Assessing the Quality and Applicability of Systematic Reviews (AQASR)

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

Session 5 – March 5, 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:

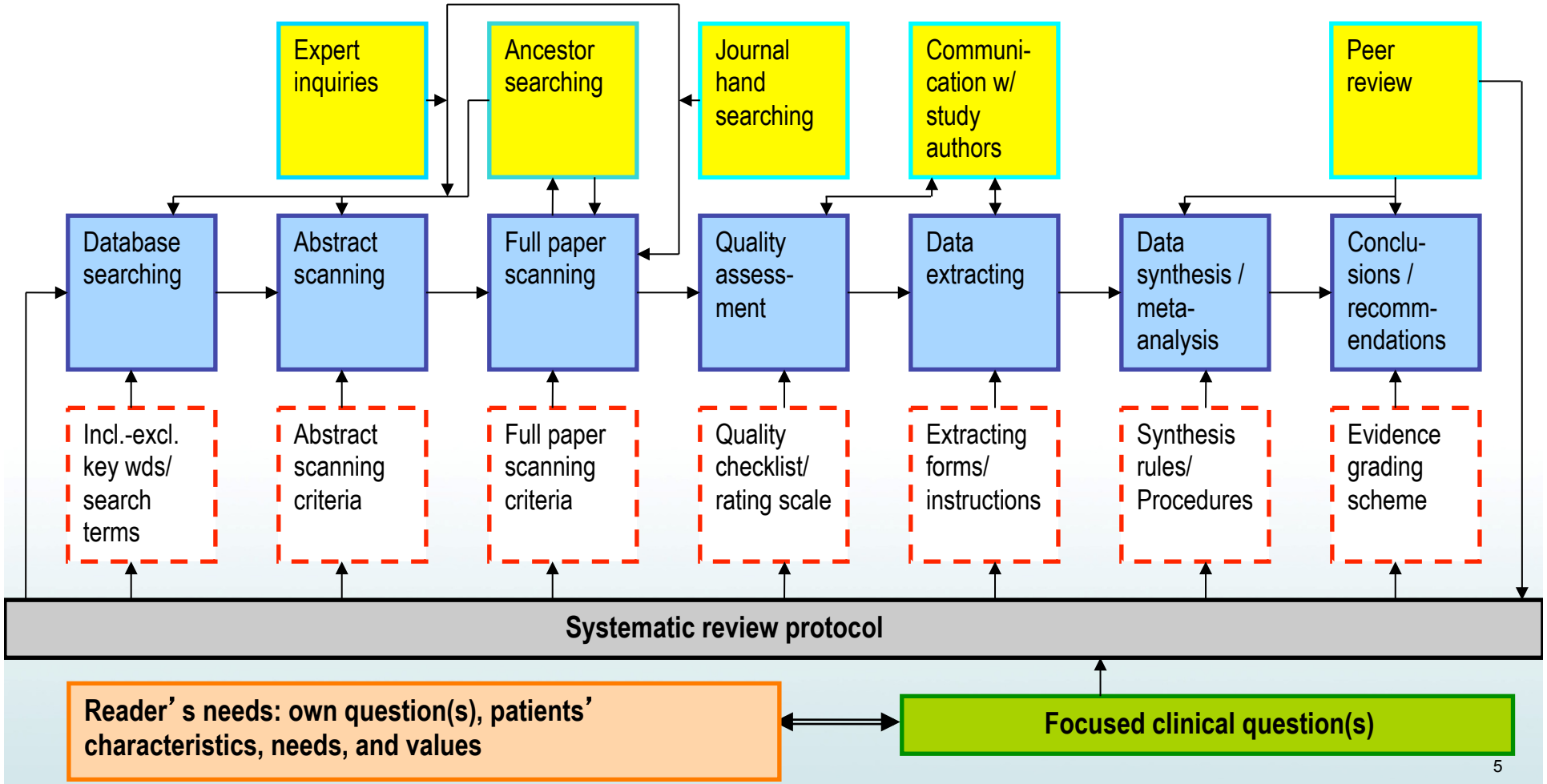
- Delineate steps and issues in the development of systematic reviews
- Introduce *Assessing the Quality and Applicability of Systematic Reviews (AQASR)* *(© SEDL/NCDDR 2011)*
- Describe how AQASR can be used in evaluating whether a particular systematic review can be trusted to provide an unbiased,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

INTERVENTION STUDIES
PREVENTION STUDIES

Intervention studies and prevention studies: same difference

	intervention study			prevention study	
	treated group	comparator group		treated group	comparator group
cured (+)	50%	40%	not diseased (+)	80%	70%
not cured (-)	50%	60%	diseased (-)	20%	30%
cases	1,000	1,000	cases	1,000	1,000

IN1. Are the interventions(s) and the comparator(s) of interest described/defined?

