Systematic reviews: From evidence to recommendation

Session 2 - June 18, 2014
Going beyond design, going beyond intervention: The American Academy of Neurology (AAN) Clinical Practice Guideline process and other approaches

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

A webinar sponsored by SEDL's Center on

Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

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

June 4 topics:

- The role of (clinical) research evidence in practitioner decision making: primary studies, EBP resources
- The process of creating a systematic review
- The meaning of 'evidence' and the need to evaluate the quality of evidence resulting from clinical research:
  - Big D design
  - Little d design
  - Research implementation

June 4 topics:

- Hierarchies for evidence relevant to interventions, developed by:
  - Sackett
  - Cicerone et al.
- Checklists and rating scales for evidence quality:
  - Jadad scale
  - PEDro scale
- Possible uses of checklists and rating scales

Questions?

There is more to EBP than RCTs

- RCTs are irrelevant to prognosis, diagnosis, screening, etc.

- Multiple hierarchies are / may be needed, each one for a specific research question

- And have been developed since Sackett published his hierarchy for intervention studies

Oxford CEBM 2011 hierarchy of evidence

question	How common is the problem?	Is this Dx or monitoring test accurate? (diagnosis)	What will happen if we do not add a therapy? (prognosis)	Does this intervention help? (treatment benefits)	What are the COMMON harms? (treatment harms)	What are the RARE harms? (treatment harms)	Is this (early detection) test worthwhile? (screening)
level 1 evidence							
level 2 evidence							
level 3 evidence							
level 4 evidence							
level 5 evidence							

http://www.cebm.net/mod_product/design/files/CEBM-Levels-of-Evidence-2.1.pdf

A note

- The Oxford CEBM (Centre for Evidence-Based Medicine) hierarchies are for clinicians conducting 'bedside' EBP, not for researchers conducting systematic reviews:
  - Systematic reviews generally hold the highest rank in the hierarchies
  - A N-of-1 trial *with the patient the clinician needs to treat* also take a high position

Oxford CEBM evidence levels for Incidence/ prevalence: How common is the problem?

Level 1	Local and current random sample surveys (or censuses)
Level 2	Systematic review of surveys that allow matching to local circumstances
Level 3	Local non-random sample
Level 4	Case-series
Level 5	n/a

Oxford CEBM evidence levels for Diagnosis: Is this diagnostic or monitoring test accurate?

Level 1	Systematic review of cross sectional studies with consistently applied reference standard and blinding
Level 2	Individual cross sectional studies with consistently applied reference standard and blinding
Level 3	Non-consecutive studies, or studies without consistently applied reference standards
Level 4	Case-control studies, or poor or non-independent reference standard
Level 5	Mechanism-based reasoning

Oxford CEBM evidence levels for Prognosis: What will happen if we do not add a therapy?

Level 1	Systematic review of inception cohort studies
Level 2	Inception cohort studies
Level 3	Cohort study or control arm of randomized trial
Level 4	Case-series or case-control studies, or poor quality prognostic cohort study
Level 5	n/a

Oxford CEBM evidence levels for Treatment Benefits: Does this intervention help?

Level 1	Systematic review of randomized trials or n-of-1 trials
Level 2	Randomized trial or observational study with dramatic effect
Level 3	Non-randomized controlled cohort/follow-up study
Level 4	Case-series, case-control studies, or historically controlled studies
Level 5	Mechanism-based reasoning

Oxford CEBM evidence levels for Treatment Harms: What are the COMMON harms?

Level 1	Systematic review of randomized trials, systematic review of nested case-control studies, n-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect
Level 2	Individual randomized trial or (exceptionally) observational study with dramatic effect
Level 3	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)
Level 4	Case-series, case-control, or historically controlled studies
Level 5	Mechanism-based reasoning

Oxford CEBM evidence levels for Treatment Harms: What are the RARE harms?

