Assessing the Quality and Applicability of Systematic Reviews (AQASR)

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

Session 3 – February 5, 2014

An online workshop sponsored by SEDL's Center on

Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

Objectives:

- Delineate steps and issues in the development of systematic reviews

- Introduce *Assessing the Quality and Applicability of Systematic Reviews (AQASR)* *(© SEDL/NCDDR 2011)*

- Describe how AQASR can be used in evaluating whether a particular systematic review can be trusted to provide an unbiased, reliable answer to one's (clinical, research, policy) question

Objectives:

- Review the various sections of AQASR and the items in each section
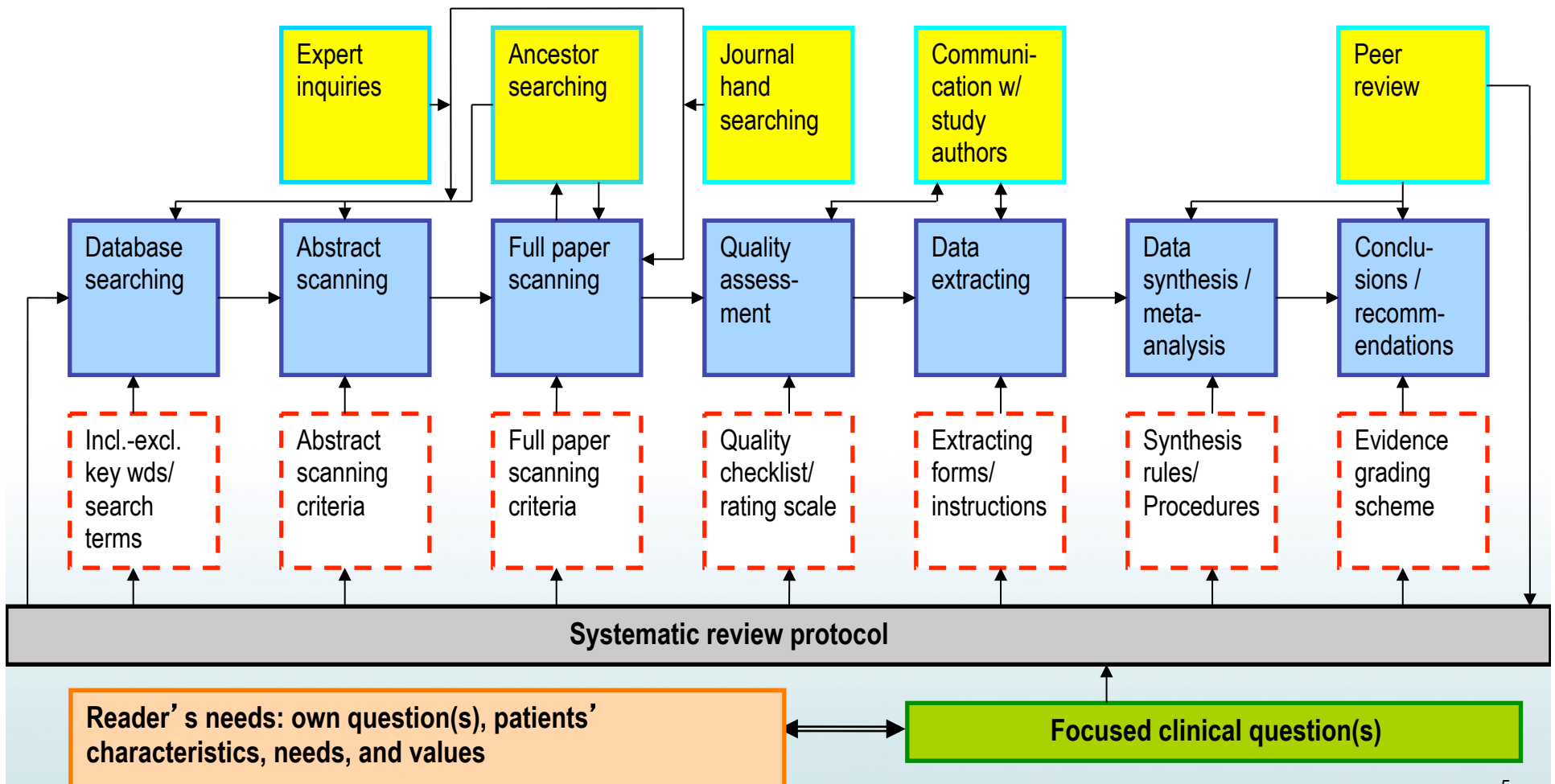- Apply the instrument to several systematic reviews to increase familiarity with its elements and application

Questions?

