Assessing the Quality and Applicability of Systematic Reviews (AQASR)

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

Session 4 – February 19, 2014

An online workshop sponsored by SEDL's Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR)

Funded by NIDRR, US Department of Education, PR# H133A120012

Objectives:

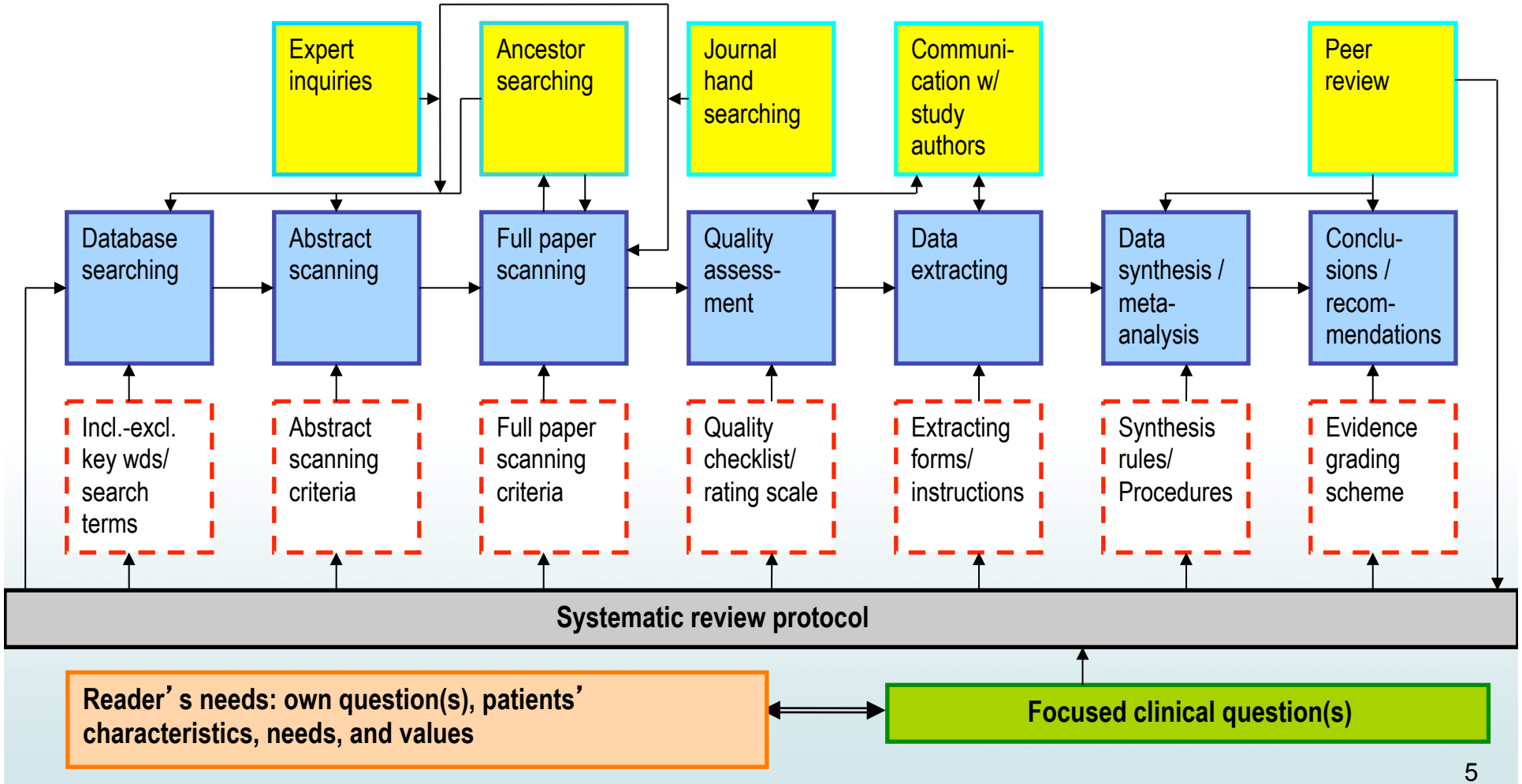
- Delineate steps and issues in the development of systematic reviews
- Introduce *Assessing the Quality and Applicability of Systematic Reviews (AQASR)* *(© SEDL/NCDDR 2011)*
- Describe how AQASR can be used in evaluating whether a particular systematic review can be trusted to provide an unbiased, 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?

