



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

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

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
4. A discussion of future developments in methods of qualifying and synthesizing evidence that might benefit disability/rehabilitation practice



Questions?

July 2 topics: GRADE

- 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 for all decisions taken transparent
- Based on RCT vs observational study (and even weaker evidence??); four evidence qualities distinguished: high, moderate, low, very low

July 2 topics: GRADE

- 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

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

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



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(11)01142-7/pdf). Reprinted by Marcel Dijkers in compliance with Elsevier's author rights.

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

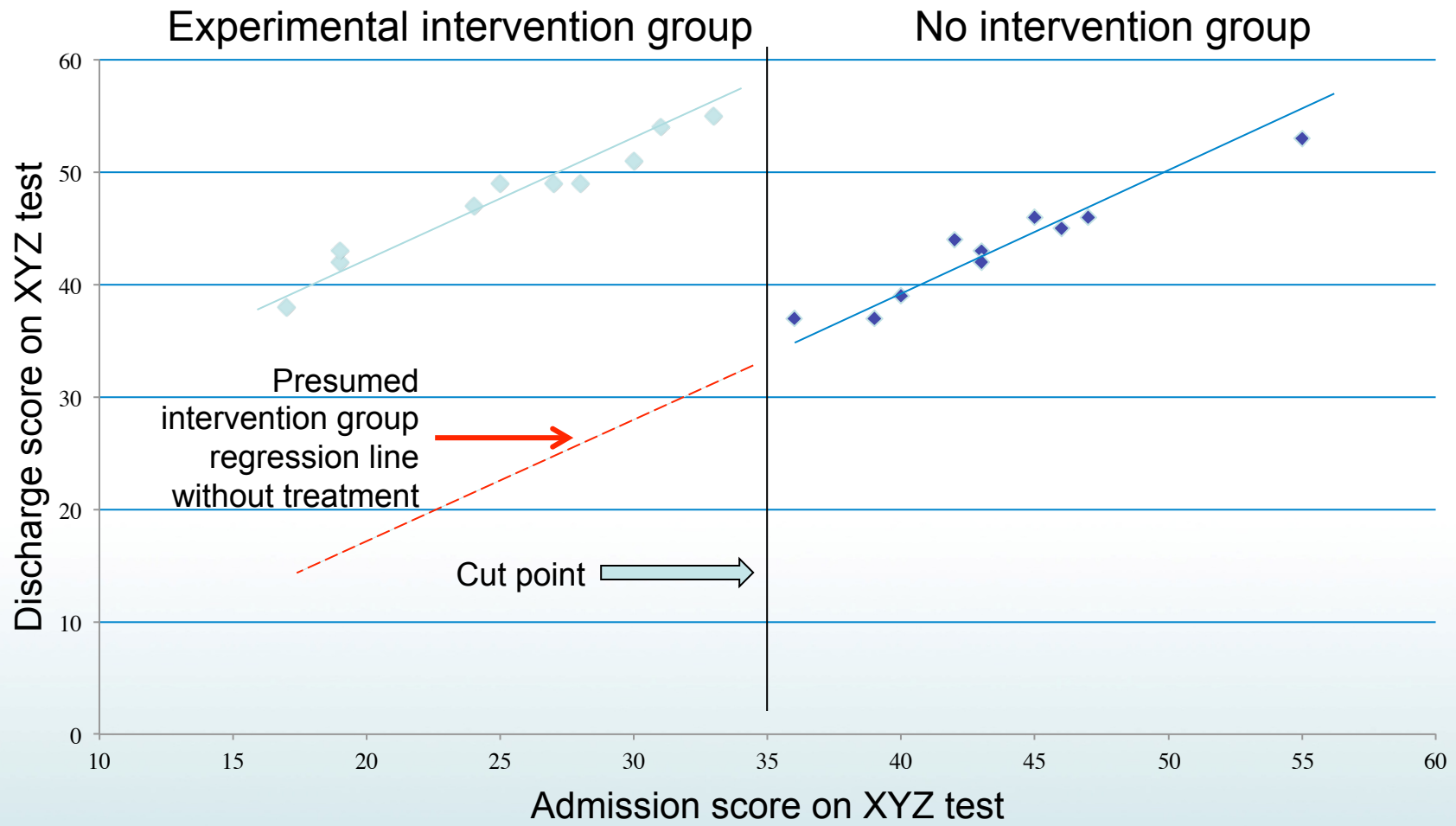
1: Define outcomes in terms meaningful and important to the persons served

