



## **Systematic reviews: From evidence to recommendation**

**Session 1 - June 4, 2014**

### **A blast from the past: Systematic reviews and the traditional evidence pyramid**

*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)

*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?

## My background

- Research professor of rehabilitation medicine at Icahn School of Medicine at Mount Sinai
- Trained as a social scientist, with expertise in statistics and methodology
- 30+ year of experience in rehabilitation research, especially TBI and SCI
- Last 10 years special focus on evidence-based practice, systematic reviewing/meta-analysis
- Co-author of a number of reviews and guidelines
- Lead author of the development of AQASR

## Logistical support

- Joann Starks, MEd, SEDL's Center on Knowledge Translation for Disability and Rehabilitation Research
- Contact with any questions: [Joann.Starks@sedl.org](mailto:Joann.Starks@sedl.org)

# Influences on a clinician's decisions

1. Training and experience (provided by / guided by experts who may or may not be up-to-date on latest clinical research)
  - A. Professional preservice training
    - I. Basic science
    - II. Clinical science didactics
    - III. Clinical science practicums
  - B. Continuing education and other inservice training
  - C. Own experiences and accumulated expertise
2. Patient values and preferences

## Influences on a clinician's decisions

### 3. Societal and health care system factors

- A. Societal values
- B. Organization of the health care system
- C. Laws and regulations specifying professional roles and privileges
- D. Feasibility of/ reimbursement for diagnostic, treatment and management actions
- E. Organizational mandates (potentially based on clinical research findings)



# Influences on a clinician's decisions

## 4. Clinical research reports

A. Primary studies

B. EBP resources

I. Systematic reviews

II. Critically assessed topics (CATs)

III. Evidence journals' summaries of primary studies

IV. Clinical guidelines

V. Etc.

(Dijkers MP, Murphy SL, Krellman J. Evidence-based practice for rehabilitation professionals: concepts and controversies. Arch Phys Med Rehabil, 2012;93:164-76)

# Evidence

(Webster's Third New International Dictionary 1966)

- 1a. An outward sign (indication, token)
- 1b. Something that furnishes or tends to furnish proof;  
medium of proof (proof, testimony)
2. (archaic) The state of being evident (clearness)

## Evidence's Latin roots

- *evidentia* (evidence; obviousness; quality of being manifest)

derived by way of

- *evidens*: (clear, distinct, plain, visible, evident)

from

- *videre* (to see)

## Criteria for judging evidence:

- **Relevance:** evidence constitutes information for (or against) a specific proposition (or relevant to a specific clinical question)
- **Sufficiency:** information needs to meet criteria for corroboration by other pieces of information on the same topic
- **Veracity:** the process of information gathering has been free from distortion, and (to the degree possible) uncontaminated by conflicts of interest (COI)