QUALITATIVE SYNTHESIS

QS1. Did the review include the right type of study (relevancy to the question)?

- **Look for:**
  - Correspondence between studies included and the studies called for by the clinical question in terms of:
    - Clinical/scientific domain
    - Research design
    - Sample characteristics (age, sex, co-morbidities, etc.)
    - Time period, political/geographic area etc.
    - Other relevant characteristics of the studies and subjects

QS1. Did the review include the right type of study (relevancy to the question)?

- **Rationale:**
  - The clinical question can only be answered if the systematic review finds and summarizes the right type of evidence
  - A shortage of evidence is never a good justification for shifting to other diagnostic groups, outcomes, study types, etc.
  - Unless: considered a priori in the protocol:
    - Designs lower than Class 1
    - 'Indirect evidence'

QS2. Is the method for data synthesis (aggregating evidence across studies) described?

- **Look for:**
  - A statement as to whether or not the data will be/are described descriptively or will be/are combined in a meta-analysis
  - IF NO META-ANALYSIS IS PERFORMED: A description of the methods and criteria employed to combine the results of various studies and draw conclusions from their joint findings

QS2. For instance: AAN methodology

- Step 1: rate strength of each study identified as Class I (strong) to Class IV (weak: case study, opinion)
- Formulate recommendations titrated to the quantity and strength of available studies
  - **High confidence** (anchor: ≥ 2 Class I studies—corresponds to a "highly likely" conclusion)
  - **Moderate confidence** (anchor: 1 Class I study or ≥ 2 Class II studies—corresponds to a "likely" conclusion)
  - **Low confidence** (anchor: 1 Class II study or ≥ 2 Class III studies—corresponds to a "possibly" conclusion)
  - **Very low confidence** (anchor: < 2 Class III studies—corresponds to an "insufficient evidence" conclusion)
  - (**Conflicting evidence** – no conclusion)

QS2. Is the method for data synthesis (aggregating evidence across studies) described?

- **Rationale:**
  - The primary studies may be more or less heterogeneous.
  - A narrowly based question will lend itself better to pooling of the data (and meta-analysis) while a more broadly based question will lend itself to descriptive tables followed by synthesis into what the entirety of the literature shows.

QS3: Were the findings (from original studies) combined appropriately and the data analyzed appropriately?

- **Look for:**
  - Descriptive tables ('evidence tables') that summarize the salient points of each study
  - Forest plots or L'Abbe plots used to illustrate the treatment effects and confidence intervals for each study

QS3: Were the findings (from original studies) combined appropriately and the data analyzed appropriately?

- **Rationale:**
  - In many cases, the studies are sufficiently heterogeneous, such as different subject populations, interventions, or outcome measures, that only a qualitative analysis is possible

QS4: Were the studies similar enough to combine? (same subjects, same or similar intervention, same or comparable outcomes?)

- **Look for:**
  - The decision to pool results based upon descriptive (subjective) rather than statistical criteria
  - Acknowledgment of heterogeneity measures, and arguments for combining

QS4: Were the studies similar enough to combine? (same subjects, same or similar intervention, same or comparable outcomes?)

- **Rationale:**
  - Systematic reviews should seek to answer a focused question, which drives the pooling of results
  - Pooling is often a judgment call of the authors, where forest of L'Abbe plots should be used to assess consistency of the results
  - Sufficient heterogeneity may preclude either statistical or qualitative synthesis

QS5. Were the results clearly reported and in sufficient detail?