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

AQASR has questions on the steps all systematic reviews have in common:

- The focused clinical question (6)
- Systematic review protocol (5)
- Literature searches (16)
- Scanning of abstracts and full papers (8)
- Assessment of the quality of the primary studies (6)
- Extracting data (4)
- Synthesizing the data qualitatively (7)
- Drawing conclusions, making recommendations (7)

And (used by some)

- Synthesizing the data quantitatively (meta-analysis) (7)

In addition, AQASR has questions relevant to the topic of the systematic review:

- Intervention/prevention (13)

- Diagnostic procedure (8)

- Measurement instrument (10)

- Prognosis (6)

- Economic evaluation (7)

Questions?

METHODOLOGICAL QUALITY ASSESSMENT AND USE

MQ1. Were studies reviewed for methodological quality?

- **Look for:**
  - A list of criteria used to evaluate methodological quality of the primary studies
  - Names of quality assessment instruments, e.g. Jadad scale, PEDro, Black and Downs, AAN approach, etc.
  - Entries in an evidence table of quality grades or scores

MQ1. Were studies reviewed for methodological quality?

- **Rationale:**
  - A clear statement of methodological quality criteria helps users of reviews determine the thoroughness of the review and the usefulness of the review and its conclusions for their own work.
  - Reference to well-established criteria may be sufficient, such as those of the AGREE Collaboration, the Campbell Collaboration, the American Academy of Neurology, the Agency for Healthcare Research and Quality, or the Cochrane Collaboration

MQ2. Was the instrument for assessing study quality identified and presented?

- **Look for:**
  - A reference to an existing instrument or the description of an ad-hoc one
  - An explanation justifying the selection of a study quality review instrument

MQ2. Was the instrument for assessing study quality identified and presented?

- **Rationale:**
  – Several well-established checklists have been developed, such as those of Jadad, Black and Downs.

  – Reporting checklists such as CONSORT sometimes are also used as methodological quality checklists or even rating scales.

  – Adoption of an established review instrument assures that the criteria have been given careful consideration by an independent organization.

MQ3. Were the results of quality assessment used, and was this use justified?

- **Look for:**
  - A summary of the quality assessment results
  - A description of how quality ratings were used
  - A justification of this use of the results

MQ3. Were the results of quality assessment used, and was this use justified?

- **Rationale:**
  - Quality assessment summaries can be reported in tabular and narrative form.
  - Readers should be able to identify key quality aspects of studies quickly and to understand the reviewers' rationale.
  - The review should also state how the evaluations of quality were used (delete poor quality research, weight studies by quality in a meta-analysis, etc.), and why this use was appropriate.

MQ4. Was study quality scored by ≥2 persons independently? Is agreement level reported? Was there a procedure for developing consensus?

- **Look for:**
  - A description of independent rating of study quality by more than one reviewer
  - A discussion about level of agreement between raters
  - The method used to assess agreement (kappa, percent agreement, ICC) and the value(s) found

16

MQ4. Was study quality scored by ≥2 persons independently? Is agreement level reported? Was there a procedure for developing consensus?

- **Rationale:**

  – The description in primary studies of the methods used is often incomplete or ambiguous. Individuals may have idiosyncratic ways of scoring the quality of studies, ways that reflect bias or carelessness.

  – Including two or more independent reviewers helps assure that quality scores are reliable.

MQ5. Is nature and training of study quality scorers/reviewers specified?

- **Look for:**
  - A statement about the nature and training of study reviewers

MQ5. Is nature and training of study quality scorers/reviewers specified?

- **Rationale:**
  - After review criteria and procedures are developed, reviewers need training to assure they understand and apply study quality criteria consistently.
  - A statement about reviewer training helps researchers replicate the findings.