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

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

Regression discontinuity design



Interrupted time series design

		Traffic deaths per 1,000,000 miles in the year																				
state	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	mean pre	mean post
AL	4	6	3	4	x	6	4	7	5	7	8	6	8	9	8	5	9	9	7	8	4.3	7.1
CO	3	4	3	5	3	2	x	4	5	8	6	7	5	7	8	6	6	6	5	7	3.3	6.2
DE	6	7	6	8	7	6	5	6	7	8	x	8	9	10	10	11	9	8	9	10	6.6	9.3
FL	6	5	7	8	5	6	7	5	6	7	6	5	6	6	x	9	8	9	10	10	6.1	9.2

x: increase of maximum speed limit to 65 miles/hour or higher

Single subject design with multiple baseline

Over subjects

score on outcome measure on successive days													
patient	1	2	3	4	5	6	7	8	9	10	11	mean pre	mean post
A	4	6	3	x	6	4	7	8				4.3	6.3
B			3	3	2	x	4	6	7			3.0	6.2
C					6	5	6	x	8	10	11	6.1	9.7
D						7	5	6	5	x	9	5.8	9.0

Over outcomes

score on outcome measure X on successive days													
outcome	1	2	3	4	5	6	7	8	9	10	11	mean pre	mean post
K	4	6	3	x	6	4						4.3	6.3
L			3	3	2	x	4	6				3.0	6.2
M				7	6	5	6	x	8	10	11	6.1	9.7
N								6	5	x	9	5.8	9.0

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

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

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

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

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

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?

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

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

7: Establish requirements to insure expertise and minimize bias of review panels

- If we eliminated all persons with any COI (financial, intellectual, other) 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

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

- GRADE

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** [<http://www.ktdrr.org/aqasr>])
- Disability and rehabilitation researchers should be especially interested
- No EBP organization has focused on this – not even the Campbell Collaboration

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)

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?

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-)morbidity
- Health system issues to consider
 - Referral patterns
 - Resources at 1^o, 2^o, 3^o care centers
 - Expertise of clinicians
 - Patient/client-clinician rapport

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

Component	A (Excellent)	B (Good)	C (Satisfactory)	D (Poor)
Generalizability to target population	Population studied is the same as the target population	Population studied is similar to the target population	Population studied is different but it is clinically sensible to apply this evidence to the target population	Population studied is different and it is hard to judge whether it is sensible to generalize
Applicability to target context	Evidence is directly applicable to the context of the target population	Evidence is applicable to the local context with few caveats	Evidence is probably applicable . . . with some caveats	Not applicable to local context

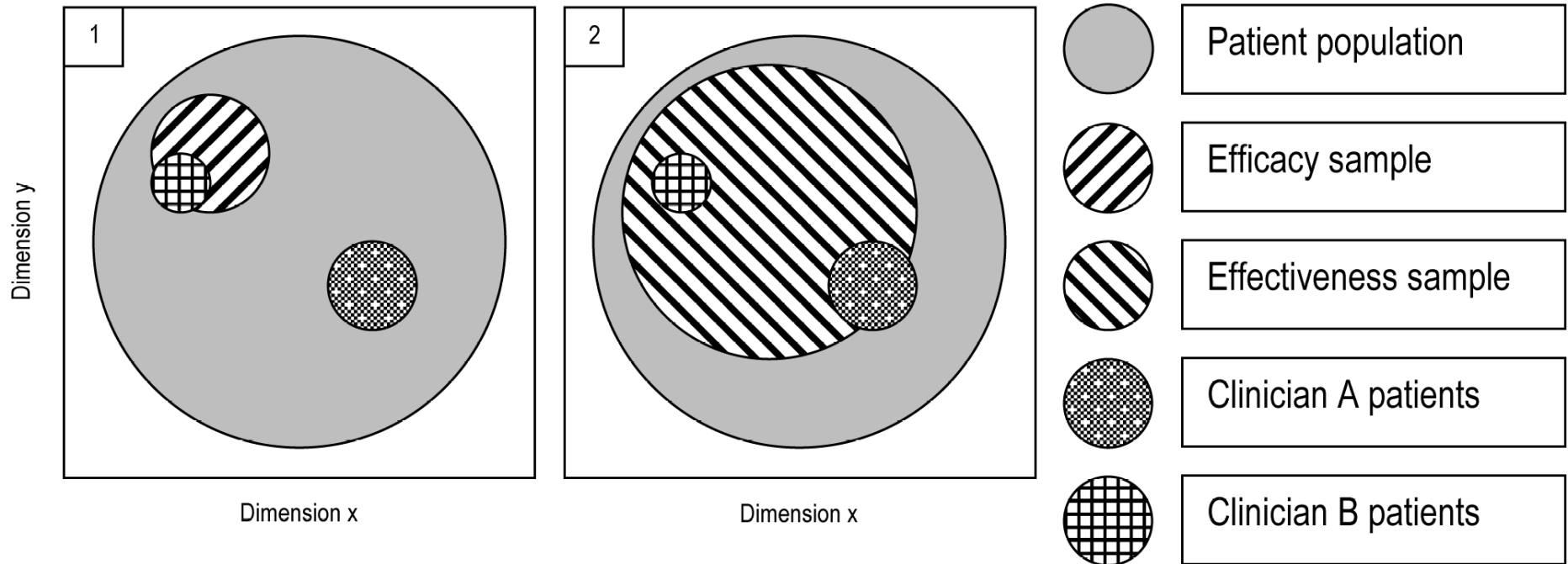
From Table 2 in "Toward improved evidence standards and methods for rehabilitation: recommendations and challenges," by M. V. Johnston and M. P. Dijkers, 2012, *Archives of Physical Medicine and Rehabilitation*, 93(8 Suppl 2), S185-189. Retrieved from [www.archives-pmr.org/article/S0003-9993\(11\)01142-7/pdf](http://www.archives-pmr.org/article/S0003-9993(11)011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