- **Look for:**
  - Qualitative descriptions of the studies in the text of the review
  - Supporting tables that summarize each study that was included
  - Forest plots, L'Abbe plots or other graphs may also be used to graphically illustrate the main effects of each study

QS5. Were the results clearly reported and in sufficient detail?

- **Rationale:**
  - There should be sufficient detail for the reader to judge homogeneity or heterogeneity of the studies and relevance to the core question
  - Tables should clearly indicate which studies found similar results
  - Because systematic reviews may produce voluminous tables and other materials, part of the information may be published on the web or by available by request from the authors

QS6. Was any sensitivity testing reported (subgroup analyses; best-studies analysis, etc.)?

- **Look for:**
  - A rationale for conducting additional analyses, such as:
    - A summary of the heterogeneity of the studies
    - Imprecision of study results (large confidence intervals)
    - Rationale for examining "best studies" which should be justified in terms of the clinical question

QS6. Was any sensitivity testing reported (subgroup analyses; best-studies analysis, etc.)?

- **Rationale:**
  - Prior to the review, there should have been a decision made about how to combine the data
  - BUT, in some justified cases (such as greater level of homogeneity than expected and it is not appropriate to pool all of the studies), additional analyses could be conducted

Questions?

Ruivo

- A ‘semiquantitative/qualitative’ description of the appraised studies is reported, as it was not possible to use formal techniques of meta-analysis because of the wide diversity in study designs, interventions, and outcome measures.
- Evidence tables are provided separately for laboratory studies (ranked by the intensity of the AVG, according to the reported/ estimated EE) and field intervention studies (with RCTs ranked by whether the AVG protocol resulted in clinical improvement)
- Text (in Results) summarizes tables using counts, percentages, means, separately for laboratory studies and field intervention studies

Ruivo

- Tables provide information on
  - Sample size and characteristics (sex, age, health)
  - Design (and methodological quality score)
  - AVG name and duration
  - AVG intensity (and difficulty)
  - Experimental/alternative conditions
  - AVG metabolic equivalents
  - Outcome measures
  - Key results
  - Gender differences

META ANALYSIS

MA1. Is it specified how missing values are handled? Is this appropriate?

- **Look for:**
  - A statement on how reports of primary studies with missing data were handled

MA1. Is it specified how missing values are handled? Is this appropriate?

- **Rationale:**
  - Papers and other primary research reports may miss crucial information needed for a meta-analysis – e.g. N of cases, standard deviations corresponding to means, etc. This may be handled by omitting the report, estimating from other studies, estimating conservative values, etc. Any decision should be justifiable.

MA2. Was the heterogeneity of studies in terms of outcomes analyzed and reported? If the studies were heterogeneous, was the random effects model used?

- **Look for:**
  - A formal test of heterogeneity, using such measures as Cochran's Q or the I^2
  - A statement on the model (fixed or random effects) used in combining study findings

MA2. Was the heterogeneity of studies in terms of outcomes analyzed and reported? If the studies were heterogeneous, was the random effects model used?

- **Rationale:**
  - If the effect sizes of the various studies to be combined are very similar, as shown using a formal test, a fixed effects model for combining can be used. If they are heterogeneous, the random effects model should be used, unless they are so dissimilar ("apples and oranges") that only a qualitative synthesis makes sense.

MA3. How are results expressed (odds ratio, relative risk, etc.) in the primary studies and in the systematic review?

- **Look for:**
  - A statement or column heading or similar indications as to what the "common denominator" of the studies that are being combined is