## Wikipedia article: 'burden of proof' article

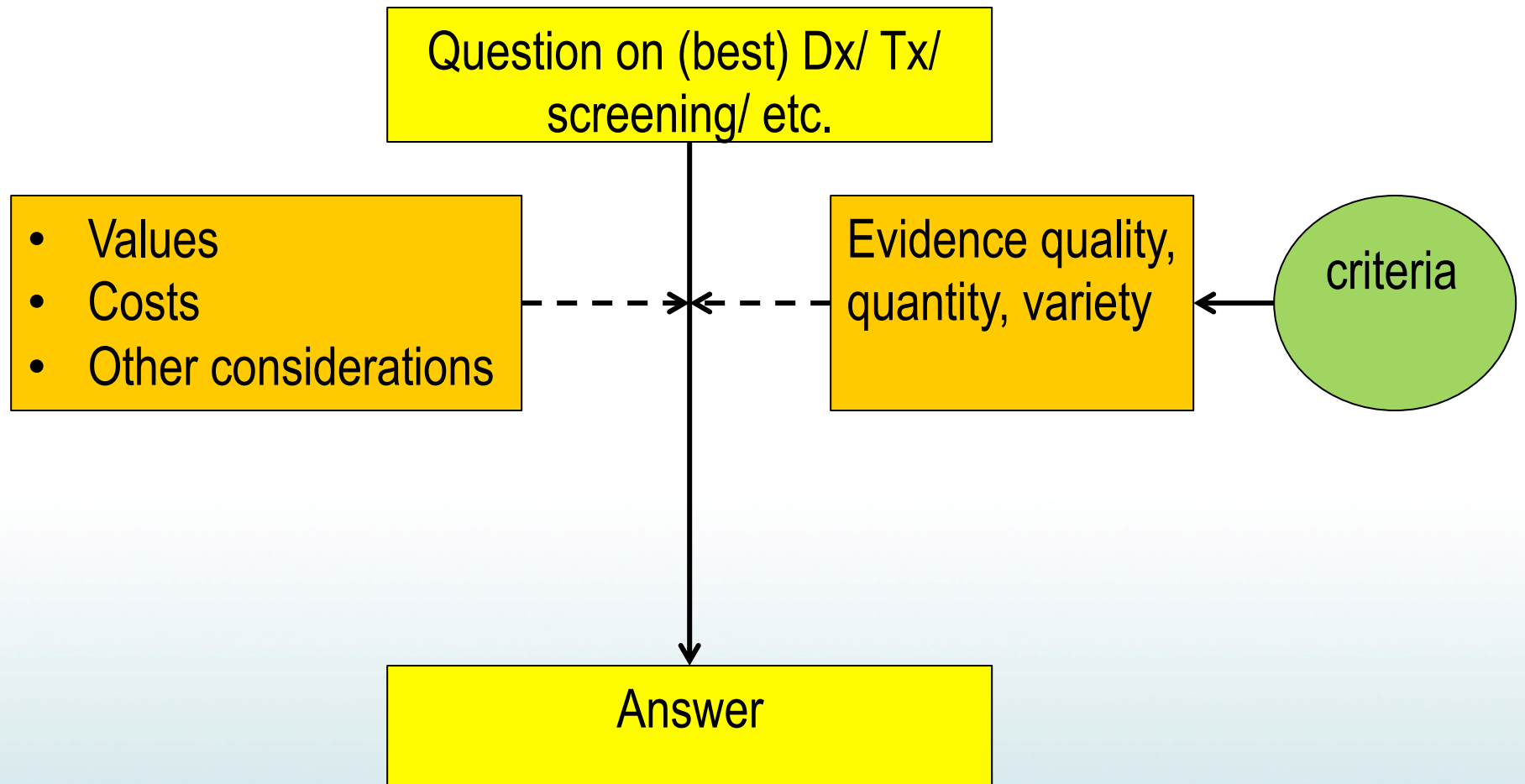
- Reasonable suspicion
- Reasonable to believe
- Probable cause
- Some credible *evidence*
- Substantial *evidence*
- Preponderance of the *evidence*
- Clear and convincing *evidence*
- Beyond reasonable doubt
- Beyond the shadow of a doubt

## Evidence is ....

- One single study (of relevance, sufficient quality, etc.)
- The body of all studies (of sufficient quality, etc.) relevant to a single clinical question, preferably consistent, etc., and summarized qualitatively or quantitatively (meta-analysis)\*

\* this is what the term tends to mean in evidence-based practice (EBP)

# The EBP process - ideally





Questions?