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

The paradox of generalizability



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.

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

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?

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)

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

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](#))
 - 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

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

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.?



Questions?

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

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/>

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.

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.

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).

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.

EPC criterion E:

Administration of the subjective PRO uses rigorous methods to minimize bias, including:

1. The same measurement procedure is used for both treatment conditions.
2. There are standardized administration and scoring procedures that contain no biasing instructions.
3. 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

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

EPC criterion G:

Investigators with COI (e.g. financial, intellectual):

1. Must not personally deliver treatment.
2. Must not collect PRO from participants.
3. Must not analyze the data; an independent statistician must do the analysis.
4. Should remain blinded to treatment condition until the final results are determined.

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.

EPC summary:

CRITERION:	Class I	Class II
A. The study compares two active treatments OR an active treatment and a face valid control treatment	√	√
B. Treatment was not blinded due to its inherent nature	√	√
C. The study contains rigorous methods to assure equal outcome expectancies across non-blinded treatments, including:	--	--
1. No statements are in the treatment protocol and no steps taken in the procedures that would bias participant expectancies differentially.	√	√
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.		
3. The treatment itself does not directly train participants to change their verbal behavior related to describing their outcomes.	√	√
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).	√	√

EPC summary:

CRITERION:	Class I	Class II
E. Administration of the subjective PRO uses rigorous methods to minimize bias including:	--	--
1. The same measurement procedure for both treatment conditions.	√	√
2. Standardized administration and scoring procedures that contain no biasing instructions.	√	√
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.		
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.		

EPC summary:

CRITERION:	Class I	Class II
G. Investigators with COI (e.g. financial, intellectual):	--	--
1. Must not personally deliver treatment.	√	√
2. Must not collect PRO from participants.	√	√
3. Must not conduct the data analysis; an independent statistician must conduct the analysis.	√	√
4. Must remain blinded to treatment condition until the final results are determined.	√	√

Application to 31 treatment studies dealing with anxiety treatment after TBI

Criterion	Clearly not done/not the case	Clearly done/ the case	No report	Not applicable
A	22	9	0	0
B	11	17	1	2
C1	9	3	13	6
C2a	16	0	11	4
C2b	9	0	6	16
C3	17	6	8	0
D	7	21	3	0
E1	3	12	7	9
E2	1	12	17	1
E3a	1	1	16	13
E3b	3	5	10	13
F1	5	4	21	0
F2	20	0	11	0
G1	4	6	21	0
G2	7	1	23	0
G3	3	1	25	2
G4	4	2	24	1



Questions?

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?

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

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)

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

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?*)

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?



Questions?

Wrapping Up

Thank you for participating!

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@sedl.org

Please fill out the brief Evaluation Form:

www.surveygizmo.com/s3/1697350/Evaluation-Webinar-Series-Session-4

Disclaimer

This webinar was developed for grant number H133A120012 from the National Institute on Disability and Rehabilitation Research (NIDRR), 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.