MQ6. Was bias or potential bias in reviewed studies addressed and presented?

- **Look for:**
  - Comments regarding the risk of bias in reviewed studies (based on an bias assessment instrument or otherwise), and when judged to be more than minimal, comments regarding the consequences of bias

MQ6. Was bias or potential bias in reviewed studies addressed and presented?

- **Rationale:**
  - It can require considerable experience and a high level of suspicion to detect studies that are not systematic in randomizing cases, delivering an intervention, monitoring the fidelity of the intervention, assessing the outcomes and conducting appropriate analyses.
  - Reviewer attention to these issues helps assure that poorly designed or implemented studies are noted and given appropriate weight in the synthesis of evidence.

Methodological assessment: Ruivo

- Data extracted on ("among others"):
  - Trial design
  - Adequacy of randomization
  - Degree of allocation concealment
  - Blinding
- By one person only

Methodological assessment: Ruivo

- Methodology data used to complete
  - PEDro
  - Romeiser Logan scale for SSDs
- Studies included if score was
  - ≥5/10 on PEDro
  - ≥7/14 on Romeiser Logan scale
- 21 of 26 papers accepted for systematic review
  - 4: insufficient quality
  - 1: has trial design and protocol only

Quality of study vs quality of thing studied

quality of assessment measure	quality of the study investigating the assessment measure		
	good	average	poor
good			
average			
poor			

Similar issue for all other systematic review topics: good / average / poor study to investigate:

- Diagnosis
  - Poor instrument (low sensitivity and/or specificity)
  - Good instrument (good sensitivity and specificity)

- Treatment
  - Good (strong, effective) treatment
  - Poor (ineffective, useless, harmful treatment

- Prognosis
  - X and Y are strong predictors of Z
  - X and Y are not predictive of Z

Study design aspects that make for a poor study (for Tx and/or Dx and/or Px and/or assessment measure development purposes)

- Small sample
- Too homogeneous a sample
- Nonrepresentative sample
- Strong random attrition
- Selective attrition
- Failure to blind
- Failure to randomize
- No concealment
- Etc.

Systematic review of measures of concept X (in population Y)

study #	quality assesment	measure 1	measure 2	measure 3	…..	measure j
1	good	x				
2	average	x				
3	average	x				
4	poor	x				
5	poor					
6	average	x	x			
7	poor		x			
8	poor		x	x		
9	average		x	x		
…						
k	average					x
		summary for 1	summary for 2	summary for 3		summary for j

standards & criteria
- validity
- reliability
- sensitivity
- …

recommendations
- overall best:
- if a short measure is needed:
- if a self-administered measure is needed:
- etc.

Methodological assessment: Dobson

- 1: Rating of study quality
  - COSMIN tool, with 4-point scoring of items
    - 10 sections with 5-18 items each
    - Sample size items not considered
    - Scored excellent, good, fair, poor each
    - Lowest item score in section taken as section score
  - Two raters, experienced in COSMIN
    - Not specified what was done in case of disagreement

Methodological assessment: Dobson

- 2: Rating of performance-based measures of physical function
  - Up to 9 psychometric qualities (table 1)
    - Reliability: 3
    - Validity: 5
    - Responsiveness: 1
  - Each rated + (good), - (poor), ? (undetermined)
    - NOTE: dichotomy: grey not allowed
  - Based on prespecified quality criteria
  - Not indicated whether done by two people

Methodological assessment: Dobson

- 3: integration (best evidence synthesis) considers:
  - Quality of studies (COSMIN)
  - Rating of 9 psychometric properties (+, -, ?) and consistency of ratings
  - Number of studies for each of 9 properties
  - Cumulative sample size of studies
  - Cumulative sample quality
  - (Studies with 'poor' COSMIN rating not considered)

Methodological assessment: Dobson

		Dobson integration of study quality and measure quality			
		quality of study design and implementation			
level	symbol	excellent	good	fair	poor
strong	+++	>1	>2		
moderate	++		>1	>2	
limited	+			1	
conflicting	±				
unknown	?				>1

31

DATA EXTRACTION

DA1. Is an extracting form and syllabus described?