Level 1	Systematic review of randomized trials or n-of-1 trial
Level 2	Randomized trial or (exceptionally) observational study with dramatic effect
Level 3	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)
Level 4	Case-series, case-control, or historically controlled studies
Level 5	Mechanism-based reasoning

Oxford CEBM evidence levels for Screening: Is this (early detection) test worthwhile?

Level 1   Systematic review of randomized trials

Level 2   Randomized trial

Level 3   Non-randomized controlled cohort/follow-up study

Level 4   Case-series, case-control, or historically controlled studies

Level 5   Mechanism-based reasoning

CEBM language: note the subjective elements

- "consistently applied"
- "poor standard"
- "poor quality cohort study"
- "dramatic effect"
- "sufficient numbers"

Oxford CEBM 2011 hierarchy of evidence: presentation

question	level 1 evidence	level 2 evidence	level 3 evidence	level 4 evidence	level 5 evidence
How common is the problem?					
Is this Dx or monitoring test accurate? (diagnosis)					
What will happen if we do not add a therapy? (prognosis)					
Does this intervention help? (treatment benefits)					
What are the COMMON harms? (treatment harms)					
What are the RARE harms? (treatment harms)					
Is this (early detection) test worthwhile? (screening)					

Footnotes to the Oxford 2011 scheme

- Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match question PICO), because of inconsistency between studies, or because the absolute effect size is very small (GRADE)

- Level may be graded up if there is a large or very large effect size (GRADE)

- The Oxford website does not offer a scheme to link the quality and quantity of the evidence to the strength of a recommendation – a consequence of focus on 'bedside' EBP?

Oxford CEBM grades of recommendation 2009 grid

A   consistent level 1 studies

B   consistent level 2 or 3 studies *or* extrapolations from level 1 studies

C   level 4 studies *or* extrapolations from level 2 or 3 studies

D   level 5 evidence *or* troublingly inconsistent or inconclusive studies
of any level

Action (must, should, may) related to letter grades not to be found!!

2011 Edition

Clinical Practice Guideline Process Manual

Prepared by

Gary S. Gronseth, MD, FAAN
Laura Moses Woodroffe
Thomas S. D. Getchius

http://tools.aan.com/globals/axon/assets/9023.pdf

American Academy of Neurology (AAN) 2011 Clinical Practice Guideline Process Manual

- Earlier edition: 2004

- Next edition: 2014?

- Used by AAN groups

- Used by many ACRM groups (but not Cicerone et al.)

EBP Process as Applied by the American Academy of Neurology (AAN)

## A. Developing the Questions

  i. PICO Format

   Or PICOTS: time frame, setting

   All other questions (Dx, Px, etc.) squeezed into PICOTS
   framework – e.g. for screening question, 'I' is doing screening,
   'C' is not doing screening

  ii. Types of clinical questions

  iii. Development of an analytic framework

An hypothetical and simple analytic framework

Dx: 90% sensitive	→	Tx: NNT is 4.8	→	Px: 5 year future 30% predicted
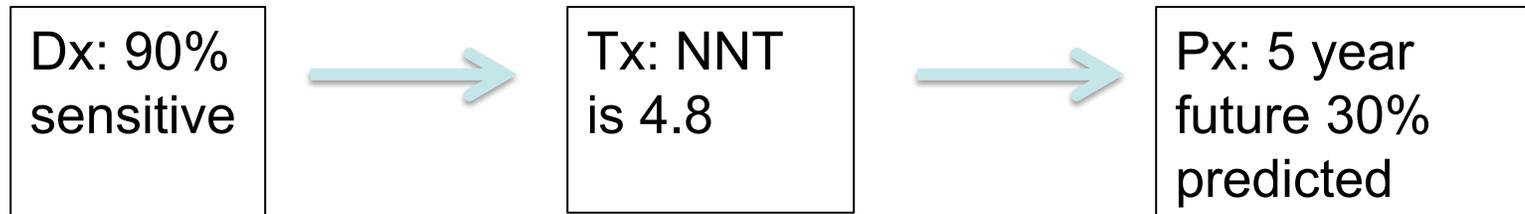
EBP Process as Applied by the American Academy of Neurology (AAN)

