Assessing the Quality and Applicability of Systematic Reviews (AQASR)

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

Session 6 – March 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:

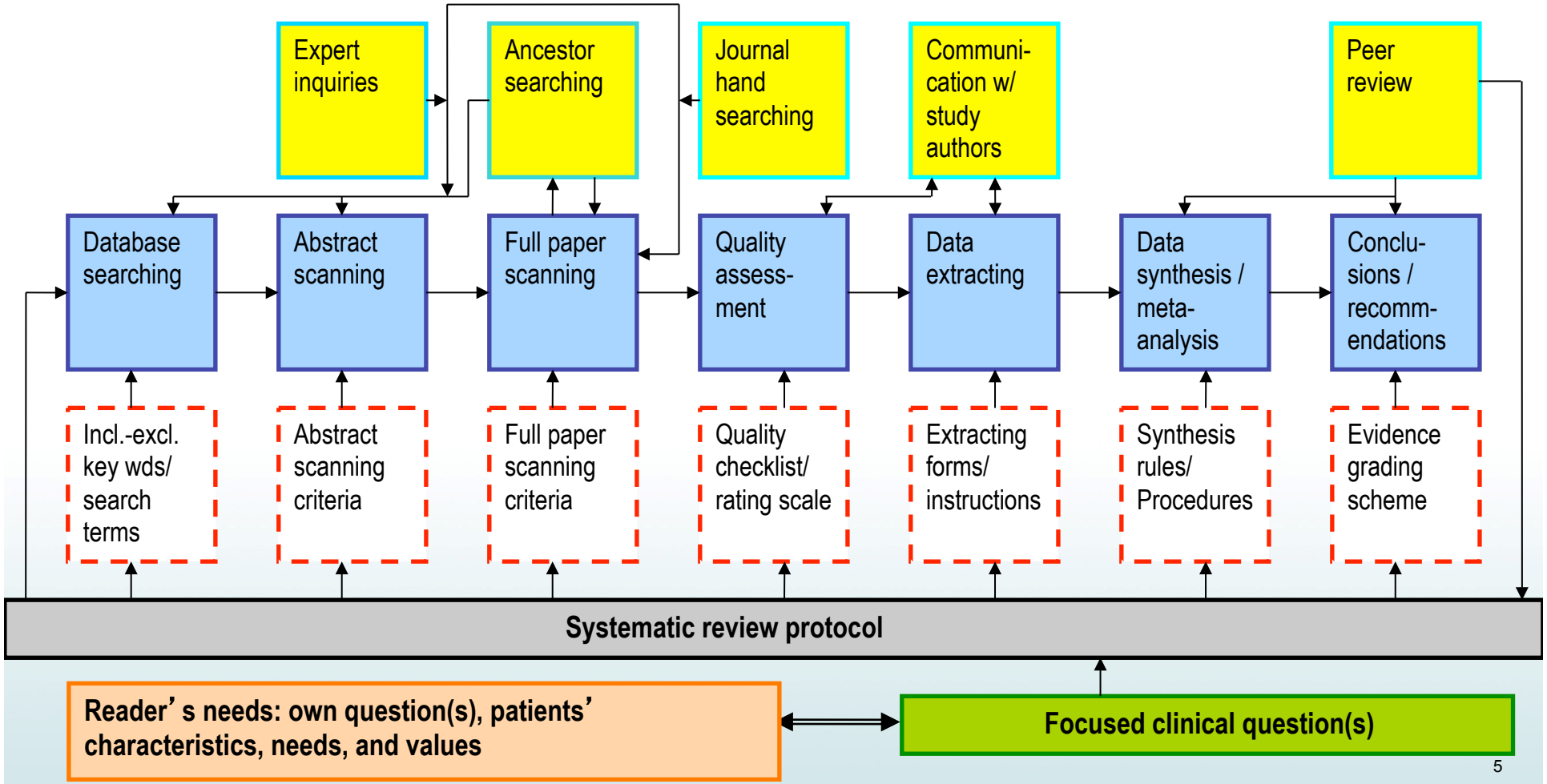
- Delineate steps and issues in the development of systematic reviews
- Introduce *Assessing the Quality and Applicability of Systematic Reviews (AQASR)* *(© SEDL/NCDDR 2011)*
- Describe how AQASR can be used in evaluating whether a particular systematic review can be trusted to provide an unbiased,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