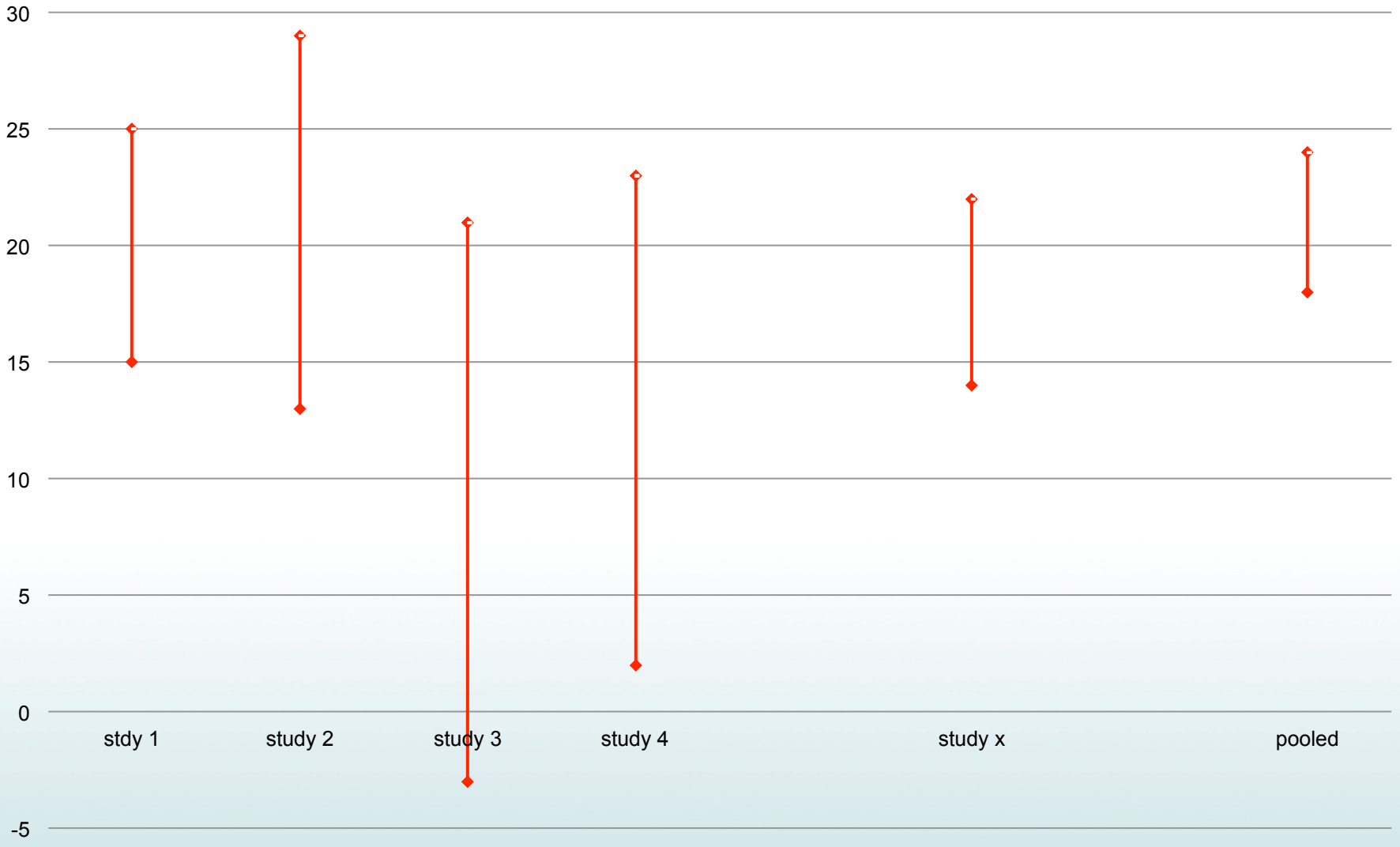
MA3. How are results expressed (odds ratio, relative risk, etc.) in the primary studies and in the systematic review?

- **Rationale:**
  - Whatever the effect size measures used in the original studies (risk difference, odds ratio, risk ratio, means and standard deviations, etc.), the systematic reviewer has to "translate" them all to a common denominator (based on information in the original reports) in order to combine them. Sometimes they cannot be translated without making assumptions; best is when all primary studies used the same outcome measures.

MA4. How large is the pooled effect? Are confidence intervals reported? How precise are the results? Would practical decisions be different/same at the low vs. high end of the confidence interval?