## MedLine: definition of “Review”

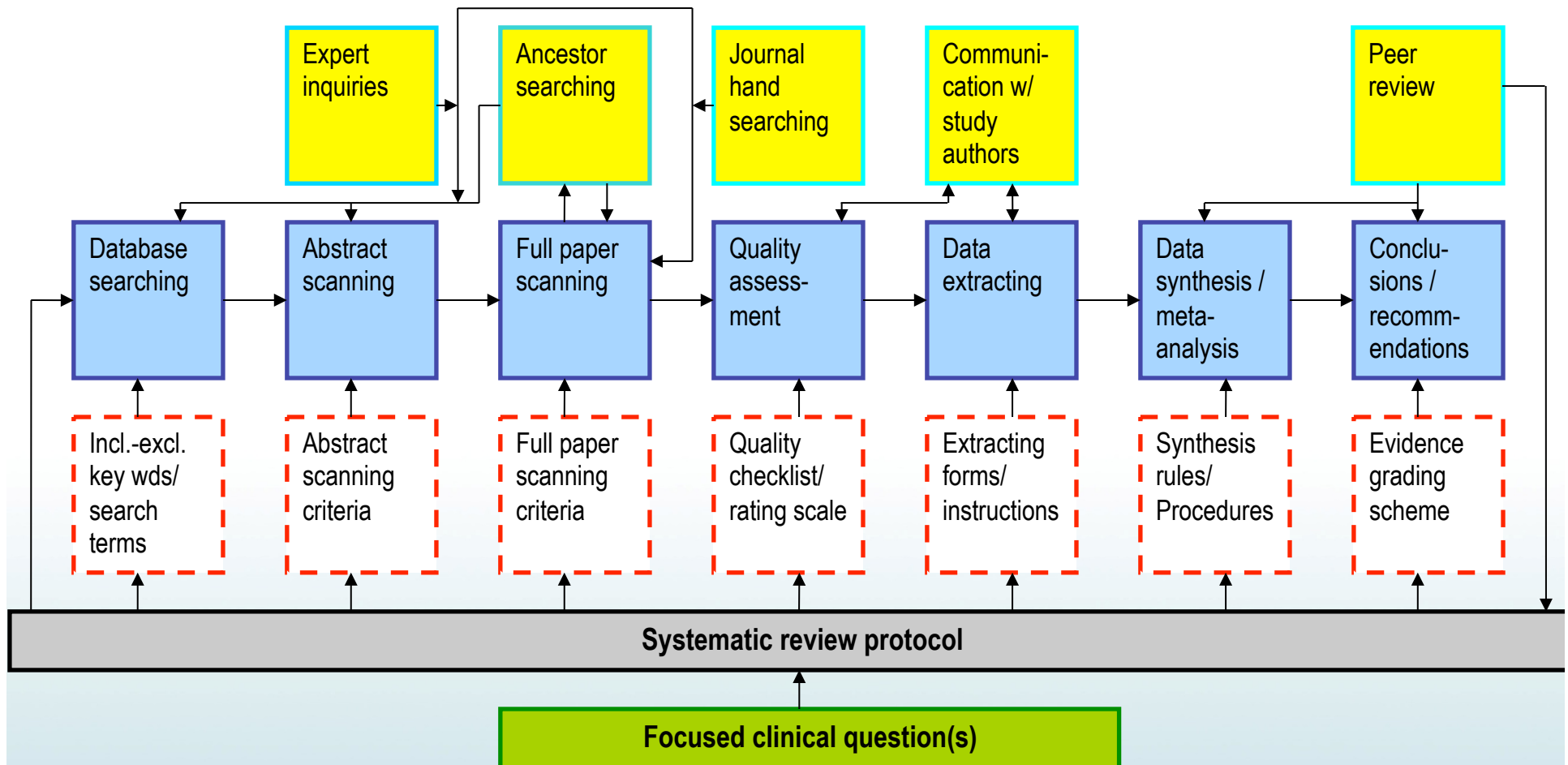
- An article or book published after examination of published material on a subject. It may be comprehensive to various degrees and the time range of material scrutinized may be broad or narrow, but the reviews most often desired are reviews of the current literature. The textual material examined may be equally broad and can encompass, in medicine specifically, clinical material as well as experimental research or case reports. State-of-the-art reviews tend to address more current matters. [...]