- **Look for:**
  - Description of the intervention, including a definition of the procedures to which the intervention will be compared (comparator)
  - Previous findings regarding effectiveness of certain types of interventions
  - Definitions of the intervention of interest, including those to be excluded
    - Information about the comparator(s), such as dose, frequency, intensity or duration

IN1. Are the interventions(s) and the comparator(s) of interest described/defined?

- **Rationale:**
  - Interventions must be specifically described so that practitioners and researchers can replicate or use them in practice or research
  - Must be presented in the context of other similar/ related interventions and standard practice
  - Most useful if they make explicit comparisons with outcomes of alternative interventions

IN2. Are the provider(s) of interest described/ defined?

- **Look for:**
  - Information on the type of people providing the intervention (physicians, nurses, educators, counselors, therapists, etc.)
  - Description of the settings and organizations in which the interventions were provided
  - If relevant, description of the training and skills needed to conduct the intervention

IN2. Are the provider(s) of interest described/ defined?

- **Rationale:**
  - Quality and feasibility of the intervention may depend on the training, skills and knowledge of the people providing the intervention
  - Characteristics of the provider organizations can affect the relevance for particular settings (e.g. primary, secondary or tertiary care)

IN3. Is treatment integrity (fidelity) of the primary studies evaluated? Was the occurrence of co-interventions noted?

- **Look for:**
  - Statement describing the methods used to evaluate treatment fidelity, when appropriate (e.g. pill counts, observer ratings, therapist checklist)
  - Descriptions of how primary studies were reviewed for occurrence of co-interventions

IN3. Is treatment integrity (fidelity) of the primary studies evaluated? Was the occurrence of co-interventions noted?

- **Rationale:**
  - Treatment fidelity refers to how well an intervention is delivered relative to a previously created study protocol; this is easier to assess in manualized protocols
  - The quality of the administration of an intervention is an important point to consider
  - Unequal co-treatments in intervention arms confounds analysis

IN4. For reviews that include RCTs: Was the integrity of randomization considered?

- **Look for:**
  - Statement describing methods used to determine whether case assignment to treatment conditions was random
  - Statement that randomization concealment in the primary studies was evaluated

IN4. For reviews that include RCTs: Was the integrity of randomization considered?

- **Rationale:**
  - The issue of effective randomization is central to assessing the quality of treatment studies using controls

IN5. Were the primary studies' method of analysis (intent-to-treat vs per-protocol) considered?

- **Look for:**
  - Statement describing consideration of intent-to-treat analyses by the primary studies
  - Statement on loss to follow-up in primary studies
  - Attention to comparison between lost and non-lost cases in primary studies

IN5. Were the primary studies' method of analysis (intent-to-treat vs per-protocol) considered?

- **Rationale:**
  - Intent-to-treat (ITT) analyses are designed to avoid misleading conclusions based on study artifacts such as higher attrition for people with more severe illnesses
  - Per-protocol analysis includes only those cases that complete the entire protocol
  - An evaluation of how the intervention works in the 'real world' should be based on ITT analysis, or minimal attrition and non-selective attrition

IN6. Was the potential of confounding in the studies included assessed (e. g. comparability of intervention vs control group)

- **Look for:**
  - Comparison of cases assigned to various treatment arms on demographic and baseline characteristics
  - Selective attrition

IN6. Was the potential of confounding in the studies included assessed (eg. comparability of intervention vs control group)

- **Rationale:**
  - Non-random assignment creates confounds and severely diminishes the value of a study
  - A comparison of groups on demographic and baseline characteristics helps assure that randomization was effective or that groups in non-randomized studies were comparable

IN7. Was blinding of patients, clinicians, outcome assessors, and analysts assessed?

- **Look for:**
  - Statements that use of blinding in the primary studies was determined
  - Was this information used in assessing their methodological quality

IN7. Was blinding of patients, clinicians, outcome assessors, and analysts assessed?

- **Rationale:**
  - Blinding is a countermeasure to "expectation effects" with respect to outcomes
  - Systematic reviews should take into account whether this occurred when weighing the evidence
  - Blinding of outcome assessors and analysts is always possible and should always be considered

IN8. Was loss to follow-up assessed?

- **Look for:**
  - Statements that drop-out percentages in treatment and control groups were recorded or calculated
  - Statement that of studies were quality scored based on a cut-off level of acceptable loss to follow-up – e.g. 15%
  - Consideration of selectivity of attrition

IN8. Was loss to follow-up assessed?