B. Finding and Analyzing Evidence

i. Finding the relevant evidence

ii. Identifying methodological characteristics of the studies

iii. Rating the risk of bias

iv. Understanding measures of association

v. Understanding measures of statistical precision

vi. Interpreting a study

EBP Process as Applied by the American Academy of Neurology (AAN)

C. Synthesizing Evidence—Formulating Evidence-based Conclusions

i. Accounting for conflicting evidence

ii. Knowing when to perform a meta-analysis

iii. Wording conclusions for nontherapeutic questions

iv. Capturing issues of generalizability in the conclusion

EBP Process as Applied by the American Academy of Neurology (AAN)

D. Making Practice Recommendations

    i. Rating the overall confidence in the evidence from the perspective of supporting practice recommendations

    ii. Putting the evidence into a clinical context

    iii. Crafting the recommendations

    iv. Basing recommendations on surrogate outcomes

    v. Knowing when not to make a recommendation

    vi. Making suggestions for future research

Questions?

AAN evidence classification schemes for:

- Therapy

- Causation

- Prognosis

- Diagnosis

- Population screening

Classification of Evidence Scheme: Therapy – class I

- Randomized, controlled clinical trial (RCT) in a representative population

- Masked or objective outcome assessment

- Relevant baseline characteristics are presented and substantially equivalent between treatment groups, or there is appropriate statistical adjustment for differences

- Also required:

  a. Concealed allocation

  b. Primary outcome(s) clearly defined

  c. Exclusion/inclusion criteria clearly defined

  d. Adequate accounting for dropouts (with at least 80% of enrolled subjects completing the study) and crossovers with numbers sufficiently low to have minimal potential for bias

  e. For noninferiority or equivalence trials …

What is 'objective'?

- Objective outcome measurement: an outcome measure that is unlikely to be affected by an observer's (patient, treating physician, investigator) expectation or bias (e.g., blood tests, administrative outcome data)

    – Level I: unmasked investigator and unmasked patient cannot

    influence measurement of the outcome

    – Level II: either the unmasked investigator or unmasked patient (but

    not both) can influence measurement of the outcome

    – Level III: unmasked investigator and unmasked patient can

    influence measurement of the outcome

- For AAN guidelines, only level I is objective > major issue for disability and rehabilitation studies, where blinding often is not possible, and either therapist or patient rates treatment outcomes

Classification of Evidence Scheme: Therapy – class II

- Cohort study meeting criteria a–e (see Class I) or an RCT that lacks one or two criteria b–e (see Class I)

- All relevant baseline characteristics are presented and substantially equivalent among treatment groups or there is appropriate statistical adjustment for differences

- Masked or objective outcome assessment

Classification of Evidence Scheme: Therapy – class III

- Controlled studies (including well-defined natural history controls or patients serving as their own controls)

- A description of major confounding differences between treatment groups that could affect outcome

- Outcome assessment masked, objective or performed by someone who is not a member of the treatment team.

Classification of Evidence Scheme: Therapy – class IV

- Did not include patients with the disease (indirectness – GRADE)

- Did not include patients receiving different interventions (i.e. comparator)

- Undefined or unaccepted interventions or outcome measures

- No measures of effectiveness or statistical precision presented or calculable

AAN approach to evidence synthesis and formulating conclusions

- Link to clinical question
- Consideration of four types of information
    - Class of evidence all studies included
    - Strength of associations between treatment and outcome (effect size) (GRADE)
    - Statistical precision (confidence intervals and statistically pooled confidence interval) (GRADE)
    - Consistency between studies (GRADE)

AAN characterization of individual studies

I   low risk of bias

II   moderate risk of bias

III   moderately high risk of bias

IV   very high risk of bias

AAN approach to evidence synthesis and formulating conclusions: therapy questions

Findings	Phrasing
in case of effective therapy	
Multiple class I studies	Tx X is highly likely to be effective
Multiple class II studies / single class I study	Tx X is likely to be effective
Multiple class III studies / single class II study	Tx X is possibly effective
Multiple class IV studies / single class III study	Insufficient evidence in favor of or against Tx X