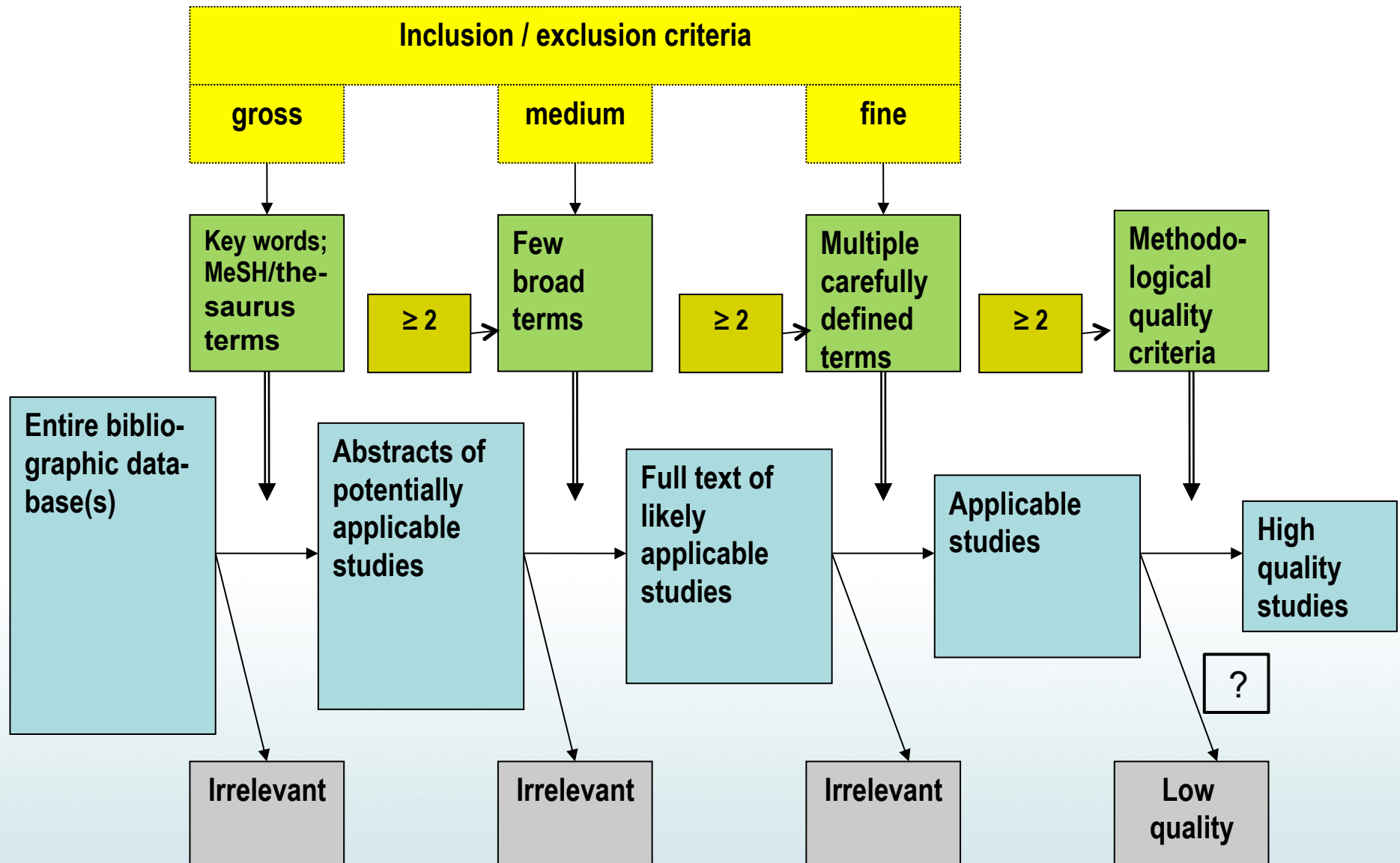
## AQASR Glossary: definition of “Systematic Review”

- A systematic review synthesizes research evidence *focused on a particular question* and following an *a priori protocol* to systematically *find* primary studies, *assess* them for quality, *extract* relevant information and *synthesize* it, qualitatively or quantitatively (meta-analysis).
- Systematic reviews *reduce bias* in the review process and *improve the dependability* of the answer to the question, through use of a protocol, extensive electronic and manual literature search, careful extracting of data and critical appraisal of individual studies.

Task Force on Systematic Review and Guidelines. (2013). *Assessing the quality and applicability of systematic reviews (AQASR)*. Austin, TX: SEDL, Center on Knowledge Translation for Disability and Rehabilitation Research.

## The steps in a systematic review: schematic overview of systematic review production and the link of the results to the reader's interests







Questions?

## Determinants of the quality of evidence – individual primary study

- Research Design (RCT, RCT with crossover, historical controls, pre-post study, single subject design, regression discontinuity design, time series, etc. etc.)
- Other design elements:
  - Quality of outcome measure(s)
  - Quality control on intervention (if any) (fidelity)
  - Blinding
    - Interventionist
    - Subjects/patients
    - Assessor
    - Statistician
  - Blocking, stratifying, matching

# Determinants of the quality of evidence – individual primary study

- Research implementation
  - Number of subjects recruited (in relationship to target number specified by power analysis)
  - Blinding failures
  - Attrition
  - Missing data (completely at random, etc.)
  - Sampling fluctuations
  - Per protocol vs intent-to-treat analysis
  - Deviations from analysis protocol (fishing expeditions)

# Determinants of the quality of evidence – entire body of relevant primary studies

- Number of studies
- Quality (grade) of individual studies
- Size of studies, individually and combined (N of subjects)
- Consistency of findings
- Effect sizes and 'average' effect size
- **Applicability to research question:**
  - Same population
  - Outcomes reported: patient-valued outcomes vs proxy measures (lab tests, etc.)
  - Feasibility of intervention