- **Rationale:**
  - Selective attrition can bias the results of an RCT or other intervention study, even if randomization was conducted correctly and the groups were balanced at baseline
  - An arbitrary standard that attrition should be below 15% is often used to distinguish high from low quality studies

IN9. Were sources of heterogeneity addressed; was the sensitivity of findings to addition/ omission of key studies considered?

- **Look for:**
  - A sensitivity analysis which test the effect of exclusion of studies where there is ambiguity as to whether the inclusion criteria are met

IN9. Were sources of heterogeneity addressed; was the sensitivity of findings to addition/ omission of key studies considered?

- **Rationale:**
  - Systematic reviews can be conducted using a variety of approaches which can change the results of a review
  - This includes the decision as to whether a study should be included or not
  - Clear justification for inclusion or exclusion should be included in the review
  - A sensitivity analysis may show how much difference inclusion/ exclusion make for final conclusions

IN10. Were the major outcomes (benefits and harms) considered?

- **Look for:**
  - Description of negative (adverse events) and positive results in the studies
  - Recommendations that take into account both types of outcomes (positive and negative)

IN10. Were the major outcomes (benefits and harms) considered?

- **Rationale:**
  - The ultimate purpose of a systematic review is to provide an evidence base for answering a focused question
  - Interventions and/or treatments involve expected benefits, as well as costs and risks of harm
  - This balance must be considered when making a judgment on the evidence

IN11. Was the generalizability of the data addressed?

- **Look for:**
  - A statement(s) considering the generalizability of the results with respect to the subject population, other populations, different interventions, and different measures of the key outcome(s)

IN11. Was the generalizability of the data addressed?

- **Rationale:**
  - Each reader of the review must be able to determine if the treatment recommendations are applicable to his/her own patient population

IN12. Were the studies cited as support sufficiently strong in quality and quality?

- **Look for:**
  - An explicit approach to specifying the levels of evidence used to support the treatment recommendations
  - A methodology (e.g. AAN) that ties strength of recommendation to quality, quantity and consistency of the evidence

IN12. Were the studies cited as support sufficiently strong in quality and quality?

- **Rationale:**
  - Treatment recommendations should not exceed the quality of the evidence that is reviewed
  - Use of an explicit system to grade the evidence maximizes the possibility that the recommendations are appropriately based upon the strongest possible evidence

IN13. Were the costs of treatment options considered?

- **Look for:**
  - Information on costs in the tables derived from the primary studies
  - A statement considering the costs of the treatment(s) considered, based on other sources

IN13. Were the costs of treatment options considered?

- **Rationale:**
  - While a particular treatment may be well justified based upon the evidence, and have no or negligible risk, it may be too costly for a particular patient or group of patients to be utilized
  - Systematic reviews that consider cost issues explicitly are of more value to readers

Questions?

Polisena et al.

- IN1: Interventions/comparators defined/described?
  - HTM and ST defined; usual care not defined
  - Descriptions of HTM, ST and usual care (if in primary study) in evidence table
- IN2: providers of interest defined/described?
  - Not in introduction
  - Some statement of staff involved in evidence table
- Treatment integrity evaluated? Co-interventions noted?
  - Study quality ratings (ref #47) include ‘compliance failures’
  - Nothing else noted/ reported

Polisena et al.

- IN4: integrity randomization
  - Study quality ratings (ref #47) include ‘methods of randomization’; no further comments
- IN5: intent-to-treat vs per-protocol
  - Study quality ratings (ref #47) include ‘statistical methods used’; no further comments
- IN6: potential for confounding
  - Study quality ratings (ref #47) include ‘statistical methods used’; ‘equivalence of intervention and control groups’, and ‘missing results’; no further comments

Polisena et al.

- IN7: blinding of participants/treaters/assessors/analysts
  - Not evaluated; no comments
- IN8: loss to follow-up
  - Study quality ratings (ref #47) include ‘drop-outs’ and ‘missing results’; imputing of some missing results described; no further comments
- IN9: sources of heterogeneity assessed; sensitivity of findings assessed
  - Subgroup analysis to explore heterogeneity promised but not reported
  - No sensitivity analysis done

Polisena et al.

- IN10: benefits and harms considered
  - Note that ‘Patient adverse events were not discussed in any of the selected studies.’
  - Conclusion: ‘In general, home telehealth had a positive impact on the use of numerous health services and glycaemic control.’
- IN11: generalizability of data addressed
  - Statistical and clinical heterogeneity of studies noted
  - Noted that most studies excluded patients with cogn. Impairment, mental illness, language barrier, etc.
  - No comments on generalizability due to treaters/ health systems/HTM and TS methods/other

Polisena et al.