SYSTEMATIC REVIEWS OF DIAGNOSTIC ACCURACY STUDIES

	reference test		
index test	**+**	**-**	**total**
+	a TP true positive	b FP false positive	a+b
-	c FN false negative	d TN true negative	c+d
total	a+c	b+d	a+b+c+d

sensitivity: a/(a+c)
specificity: d/(b+d)
accuracy: (a+d)/(a+b+c+d)

What do we call a mild Traumatic Brain Injury (mTBI)?

- TBI=incident, injury to brain happening at a specified place and time, with a specified etiology (fall, MVA, blast, etc.)
- TBI=potentially continuing injured state, with manifestations (symptoms, 'broken' neural connections, specific 'marker' cells in the blood, etc.) that often disappear with passage of hours/ days/ weeks, but not in all cases

What do we call a mild Traumatic Brain Injury (mTBI)?

- Eight case definitions of/clinical criteria for mild TBI
- Describing what should have been observed/ observable at the time of the TBI:
  - Loss of consciousness (LOC)
  - Alteration of consciousness (AOC)
  - Post-traumatic amnesia (PTA)
- Differing in:
  - Minimum/maximum duration of LOC, AOC, PTA
  - Glasgow Coma Score maximum, at a specified time after the injury

Glasgow Coma Scale

Score	Eyes	Verbal	Motor
1	Does not open eyes	Makes no sounds	Makes no movements
2	Opens in response to painful stimuli	Incomprehensible sounds	Extension to painful stimuli
3	Opens in response to voice	Utters inappropriate words	Abnormal flexion to painful stimuli
4	Opens spontaneously	Confused, disoriented	Flexion / Withdrawal to painful stimuli
5	N/A	Oriented, converses normally	Localizes painful stimuli
6	N/A	N/A	Obeys commands

Pape et al.'s question:

- How well do proposed index tests of TBI effects (symptoms, functioning capacity) correspond with reference tests?
- Index tests specified, different for each study
- Reference test never specified (in systematic review paper as printed) presumably are based on physical exam and/or interview and/or more complete neuropsychological examination

DS1. Did the systematic reviewers select studies that were the same with respect to patient factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Look for:**
  - Mention that studies were selected based on patient subgroups, spectrum of disease, co-morbidities, clinical setting (especially primary vs. secondary vs. tertiary care),
  - A subgroup analysis or sensitivity analysis that explores the role of these factors

DS1. Did the systematic reviewers select studies that were the same with respect to patient factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Rationale**
  - Sensitivity and specificity, as well as other measures used to evaluate test accuracy and utility, are not fixed properties of a test, but very much dependent on the sample of patients they are used with.
  - "Averaging" over the results of heterogeneous samples may not be warranted.

DS1. Did the systematic reviewers select studies that were the same with respect to patient factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Pape et al.:**
  - No selection
  - No statistical control
  - Sensitivity/specificity discussed in relation to:
    - Time since injury
    - Cut-off point selected

DS2. Did the systematic reviewers select studies that were the same with respect clinician factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Look for:**
  - Mention that studies were selected based on the training and expertise of any test administrators/ readers-interpreters (e.g. radiologists), if applicable
  - Indications of the availability to the test administrators/ readers of any supplemental information on the patients that is/is not available in routine clinical practice or that differed from one primary study to the next
  - A subgroup analysis or sensitivity analysis that explores the role of these factors

DS2. Did the systematic reviewers select studies that were the same with respect clinician factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Rationale**
  - Sensitivity and specificity, as well as other measures used to evaluate test accuracy, are not fixed properties of a test, but very much dependent (for tests that require interpretation by a human) on the training and experience of the test readers.
  - "Averaging" over the results of heterogeneous samples may be unwarranted.