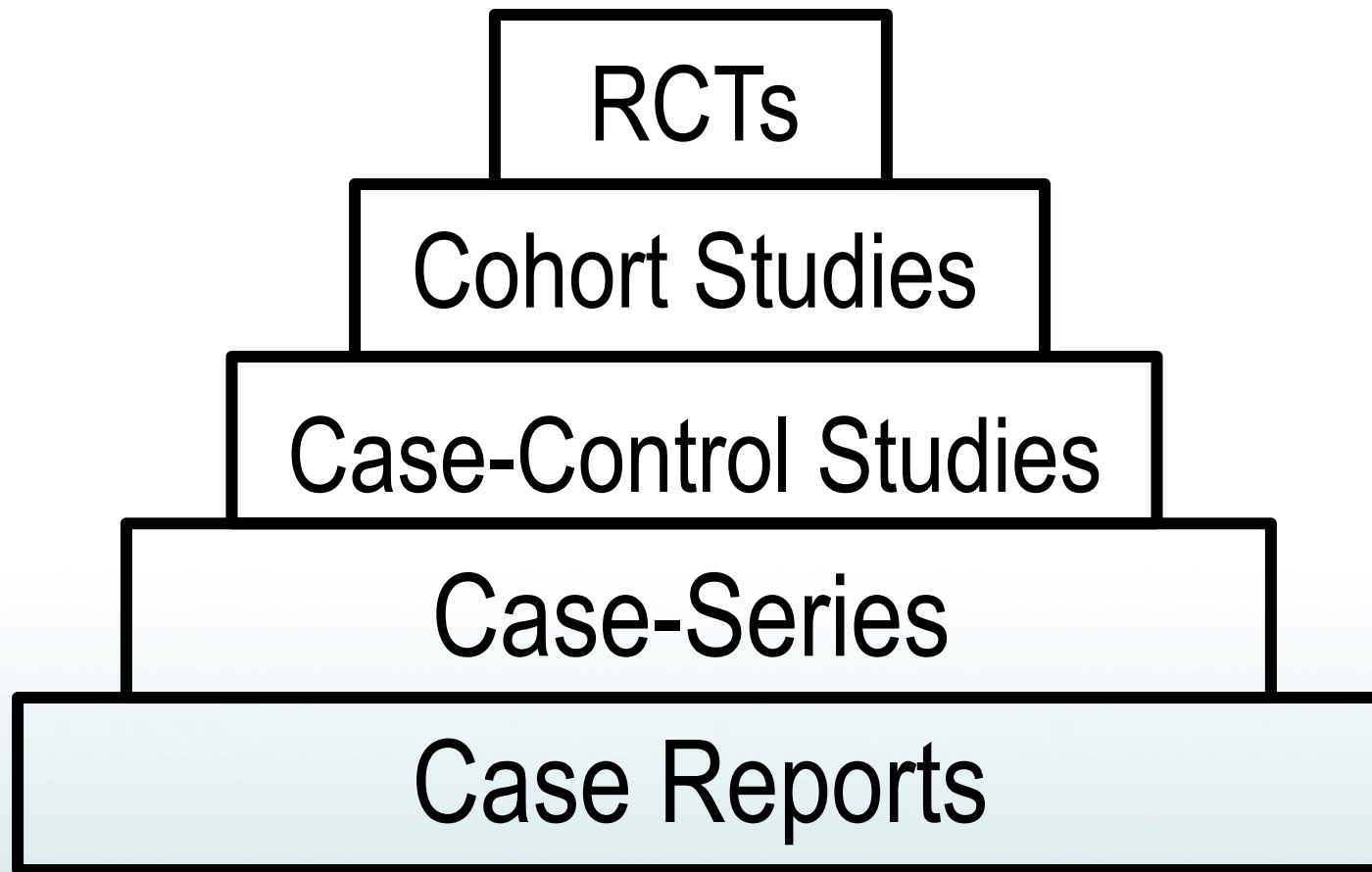


- For many decades, reviews had been qualitative
  - Authors picked the primary studies ‘randomly’
  - Authors weighted primary studies randomly
  - And drew conclusions that might correspond to their prior convictions/COIs/etc.
- First meta-analyses (in 1970s, in social sciences) might follow the same lack of rules
- Systematic reviewing ‘invented’ in medicine in the 1980s
- First classification of primary studies as to strength of evidence they offered by Sackett in 1986