- IN12: primary studies sufficiently strong in quality/ quantity
  - Studies quality screened, only RCTs and observational studies with score C/higher included
  - Weakness and small number of studies (for some outcomes) noted
- IN13: costs of treatment options considered
  - No mention

Questions?

EC1. Does the systematic review specify what the economic question addressed is – cost, cost-effectiveness, cost-benefit, cost-utility – and maintain this focus throughout?

- **Look for:**
  - identification of the specific question(s) in the introduction
  - consistency of the literature collected with this question
  - evidence tables that provide information relevant to the question
  - conclusions or recommendations that do not stray from the narrow area of interest

Four types of cost studies

- Costs study: what is the expense connected to this intervention/ prevention/screening/etc. ($ per case treated, e.g.)
- Cost-utility analysis (CUA): what is the cost relative to outcomes (falls, QALYs, etc.) ($ per fall prevented, e.g.)
- Cost-benefit analysis (CBA) (also called benefit-cost analysis - BCA): what is the cost relative to the monetary value of the benefits ($ saved/gained per $ invested, e.g.)
- Cost-effectiveness analysis: what is the incremental cost effectiveness ratio (ICER) of one intervention/prevention/ screening compared to usual care/nothing/etc., relative to outcome(s)

  [$new - $comparator] / [outcomes new – outcomes comparator]

EC1. Does the systematic review specify what the economic question addressed is – cost, cost-effectiveness, cost-benefit, cost-utility – and maintain this focus throughout?

- **Rationale**
  - The costs and outcomes that are related to one another differ widely in these four types of economic studies, and the authors should be clear about which type of primary studies they are interested in locating, evaluating, selecting and synthesizing.

EC2. Does the systematic review specify which perspective – patient, insurer, society, etc. – and which time horizon, are of interest in answering the economic question, and does it maintain that focus throughout?

- **Look for:**
  - identification of the specific perspective(s) and time horizon in the introduction
  - consistency of the literature collected with this perspective and horizon
  - evidence tables that provide information relevant to the perspective and horizon
  - conclusions or recommendations that do not stray from the perspective and horizon taken

EC2. Does the systematic review specify which perspective – patient, insurer, society, etc. – and which time horizon, are of interest in answering the economic question, and does it maintain that focus throughout?

- **Rationale**
  - What is a cost and what a benefit depends very much on the entity whose perspective is taken.
  - Most experts recommend the society perspective, because it results in the most complete enumeration of costs and benefits
  - Other perspectives are legitimate
  - Focus on short-term and/or long-term benefits needs to be made explicit

EC3. Have the various studies considered been evaluated for their methodological quality by means of a checklist or rating scale specific to economic evaluations?

- **Look for:**
  - mention of the CHEC (Consensus on Health Economic Criteria), the PQAQ (Pediatric Quality Appraisal Questionnaire) or another instrument
  - specification of a list of key questions, apart from or in addition to the CHEC, PQAQ or other instrument, which is used to evaluate the primary studies with respect to their evidence

EC3. Have the various studies considered been evaluated for their methodological quality by means of a checklist or rating scale specific to economic evaluations?

- **Rationale**
  - A large number of instruments have been proposed, by individual investigators or by official or self-appointed panels, to evaluate the methodological quality of economic studies
  - Because key to the quality of the evidence produced by such studies are a number of factors that do not play in systematic reviews of interventions or diagnostic tests, a special checklist or instrument needs to be used.

EC4. Have all important and relevant costs been identified for all alternative interventions or other programs being evaluated or compared?

- **Look for:**
  - a listing of all costs the systematic reviewer considers relevant
  - use of a checklist to review inclusion of all those costs in the primary studies
  - estimation of omitted costs from other studies

EC4. Have all important and relevant costs been identified for all alternative interventions or other programs being evaluated or compared?

- **Rationale**
  - Most health care interventions have a number of direct and indirect costs, the nature of which depend on a variety of factors, primarily the organization of the heath care system in which they are embedded.
  - A systematic review needs to ensure that all studies considered include the same cost categories, or adjust the findings of studies that omit certain costs (imputation).

EC5. Have the entries in the evidence table been adjusted, to the degree possible and in a proper fashion, for those factors that make the results of various primary studies incomparable?