- **Look for:**
  - An effect size for the pooled studies
  - A confidence interval around this effect size
  - The distance between the high (low) end of the CI and the point of "no difference"

Confidence interval for individual studies and combined

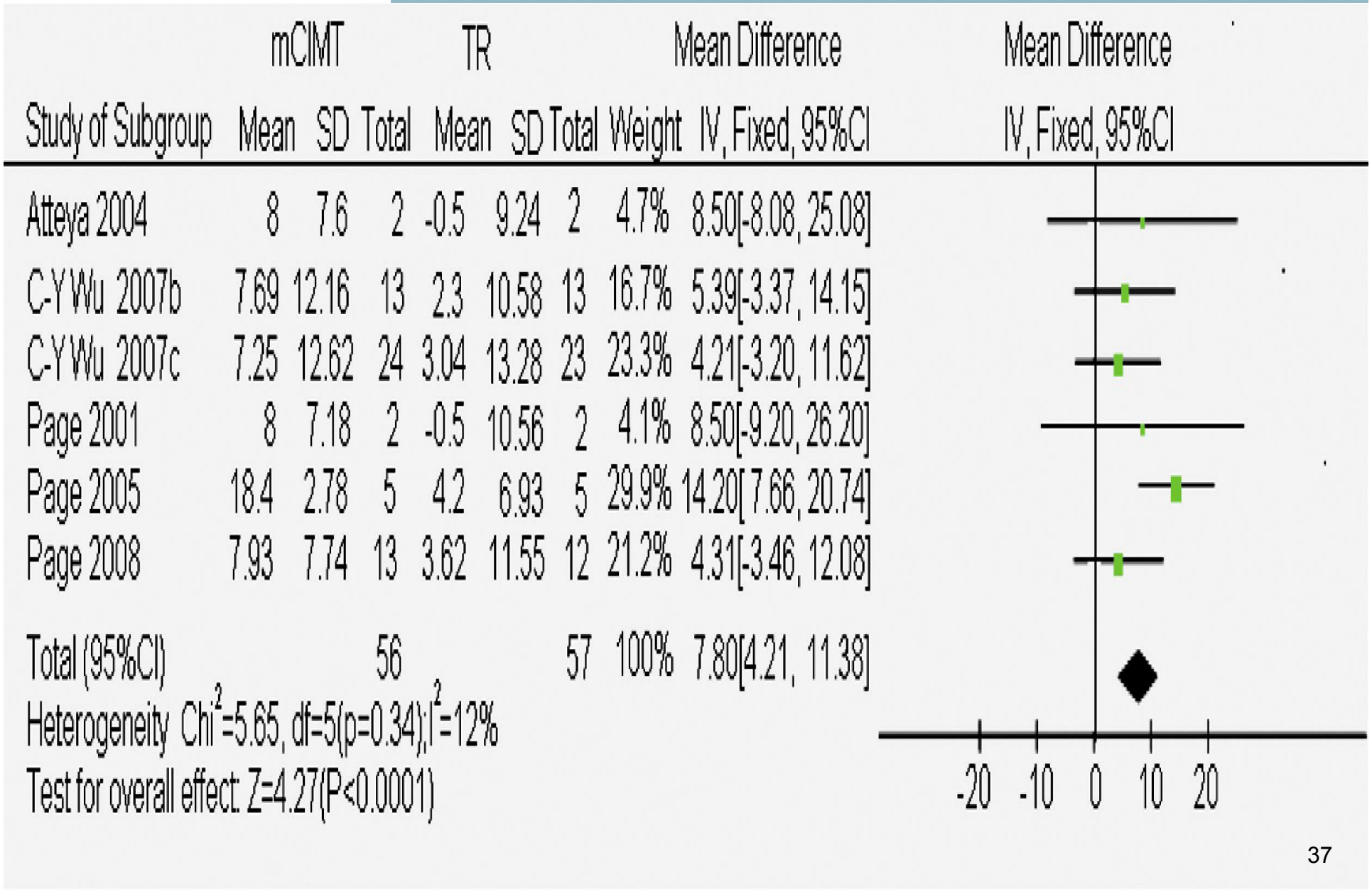


MA4. How large is the pooled effect? Are confidence intervals reported? How precise are the results? Would practical decisions be different/same at the low vs. high end of the confidence interval?