AAN approach to evidence synthesis and formulating conclusions: therapy questions

Findings	Phrasing
in case of ineffective therapy	
Multiple negative, adequately powered class I studies	Tx X is highly likely NOT to be effective
Multiple negative, adequately powered class II studies/single adequately powered class I study	Tx X is likely NOT to be effective
Multiple negative, adequately powered class II studies/single adequately poweredclass II study	Tx X is possibly NOT effective
Multiple negative class IV studies/single adequately powered class III study/ negative, inadequately powered class I, II or III studies	Insufficient evidence in favor of or against Tx X

Questions?

AAN Classification of Evidence Scheme: Prognosis – class I

- Cohort survey with prospective data collection

- Includes a broad spectrum of persons at risk for developing the outcome

- Outcome measurement is objective or determined without knowledge of risk factor status

- Also required:

  a. Inclusion criteria defined

  b. At least 80% of enrolled subjects have both the risk factor and outcome measured

AAN Classification of Evidence Scheme: Prognosis – class II

- Cohort study with retrospective data collection or case-control study. Study meets criteria a and b (see Class I)

- Includes a broad spectrum of persons with and without the risk factor and the outcome

- The presence of the risk factor and outcome are determined objectively or without knowledge of one another

AAN Classification of Evidence Scheme: Prognosis – class III

- Cohort or case control study

- Narrow spectrum of persons with or without the disease

- The presence of the risk factor and outcome are determined objectively, without knowledge of the other or by different investigators

AAN Classification of Evidence Scheme: Prognosis – class IV

- Did not include persons at risk for the outcome

- Did not include patients with and without the risk factor

- Undefined or unaccepted measures of risk factor or outcomes

- No measures of association or statistical precision presented or calculable

Questions?

AAN Classification of Evidence Scheme: Diagnosis – class I

- Cohort survey with prospective data collection

- Includes a broad spectrum of persons suspected of having the disease

- Disease status determination is objective or made without knowledge of diagnostic test result

- Also required:

  a. Inclusion criteria defined

  b. At least 80% of enrolled subjects have both the diagnostic test and disease status measured

AAN Classification of Evidence Scheme: Diagnosis – class II

- Cohort study with retrospective data collection or case-control study. Study meets criteria a and b (see Class I)

- Includes a broad spectrum of persons with and without the disease

- The diagnostic test result and disease status are determined objectively or without knowledge of one another

AAN Classification of Evidence Scheme: Diagnosis – class III

- Cohort or case control study

- Narrow spectrum of persons with or without the disease

- The diagnostic test result and disease status are determined objectively, without knowledge of the other or by different investigators

AAN Classification of Evidence Scheme: Diagnosis – class IV

- Did not include persons suspected of the disease

- Did not include patients with and without the disease

- Undefined or unaccepted independent reference standard

- No measures of diagnostic accuracy or statistical precision presented or calculable

Questions?

AAN approach to evidence synthesis and formulating conclusions

- Link to clinical question

- Consideration of four types of information

  - Class of evidence (I, II, III or IV) for all studies

  - Strength of associations between treatment and outcome (effect sizes of individual studies and pooled studies)

  - Statistical precision (confidence intervals of individual studies and pooled studies)

  - Consistency between studies

Conflicting evidence: what to do?

- Approach 1: try to explain inconsistencies from systematic or random error
  - Bias: disregard the weaker studies (class III), and base conclusion on stronger ones, if these are consistent
  - Random error: disregard the 'out-of-line' studies if they have low power, even if of same class as the 'in-line' studies
  - Meta-analysis: do a formal meta-analysis (if homogeneity of studies is adequate) to see whether the pooled studies support a conclusion

Conflicting evidence: what to do?