## Sackett's levels of evidence hierarchy (1986)

- I RCTs with low false-positive ( $\alpha$ ) and low false-negative ( $\beta$ ) errors (high power)
- II RCTs with high false-positive ( $\alpha$ ) and/or high false-negative ( $\beta$ ) errors (low power)
- III Nonrandomized concurrent cohort comparisons between contemporaneous patients who did and did not receive Tx X
- IV Nonrandomized historical cohort comparisons between current patients who did receive Tx X and former patients (from the same institution or from the literature) who did not
- V Case series without controls

## A simple evidence hierarchy



## Sackett's relation between levels of evidence and grade of recommendation (1986)

grade of recommendation	criteria
A	supported by at least one, and preferably more, Level I RCTs
B	supported by at least one level II RCT
C	supported only by level III, IV or V evidence

(The grades apparently have no specific meaning; presumably a recommendation marked A is stronger than one marked B which is stronger than one marked C.

## A simple hierarchy surgically improved

RCTs

---

Cohort Studies

Case-Control Studies

Case-Series

Case Reports

## Cicerone et al.'s definitions of three levels evidence (2000)

- I Well designed, prospective, RCTs, with true randomization or with “quasi-randomized” assignment to treatment conditions, such as prospective assignment of subjects to alternating conditions
- II Prospective, nonrandomized cohort studies; retrospective, nonrandomized case-control studies; or clinical series with well-designed controls that permitted between-subject comparisons of treatment conditions, such as multiple baseline across subjects
- III Clinical series without concurrent controls, or studies with results from one or more single cases that used appropriate single-subject methods, such as multiple baseline across interventions with adequate quantification and analysis of results

## Cicerone et al.'s three levels of recommendations

### Practice Standards

Based on at least 1, well-designed Class I study with an adequate sample, or overwhelming Class II evidence, that directly addresses the effectiveness of the treatment in question, providing good evidence to support a recommendation as to whether the treatment be specifically considered for persons with acquired neurocognitive impairments and disability.

### Practice Guidelines

Based on well-designed Class II studies with adequate samples, that directly address the effectiveness of the treatment in question, providing fair evidence to support a recommendation as to whether the treatment be specifically considered for persons with acquired neurocognitive impairments and disability.

### Practice Options

Based on Class II or Class III studies, with additional grounds to support a recommendation as to whether the treatment be specifically considered for persons with acquired neurocognitive impairments and disability, but with unclear clinical certainty.

Cicerone KD, Dahlberg C, Kalmar K, et al. Evidence-based cognitive rehabilitation: recommendations for clinical practice. Arch Phys Med Rehabil. 2000 Dec;81(12):1596-615.



Questions?



# An RCT is an RCT is an RCT: the Jadad scale

(with thanks to Gertrude Stein)

1. is the study described as randomized?
  2. is the method of generating the randomization sequence appropriate?
  3. Is double blinding used?
  4. is the method of double binding appropriate?
  5. is there a description of withdrawals and dropouts?
- 5 items with 0 vs. 1 points > total scale score of 0 to 5 points

## An RCT is an RCT is an RCT: the PEDro scale

1. eligibility criteria were specified (not scored: external validity)
2. subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)
3. allocation was concealed
4. the groups were similar at baseline regarding the most important prognostic indicators
5. there was blinding of all subjects
6. there was blinding of all therapists who administered the therapy
7. there was blinding of all assessors who measured at least one key outcome

## An RCT is an RCT is an RCT: the PEDro scale

8. measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups
9. all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analyzed by “intention to treat”
10. the results of between-group statistical comparisons are reported for at least one key outcome
11. the study provides both point measures and measures of variability for at least one key outcome
  - 10 items with 0 vs. 1 points > total scale score of 0 to 10 points

Physiotherapy Evidence Database - <http://www.pedro.org.au/english/downloads/pedro-scale>

# Judging the quality of RCTs (and other study designs)

- Checklist
  - Identify weaknesses of study design and implementation
  - Use individual items in evidence synthesis
    - Eliminating studies
    - Weighting studies
    - Sensitivity analysis of impact of study quality on conclusions

# Judging the quality of RCTs (and other study designs)

- Rating scale
  - Quantify strength of individual design elements
  - Combine item scores to calculate study quality score
  - Use scale score in evidence synthesis
    - Eliminating studies (e.g. PEDro < 6)
    - Weighting studies
    - Sensitivity analysis of impact of study quality on conclusions
  - How to combine items?
    - Adding (sumscore)
    - Multiplying
    - Lowest item



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/s3/1681053/Evaluation-Webinar-Series-Session-1](http://www.surveygizmo.com/s3/1681053/Evaluation-Webinar-Series-Session-1)