carrying out an activity)
* Attitudinal evidence (assessment of the intervention by users/experts)
* Economic evidence (relation between costs and benefits)\*
* Ethical evidence (knowledge about value questions)

\* application of systematic reviews

Slide 57: LATEST NEWS (from November 2013)

* COCHRANE publishes a review of qualitative evidence (Glenton C et al. *Cochrane Database of Systematic Reviews* 2013;10:CD010414.)
* It is not about the effectiveness of health interventions, or the accuracy of diagnostic tests/ screening tests
* It is about barriers to and facilitators of the implementation of lay health worker programs
* It was used in combination with the outcomes of a Cochrane *effectiveness* review on the use of lay health workers in community health care for maternal and child health, providing a comprehensive assessment of this strategy
* This supports claims by others (e.g. Hansen, 2014) that evidence of effectiveness is not enough, we need evidence on
	+ the process of service delivery (*how does it work?)*
	+ appropriateness of care (*is this the right service for these clients/patients?)*
	+ acceptability (*will patients/clients want it?)*
	+ satisfaction with care (*are stakeholders satisfied with the service?)*

Slide 58: There is more to be done: discussion of / developments of methods incorporating:

* Johnston/Dijkers recommendations
* Issue of generalizability, and grading studies based on external validity
* ACRM EPC PRO recommendations, and incorporation into AAN, other evidence grading systems
* Grading the quality of evidence of qualitative and mixed-methods studies
* Methods for incorporating into phrasing of recommendations qualitative, quantitative and mixed-methods evidence
* Role of evidence in a larger context and evidence for KT

**Who is ready to start the discussion?**

Slide 59: Questions?

Slide 60: Wrapping Up

We invite you to:

* Provide your input on today’s session
* Share your ideas for future sessions
* Participate in the Community of Practice to continue the dialogue
* PLEASE CONTACT US: joann.starks@air.org

Please fill out the brief Evaluation Form:

<http://www.surveygizmo.com/s3/1697350/Evaluation-Webinar-Series-Session-4>

Slide 61: Disclaimer

This webinar was developed for grant number H133A120012 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR), Office of Special Education and Rehabilitative Services (OSERS), U.S. Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the federal government.