- **Rationale:**
  - The (95%) confidence interval specifies the likely range of values in which the true effect is to be found.
  - When there are few or small studies to be combined, or when study outcomes are heterogeneous, the confidence interval may be rather wide.
  - Clinicians may make different decisions based on whether they assume the effect is at the high vs. at the low end of this range. Because both extremes are equally likely (or unlikely), the authors ought to carefully consider the implications of all possible values in the range.

MA5. Are appropriate tables and graphs provided?

- **Look for:**
  - A table and/or forest plot offering the effect sizes (plus confidence intervals) for all individual studies and the studies combined


mCIMT
TR
Mean Difference
Mean Difference
Study of Subgroup Mean SD Total Mean SD Total Weight IV, Fixed, 95%CI
IV, Fixed, 95%CI
Atteya 2004 8 7.6 2 -0.5 9.24 2 4.7% 8.50[-8.08, 25.08]
C-Y Wu 2007b 7.69 12.16 13 2.3 10.58 13 16.7% 5.39[-3.37, 14.15]
C-Y Wu 2007c 7.25 12.62 24 3.04 13.28 23 23.3% 4.21[-3.20, 11.62]
Page 2001 8 7.18 2 -0.5 10.56 2 4.1% 8.50[-9.20, 26.20]
Page 2005 18.4 2.78 5 4.2 6.93 5 29.9% 14.20[7.66, 20.74]
Page 2008 7.93 7.74 13 3.62 11.55 12 21.2% 4.31[-3.46, 12.08]
Total (95%CI) 56 57 100% 7.80[4.21, 11.38]
Heterogeneity Chi²=5.65, df=5(p=0.34); I²=12%
Test for overall effect: Z=4.27(P<0.0001)
-20 -10 0 10 20

MA5. Are appropriate tables and graphs provided?

- **Rationale:**
  - A table summarizing all data and especially a forest plot offer an "at a glance" summary, with the value and confidence interval for all studies as well as their combination lined up, typically in relationship to a "no effect" line.
  - Readers should investigate these tables/graphs for their "reasonableness" and support for the conclusion drawn by the authors.

MA6. Were subgroup analyses (if any) specified a priori?

- **Look for:**
  - A statement that subgroup analyses were considered on beforehand, either absolutely or depending on heterogeneity testing results

MA6. Were subgroup analyses (if any) specified a priori?

- **Rationale:**
  - Doing pre-specified separate subgroup analyses is justified, and feasible if the number of studies is large enough.
  - However, especially if the results of the primary studies are heterogeneous, there is a temptation to use ad-hoc analyses to identify factors that might explain heterogeneity. As is the case with all post-hoc analyses, the results of these efforts are suspect.
  - Meta-regression, a method of combining studies based on continuous variables rather than dichotomies, similarly should be pre-planned.
  - If these are not pre-specified, the findings at best are suggestive and need to be confirmed by new large primary studies or a systematic review of new primary studies.

MA7. Is lack of power considered? I.e. was a prospective power analysis done to assess whether the combined studies have enough cases, calculated on the basis of a minimally acceptable effect size?

- **Look for:**
  - A power analysis, performed before or possibly after completion of the meta-analysis

MA7. Is lack of power considered? I.e. was a prospective power analysis done to assess whether the combined studies have enough cases, calculated on the basis of a minimally acceptable effect size?

- **Rationale:**
  - Just like a primary study may lack power to demonstrate the effect of an intervention or the utility of a prognostic variable, so the studies combined in a meta-analysis may.
  - Especially in rehabilitation, where studies tend to be few and small, this may occur. When the conclusion of the meta-analysis is one of “no effect”, a power analysis should have been done or be done to make sure this conclusion can be relied on not to involve a type-II error.