DS2. Did the systematic reviewers select studies that were the same with respect clinician factors impacting test sensitivity and specificity, and/or did they control for these factors statistically?

- **Pape et al.:**
  - No selection
  - No statistical control
  - Issue not discussed

DS3. Does the systematic review include discussion/ specification/ tabulation of other factors that may impact diagnostic accuracy parameters?

- **Look for (1):**
  - Specification of the cut-off point selected (on the index test and the reference standard) to differentiate between “positive” and “negative”
  - Information on the time elapsed between the index test and the reference test in each primary study
  - Discussion of the frequency and disposal of uninterpretable/ intermediate results for index test and the reference test

DS3. Does the systematic review include discussion/ specification/ tabulation of other factors that may impact diagnostic accuracy parameters?

- **Look for (2):**
  - Selection criteria for primary studies that include any of these characteristics
  - A column in an evidence table specifying this information for individual studies
  - Subgroup analysis and/or sensitivity analysis that explores the impact of these factors on estimated pooled values for sensitivity, specificity or other accuracy indicators.

DS3. Does the systematic review include discussion/ specification/ tabulation of other factors that may impact diagnostic accuracy parameters?

- **Rationale:**
  - Because they utilize "simple" dichotomies, the results of diagnostic studies are very sensitive to minor differences in protocols for obtaining and processing the results of the index test and reference standard.
  - Consequently, systematic reviewers need to be very careful comparing like with like, and/or using statistical means to eliminate the confounding effects of differences between studies.

DS3. Does the systematic review include discussion/ specification/ tabulation of other factors that may impact diagnostic accuracy parameters?

- **Pape et al.:**
  - Cut-off points discussed (mean for ‘normals’ - 1 SD, for instance)
  - Time between reference test and index test not discussed
  - Uninterpretable test results not discussed
  - No further relevant information provided
  - ‘Sensitivity analysis’ as to the effect of including more measures in the index test

DS4. Was the methodological quality of the studies considered for (and included in) the systematic review evaluated using an appropriate instrument such as the QUADAS? IF SO, was calculation and use of a total score avoided?

- **Look for:**
  - Mention of a diagnostic study-specific methodological quality assessment measure
  - Specification of individual key methodological characteristics (for instance, blinding of index test reader to reference test and vice versa)
  - Use of findings of such assessments in qualitative analysis or meta-analysis

DS4. Was the methodological quality of the studies considered for (and included in) the systematic review evaluated using an appropriate instrument such as the QUADAS? IF SO, was calculation and use of a total score avoided?

- **Rationale**
  - Following a proper methodology is a requirement for diagnostic studies as for all research. The QUADAS was developed to help systematic reviewers assess study quality.
  - However, use of a total score in the analysis is not recommended, as some shortcomings may increase a study's sensitivity, and others decrease it. A more fine-grained use of quality assessment results is recommended.

DS4. Was the methodological quality of the studies considered for (and included in) the systematic review evaluated using an appropriate instrument such as the QUADAS? IF SO, was calculation and use of a total score avoided?

- **Pape et al.:**
  - Standards for Reporting of Diagnostic Accuracy (STARD) used (in modified form)
  - Consensus of 7 raters used
  - No total score calculated
  - STARD information NOT used to select/weigh studies

DS5. Did the systematic review identify how the primary studies recruited subjects? Did it determine whether subjects in the primary studies were a consecutive series, or whether additional criteria were used to select them?

- **Look for:**
  - General criteria for the types of studies selected
  - Comments on individual studies that in patient recruitment deviated from the ideal
  - Statistical manipulation that takes these limitations into account

DS5. Did the systematic review identify how the primary studies recruited subjects? Did it determine whether subjects in the primary studies were a consecutive series, or whether additional criteria were used to select them?

- **Rationale**
  - While ideally a series of consecutive patients typical of those with whom the index test will be used is recruited to study test accuracy, logistical, financial or ethical problems sometimes make doing so difficult.
  - However, subject selection on another basis seriously affects whether sensitivity and specificity calculation makes sense, or the size of these parameters, if calculated.