- **Look for adjustments for:**
  - Currency exchange rates, if studies from multiple economies are considered
  - Inflation adjustments, using the consumer price index (CPI), the medical consumer price index (MCPI), or another suitable index
  - The discount rate used by the primary study
  - Cost categories that the authors of the primary study omitted
  - Sensitivity analyses to assess the impact of assumptions underlying the adjustments made

EC5. Have the entries in the evidence table been adjusted, to the degree possible and in a proper fashion, for those factors that make the results of various primary studies incomparable?

- **Rationale**
  - Primary studies from various countries and time periods can be made compatible, to a degree, by making adjustments to the various costs and (sometimes) outcomes reported.
  - Minor changes in the values used may have big impacts, especially if data from widely different years are used
  - Consequently, a sensitivity analysis should be provided, for all these types of adjustments.

EC6. For studies that compare cost-effectiveness of interventions for disparate health problems: have the outcomes all been expressed in a proper and comparable common metric?

- **Look for:**
  - use of quality-adjusted life years (QALYs), disability-adjusted life years (DALYs) or similar "universal metrics", with or without adjustment for diminished-quality years of life
  - information on thousands of dollars per QALY/DALY produced or QALY/DALY loss prevented
  - a justification of the appropriateness of this metric and of comparability of outcome data across studies
  - a sensitivity analysis for any adjustments to the results of studies that used disparate outcome measures

EC6. For studies that compare cost-effectiveness of interventions for disparate health problems: have the outcomes all been expressed in a proper and comparable common metric?

- **Rationale**
  - Studies of the value of investments in treating different disorders with varied outcomes need to use a common metric.
  - QALYs and DALYs are often used to provide a common denominator.
  - Even if all available studies used the same metric, systematic reviewers should be careful to assess whether these truly were collected and interpreted similarly in all primary studies.

EC7. Does the systematic review acknowledge differences between primary studies that cannot be adjusted for, because of lack of information?

- **Look for:**
  - statements on incomparability of either the costs or the outcomes of economic studies, as footnotes to evidence tables or in the text.

EC7. Does the systematic review acknowledge differences between primary studies that cannot be adjusted for, because of lack of information?

- **Rationale**
  - Because of
    - differences between health care systems in which programs operate
    - differences in cost assumptions or outcomes that cannot be adjusted for
  - often claims of comparability of costs and/or outcomes should not be made
  - and careful systematic reviewers will not make them.

Questions?

Davis et al.

- EC1: specification of nature of economic question
  - 'value for money of fall prevention strategies' (abstract) (suggests CBAs are of interest)
  - 'home-based strength/balance program best value for money'? (title) (suggests ICERs are of interest)
  - cost-effectiveness, cost-utility and cost-benefit studies identified
- EC2: specification of perspective and time horizon
  - Perspective and time horizon of primary studies noted in evidence table

Davis et al.

- EC3: evaluation of quality of primary studies
  - Drummond et al. checklist
  - Quality of Health Economic Studies (modified)
- EC4: all important / relevant costs identified
  - Cost items measured in primary studies noted, provided in evidence table (table 3)
- EC5: adjustments for time / currency / discount rate
  - Time adjustments using CPI specific to country
  - Currency adjustments using currency exchange rates
  - Discount rate noted in evidence table, not adjusted for

Davis et al.

- EC6: (IF cost-effectiveness study of interventions for disparate health problems): proper / comparable common metric
  - NA: falls prevented is common metric
- EC7: acknowledgment of (in)comparability of studies which cannot be brought on equal footing
  - Acknowledges difficulty combining/comparing due to differences in:
    - Health system; cost items included
    - Perspectives taken
    - Time frame for collection/measurement of costs
  - No meta-analysis performed
  - Yet makes ‘best value for money’ recommendation
  - Justified by enormous differences in $/fall prevented?

Questions?

For the next session (March 19):

- Read manual sections/AQASR questions on
  - Diagnostic studies
  - Prognostic studies

- Read:
  - Pape, T. L., High, W. M., Jr., St. Andre, J., et al. (2013). Diagnostic accuracy studies in mild traumatic brain injury: A systematic review and descriptive analysis of published evidence. *PM&R*, *5*(10), 856-881, with a focus on the sections/information on evaluation/synthesis of information on diagnosis. http://www.pmrjournal.org/article/S1934-1482%2813%2900348-1/abstract
  - Clay, F. J., Newstead, S. V., & McClure, R.J. (2010). A systematic review of early prognostic factors for return to work following acute orthopaedic trauma. *Injury*, *41*(8), 787-803, with a focus on the sections/information on evaluation/synthesis of prognostic factors. http://www.injuryjournal.com/article/S0020-1383%2810%2900257-3/abstract

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