- Approach 2: try to explain inconsistencies from systematic or random error

  – Study differences: evaluate whether disagreements between studies (of same class and adequately powered) can be explained by creating subgroups based on population, intervention, comparator, and/or outcome measure (PICO)

Considering generalizability issues

- Assess whether generalizability of the evidence is limited because the combined studies were limited in
    - Subgroups studied (e.g. only or predominantly males)
    - Intervention strengths (dose) or varieties studied
    - Comparators studied
    - Time points after intervention termination studied
    - Outcomes studied
- Phrase recommendation so as to reflected the limit(s) of the evidence
    - e.g. "it is highly recommended that women with problem X receive Tx Y, but the effectiveness of Tx Y for men has not been established"

AAN: interpretation of the levels of evidence

I   (low risk of bias)

II   (moderate risk of bias)

III   (moderately high risk of bias)

IV   (very high risk of bias)

Operationalization of AAN criteria for therapy

Article ID#: _____   Authors and year:_____

Criterion A: Design type of the study (see AAN manual Appendix 3 for details)

- *If a randomized controlled trial, maximum Class I*

- *If not a randomized trial, maximum Class II*

- *If well-defined natural history controls or patients serving as their own controls, maximum Class III*

- *If no comparison group, single subject design, case study or other (e.g., qualitative), maximum Class IV*

Criterion A rating: ____

Operationalization of AAN criteria for therapy

Criterion B: Blinding and assessor independence

1. Was the person doing the outcome assessment blinded to management strategy? (i.e., treatment arm the subject was assigned to)

Yes / No / Not stated / Ambiguous

2. Was the outcome measure objective? *(Objective: The determination of the outcome is unlikely to be affected by observer expectations). NOTE: If the subjects knows the treatment they received AND the outcome is self-reported, it is not objective or masked.*

Yes / No / Not stated / Ambiguous

3. Was the person doing the outcome assessment independent? *(Independent: The investigator assessing outcome was different than the treating clinicians. A subject giving self-report measure of fatigue is independent)*

Yes / No / Not stated / Ambiguous

Criterion B *For the outcome rating consider items 1-3:*
- *If B1 or B2 = YES, maximum is Class I*
- *If only B3 = YES, maximum is Class III*
- *If all (B1, 2, 3) = NO/NOT STATED: Class IV*

Criterion B rating: ____

Operationalization of AAN criteria for therapy

Criterion C: Other Therapeutic Study Characteristics

4. Was treatment allocation concealed (Check "no" if not an RCT): *"Concealed Allocation": Investigators could not manipulate treatment assignment. Examples of concealed allocation include consecutively numbered sealed, opaque envelopes containing a predetermined, random sequence for treatment assignment or an independent center that an investigator contacts to obtain the treatment assignment.*

Yes / No / Not stated / Ambiguous

5. Primary outcome measure(s) was specified

Yes / No / Not stated / Ambiguous

6. Explicit inclusion and exclusion criteria were used

Yes / No / Not stated / Ambiguous

7. Patients in different treatment arms were similar at baseline or appropriate statistical adjustments were made for baseline differences

Yes / No / Not stated / Ambiguous   (cont.)

Operationalization of AAN criteria for therapy

Criterion C: Other Therapeutic Study Characteristics (cont.)

8. Less than 20% of patients were lost to follow-up

Yes / No / Not stated / Ambiguous

(Percentage lost to follow-up: ___)

***Criterion C** For criterion C consider items 4-8: (see AAN manual for details)*

If all 5 (C4, 5, 6, 7, 8) = "yes," maximum is Class I.

If only three or four = "yes," maximum is Class II.

If < three = "yes," maximum is Class III.

Criterion C rating: ___

Operationalization of AAN criteria for therapy

Criterion A rating: _____

Criterion B rating: _____

Criterion C rating: _____

Final Rating: Select worst maximum therapy class from above
(criteria A, B and C): overall rating: _____

    I   (low risk of bias)

    II   (moderate risk of bias)

    III   (moderately high risk of bias)

    IV   (very high risk of bias)

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/xxxxxx/Evaluation-Webinar-Series-Session-2

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.