DS5. Did the systematic review identify how the primary studies recruited subjects? Did it determine whether subjects in the primary studies were a consecutive series, or whether additional criteria were used to select them?

- **Pape et al.:**
  - Problem of selection of (non)consecutive cases noted in general
  - No mention for individual studies in the tables (except STARD yes/no)

DS6. Does the systematic review provide a description of the nature of the index test and the reference standard and of the reproducibility (test-retest reliability) of these tests?

- **Look for:**
  - Careful descriptions of the index test and the reference standard, including any study-to-study differences
  - Tabulations of test-retest reliability of the index and reference test, alongside listing of the sensitivity, positive predictive value, etc. parameters derived for the index test from the comparison of the results of the two
  - Values of the reproducibility of index test and reference standard from other sources
  - Discussion of the importance of reproducibility to estimates of diagnostic accuracy

DS6. Does the systematic review provide a description of the nature of the index test and the reference standard and of the reproducibility (test-retest reliability) of these tests?

- **Rationale:**
  - If both index test and reference test are not well reproducible, a high sensitivity and specificity cannot be expected. Information on the test-retest correlation of the two tests may be derived from the studies included in the review, or from yet other sources.

DS6. Does the systematic review provide a description of the nature of the index test and the reference standard and of the reproducibility (test-retest reliability) of these tests?

- **Pape et al.:**
  - Index tests described
  - No information on their test-retest reliability
  - Reference test(s) not described
  - No information on its/their test-retest reliability

DS7. Did the systematic review avoid estimating a pooled value separately for sensitivity and specificity?

- **Look for:**
  - "averaging" of sensitivity and specificity separately (without indications that the authors are aware of their being linked phenomena)
  - use of side-by-side forest plots for the two
  - use of summary ROC curves

DS7. Did the systematic review avoid estimating a pooled value separately for sensitivity and specificity?

- **Rationale:**
  - Sensitivity and specificity are by definition negatively correlated, in that one can always improve sensitivity (by setting a higher cutoff score for "diseased"), but at a cost of worsened specificity.
  - An appropriate pooling the reported values from individual studies uses a summary receiver operating characteristic (ROC) curve.

DS7. Did the systematic review avoid estimating a pooled value separately for sensitivity and specificity?

- **Pape et al.:**
  - No pooling of sensitivity, specificity, ROC or accuracy reported

DS8. Are the findings with respect to the index test discussed in the context of its use in clinical practice, including costs, possible treatment strategies for the disease, harms, alternative tests, use in a sequence of tests (screening, add-on, etc.), treatment decisions?

- **Look for:**
  - A discussion that goes well beyond a restatement of specificity and sensitivity and other diagnostic accuracy parameters
  - References to other (systematic) reviews of the index test, the reference standard and alternatives that discuss the wider context

DS8. Are the findings with respect to the index test discussed in the context of its use in clinical practice, including costs, possible treatment strategies for the disease, harms, alternative tests, use in a sequence of tests (screening, add-on, etc.), treatment decisions?