Polisena et al.

- Used medium and better quality RCTs and observational studies
- MA1: Missing values:
  - Missing data were imputed for RCTs
  - Data imputation was not conducted for observational studies because of the variation in designs and inherent risk of bias
- MA2: Heterogeneity and model used:
  - Where the quantitative pooling of results was appropriate ($I^2 <$ 50%), the random effects model was used (HTM vs. UC)
  - If statistical heterogeneity was substantial ($I^2 \geq 50\%$; TS vs. UC) attempts were made to explain heterogeneity by means of subgroup analyses (none actually provided)

Polisena et al.

- MA3: Expression of results
  - Mean and standard deviation for glycated hemoglobin in Figure 2
- MA4: pooled effect size and confidence interval
  - Figure 2 reports 95% confidence interval for pooled effect size
  - Confidence interval is narrow (-0.35 to -0.08); no concern about different decisions at low vs. high end
- MA5: Appropriate tables and graphs
  - Evidence table with:
    - Design and length of follow-up
    - Study quality score: claimed but missing (yet provided in text)
    - Comparison arms (n of cases, patient baseline characteristics)
    - Description of intervention and comparator
  - Text narration of outcomes that were not pooled
  - Table and forest plot for glycated hemoglobin (Figure 2)

Polisena et al.

- MA6: Subgroup analyses
  - Subgroup analysis to explore statistical heterogeneity promised but not reported
  - Sensitivity analysis exploring effect size differences between RCTs and observational studies promised but not reported

DISCUSSION OF FINDINGS AND RECOMMENDATIONS

DI1. Are study limitations discussed (e.g. publication bias, strength of studies, decisions on synthesis)

- **Look for:**
  - A subsection of the discussion section labeled ‘study limitations’
  - One or more paragraphs in the discussion section that address limitations
  - Occurrence of such terms as publication bias, selective outcome reporting, attrition bias, funding bias

DI1. Are study limitations discussed (e.g. publication bias, strength of studies, decisions on synthesis)

- **Rationale:**
  - Authors of good reviews are aware of the weaknesses of the materials they had to work with and the impact of decisions they made, such as a crucial decision that may have increased the effect sizes of an intervention
  - An informative discussion of effects on findings and conclusions of selective publication adds to confidence in the systematic review

DI2. Was publication bias assessed? Were other biases assessed?

- **Look for:**
  - A statement that all studies that met inclusion/ exclusion criteria were considered in the review, including
    - Negative outcomes (publication bias)
    - Unaccounted loss to follow-up (attrition bias)
    - Funding from commercial interests (funding bias)
  - Presentation of a funnel plot to assess selection bias of primary studies

DI2. Was publication bias assessed? Were other biases assessed?

- **Rationale:**
  - Tendency for negative findings not to be published
  - Tendency to focus only on significant outcome measures (within-study publication bias)
  - Loss to follow-up not adequately addressed (attrition bias)
  - Funding by commercial interests can favor the studied intervention and report fewer harms

DI3. Are the results interpreted in light of the totality of available evidence? Are alternative considerations/explanations for the results considered, e.g. publication bias?

- **Look for:**
  - A balanced discussion reflecting that even strong support by most studies reviewed needs to be qualified in terms of other studies
  - Considered of plausible alternative explanations for the results

DI3. Are the results interpreted in light of the totality of available evidence? Are alternative considerations/explanations for the results considered, e.g. publication bias?

- **Rationale:**
  - In most cases, support is divided or some primary studies are methodologically weak
  - Conclusions need to be qualified in terms of circumstances
  - For intervention studies, publication bias (studies that failed to support the intervention did not make it into print) is a valid concern

DI4. Is the generalization of the conclusions appropriate?

- **Look for:**
  - Recommendations that do not go beyond the type of subjects, interventions, health care/rehabilitative systems, etc. that were included in the primary studies that were reviewed

DI4. Is the generalization of the conclusions appropriate?