If so, is pilot-testing of the form/ syllabus described?

- **Look for:**
  - A data extraction form created prior to beginning the process of extracting information from articles reviewed
  - The mention of a syllabus, a set of explicit, clear instructions to ensure that all reviewers completed the form in the same manner
  - Statement that the form/syllabus was pilot tested
  - A brief (likely one sentence) statement that reviewers practiced extracting data from a few articles prior to beginning the actual review

33

DA1. Is an extracting form and syllabus described?

If so, is pilot-testing of the form/ syllabus described?

- **Rationale:**
  - If reviewers did not follow standard procedures in extracting data, data collected may be incomplete, inaccurate or biased.
  - This would be similar to conducting a primary study in which different data collectors used different procedures for collecting study outcomes.
  - The inconsistency between data collectors would be likely to invalidate the study.
  - Practice with the data collection form and syllabus provides the authors with an indication of whether the form can be completed reliably by all reviewers.
  - If this is not the case, changes can be made prior to beginning the actual review.

34

DA2. Were study data extracted by two or more persons independently? Is agreement measure and level reported?

- **Look for:**

  - A brief statement that all articles, or at least a large sample of articles, were reviewed and data extracted by at least two reviewers.

  - A statement that reviews were completed independently.

  - Information quantifying the agreement between the independent reviewers, e.g., using percent exact agreement or kappa

DA2. Were study data extracted by two or more persons independently?

- **Rationale:**

  – Prior training and practice during the piloting of the data extraction form should have minimized inter-reviewers differences.

  – Data extraction is sometimes a matter of judgment so that different reviewers may have dissimilar results.

  – Having each article reviewed by multiple reviewers ensures that one reviewer's biases will not overly affect the overall review findings.

  – Completion of reviews independently helps ensure that one reviewer does not simply defer to the other.

36

DA3. Is there a description of how disagreements between data extracters were resolved?

- **Look for:**
  - An explicit statement of how disagreements were resolved.

DA3. Is there a description of how disagreements between data extracters were resolved?

- **Rationale:**
  - Disagreements between the two independent reviewers should be resolved in a standard way to minimize bias so that the extracted data best represents the "truth" of the evidence produced by the studies.
  - Common ways of resolving disagreements include a discussion between the original reviewers to try to reach an consensus and obtaining additional review(s) to clarify which of the original reviewers was "correct."

DA4. Is the nature and training of the data extracters specified?

- **Look for:**
  - A statement of the qualifications reviewers brought to the process
  - Training conducted after reviewers were identified to ensure that they followed the a priori protocol for reviewing studies and extracting data

DA4. Is the nature and training of the data extracters specified?

- **Rationale:**
  - The quality of the results is dependent of the expertise of those conducting the data extraction.
  - For most systematic reviews, both methodology specialists and clinical specialists should be used.
  - Training on the protocol may coincide with efforts to pilot the data extracting form and syllabus, or is separate from it because the form and instructions had been fine-tuned previously.

Questions?

For the next session (Feb. 19):

- Read manual sections/AQASR questions on
    - Qualitative synthesis
    - Meta-analysis
    - Discussion
- Read Polisena et al., with a focus on the sections/ information on meta analysis

Polisena et al.

Polisena, J., Tran, K., Cimon, K., Hutton, B., McGill, S., & Palmer, K. (2009). Home telehealth for diabetes management: a systematic review and meta-analysis. *Diabetes, Obesity and Metabolism, 11(10)*, 913–930.

http://onlinelibrary.wiley.com/doi/10.1111/j.1463-1326.2009.01057.x/abstract

For the next session:

- Identify three systematic reviews you would like to use/might use in your work, one each focusing on

    – Prevention

    – Diagnosis

    – Prognosis

- Not too long; of good, middling or poor quality

- Of presumed general interest (not too specialistic)

- Send ABSTRACTS to Joann Starks

- We will pick one of each to use in later sessions

Thank you for participating!

Wrapping Up

We invite you to:

- Provide your input on today's session

- Share your ideas for future sessions

- Describe special needs you may have

- PLEASE CONTACT US:

joann.starks@sedl.org