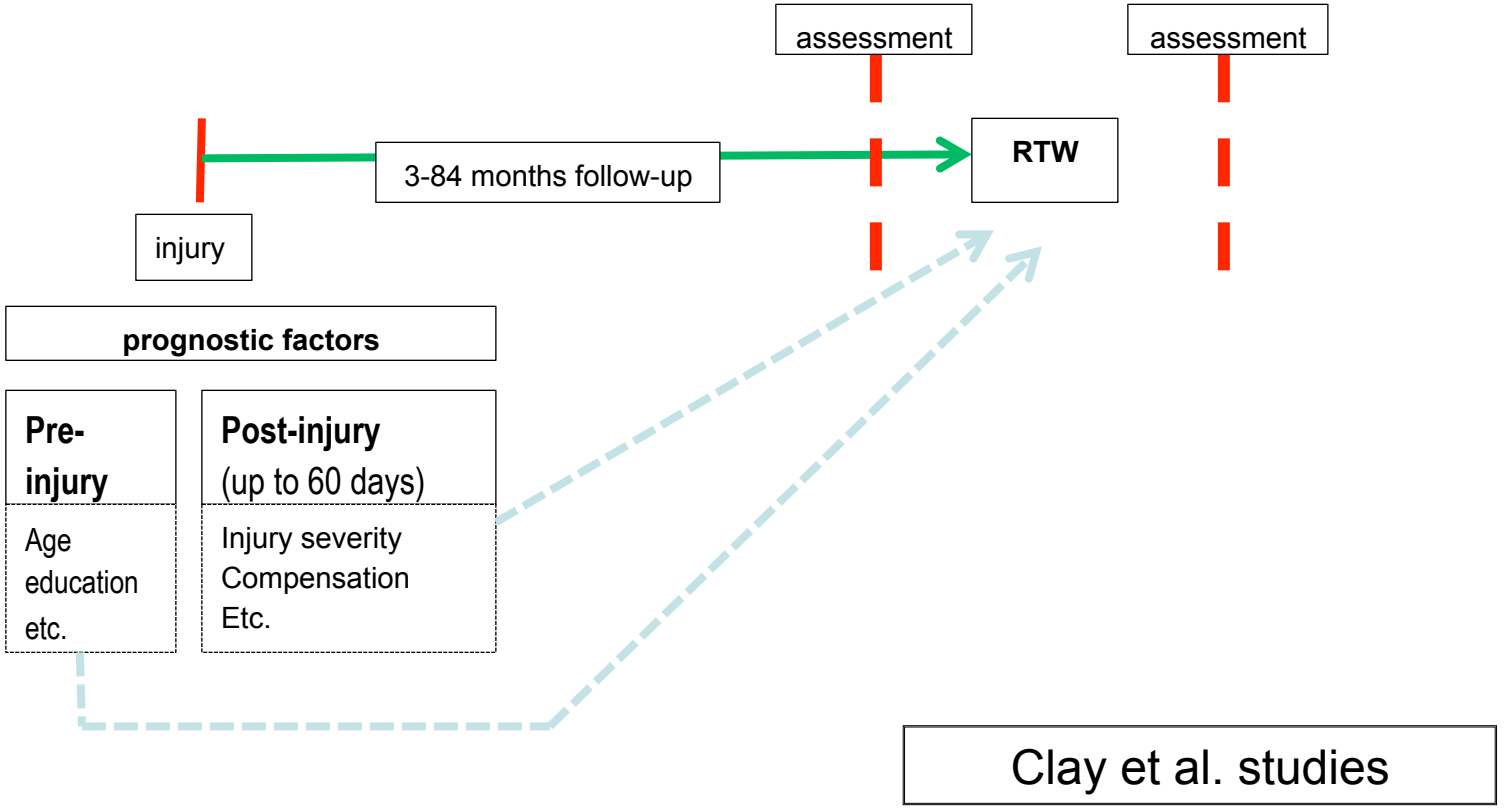
- **Rationale**
  - Because often high-risk and high-cost decisions on further testing or on treatment are based on test results, a quality systematic review will put its findings with respect to the index test in a wider perspective, to assist clinicians in making use of the test within a careful assessment-screening-testing-treating protocol.

DS8. Are the findings with respect to the index test discussed in the context of its use in clinical practice, including costs, possible treatment strategies for the disease, harms, alternative tests, use in a sequence of tests (screening, add-on, etc.), treatment decisions?

- **Pape et al.:**
  - No mTBI test recommended
  - No discussion of uses of any/all tests in clinical practice
  - Focus is on development of a new test to be used in the authors' research

QUESTIONS?

QUESTIONS ONLY RELEVANT TO SYSTEMATIC REVIEWS OF PROGNOSTIC STUDIES


assessment
assessment
3-84 months follow-up
RTW
injury
prognostic factors
Pre-injury
Age
education
etc.
Post-injury
(up to 60 days)
Injury severity
Compensation
Etc.
Clay et al. studies

The problem with prognostic studies

Prognostic factor correlations						Beta for:		
	A	**B**	**C**	**study**	A	B	C
A	-	0.6	0.4	**1**	.18	.13	.07
B		-	0.3	**2**	.22	.14	-
C			-	**3**	.25	-	.08
				4	-	.15	.11
				5	.31	-	-
				6	-	.22	-
Outcome variable: RTW				**7**	-	-	.15

PS1. Do the authors define the population of interest and do they specify criteria to make sure that all the primary studies involved dealt with (a sample from) the same population?