- **Rationale:**
  - Since systematic reviews are typically based on multiple studies with differing characteristics, the conclusions are likely more suited for generalizing
  - But careful consideration of the match between the studies reviewed and the conclusions should be made
  - To apply to a patient population, the clinician should ask "do my patients/clients fit this profile?

DI5. Are the results clinically meaningful in terms of the focused clinical question that was the basis for the review?

- **Look for:**
  - A paragraph in the discussion that addresses how and to what degree the results provide an answer to the clinical question that led to the review

DI5. Are the results clinically meaningful in terms of the focused clinical question that was the basis for the review?

- **Rationale:**
  - A good review should be able to provide some guidance to the clinician, unless absolutely no primary studies were identified
  - Refusal to make recommendations, however carefully worded, is not helpful to the clinician
  - This is especially the case in rehabilitation where RCTs are not common

DI6. If there were earlier systematic reviews in this area: do the authors discuss similarity or differences in findings and try to explain the differences?

- **Look for:**
  - A reference to other reviews in introduction or discussion
  - A paragraph in discussion section specifying similarities/ differences
  - If discrepancies: a paragraph in the discussion explaining why there were differences, suggest plausible reasons

DI6. If there were earlier systematic reviews in this area: do the authors discuss similarity or differences in findings and try to explain the differences?

- **Rationale:**
  - Systematic reviewers should be aware of prior reviews in their area, study and learn from their methods, and explicitly discuss comparative results, especially if there is a discrepancy between prior work and their own

DI7. Were directions for future research proposed?

- **Look for:**
  - One or more paragraphs in the discussion section where the authors make recommendations for future primary studies or systematic reviews

DI7. Were directions for future research proposed?

- **Rationale:**
  - Authors become very familiar with what is known and not known with respect to the area of the systematic review
  - They can and should be able to make recommendations

Ruivo

- DI1: Study limitations
  - 'Indirect evidence' being used
  - Small sample sizes in primary studies (the usual: problem is with the primary studies)
  - Omission of observational studies
  - Omission of studies of not-off-the-shelf AVGs
  - NOT: Single quality rater, extractor, conclusion drawer, publication bias
- DI2: publication and other biases assessed
  - Term bias does not occur
  - Concept of bias also not discussed under other labels

Ruivo

- DI3: totality of evidence considered? Alternative explanations considered?
  - 'Identified benefits and risks of AVGs' discussed, based on subjective interpretation of evidence table information
    - Physical activity and energy expenditure
    - Cardiorespiratory strain
    - Balance
    - Enjoyment, Social facilitation and Mood
    - Quality of life
  - Some alternative explanations considered as possible
- DI4: Generalization conclusions appropriate?
  - Author is clear that for transfer of findings from healthy subjects or individuals with other diagnoses to cardiovascular rehabilitation patients he is using 'indirect evidence'

Ruivo

- DI5: Results clinically meaningful?
  - Feasibility, efficacy and safety of AVGs in cardiac rehab discussed
- DI6: similarities/differences with earlier systematic reviews discussed?
  - Mentioned, no comparisons made
- DI7: directions for future research
  - Just a bland "before [the application of non-cardiac research to cardiac rehabilitation], future research should carefully assess the feasibility, efficacy, and safety of such strategy in light of the specificities of the cardiac population"
  - Yet no clinical direction 'don't do it' !!!

For the next session (March 5):

- Read manual sections/AQASR questions on
  - Studies of Interventions
  - Studies of Prevention
  - Studies of economic evaluation
- Read Polisena et al., with a focus on the sections/ information on intervention
- Read Davis et al., with a focus on the sections/ information on (1) prevention, and (2) economic evaluation

- Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab. 2009 Oct;11(10):913-30

- Davis JC, Robertson MC, Ashe MC, Liu-Ambrose T, Khan KM, Marra CA. Does a home-based strength and balance programme in people aged > or =80 years provide the best value for money to prevent falls? A systematic review of economic evaluations of falls prevention interventions. Br J Sports Med. 2010 Feb;44(2):80-9.

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