- **Look for:**
  - A specific definition of the population of interest.
  - A set of criteria that help determine whether the samples studied satisfy the definition.
  - A checklist or other mechanism for assessing whether the samples being followed over time are representative of the original samples.

PS1. Do the authors define the population of interest and do they specify criteria to make sure that all the primary studies involved dealt with (a sample from) the same population?

- **Rationale:**
  - Readers (clinicians, researchers, others) have to be able to determine whether the findings will generalize to relevant patient groups.
  - Provision of clear criteria (inclusion/exclusion) allows the reader to determine whether study samples from the various papers reviewed were representative of the population of interest.

PS1. Do the authors define the population of interest and do they specify criteria to make sure that all the primary studies involved dealt with (a sample from) the same population?

- **Clay et al.:**
  - Yes: employed persons incurring an orthopedic injury
  - Further inclusion/exclusion criteria

PS2. Do the authors assess loss to follow-up (from first assessment of study subjects to last evaluation of the outcome of interest) in the primary studies and do they assess whether loss to follow-up was selective?

- **Look for:**
  - A column in the evidence table with rates of loss to follow-up.
  - Statement of a maximal acceptable rate of loss to follow-up.
  - Comparison of subjects followed and subjects lost
  - Whether or not primary studies were eliminated based on bias due to loss to follow-up (selective, excessive)

PS2. Do the authors assess loss to follow-up (from first assessment of study subjects to last evaluation of the outcome of interest) in the primary studies and do they assess whether loss to follow-up was selective?

- **Rationale:**
  - Selective or excessive attrition can result in a biased sample limiting generalizability of results.
  - Factors affecting attrition may also effect the outcome.

PS2. Do the authors assess loss to follow-up (from first assessment of study subjects to last evaluation of the outcome of interest) in the primary studies and do they assess whether loss to follow-up was selective?

- **Clay et al.:**
  - Follow-up percentage listed in evidence table, and 80% used as a study quality indicator
  - Studies with <80% follow-up not eliminated
  - Selectivity of attrition not noted, let alone analyzed

PS3. Do the authors specify criteria for the measurement of prognostic factors in the primary studies?

- **Look for:**
  - Definitions of prognostic factors used in primary studies.
  - A listing of measures/tests accepted as reliable and valid for these factors.
  - Operational definitions primary researchers may have used such as cut-points, specification of treatments, etc.
  - Studies excluded from review due to inadequate specification or measurement of prognostic factors.

PS3. Do the authors specify criteria for the measurement of prognostic factors in the primary studies?

- **Rationale:**
  - Different investigators may study similar or overlapping prognostic indicators or operationalize the same indicators in different ways.
  - The reader must be able to determine the comparability of prognostic factors and measures across studies.

PS3. Do the authors specify criteria for the measurement of prognostic factors in the primary studies?

- **Clay et al.:**
  - Focus is on different ways RTW has been operationalized (Y/N at a time point; time elapsed until; work time lost)
  - No comment on operationalization of prognostic factors
  - No information on (un)reliability of measurement of prognostic factors
  - No studies eliminated based on prognostic factor measurement issues

PS4. If the outcome is subjective, do the authors report on blinding of the outcomes assessors to prognostic factors?

- **Look for:**
  - The nature of the outcome. Is it objective (e.g., survival, a result on a laboratory test) or subjective (e.g., rated level of community independence, rated competence on job tasks).
  - A description of how outcome assessors were blinded to (masked to) the prognostic factors.

PS4. If the outcome is subjective, do the authors report on blinding of the outcomes assessors to prognostic factors?

- **Rationale:**
  - Investigators are rarely disinterested parties to their own studies. Knowledge of prognostic factors and study hypotheses could result in biased assessment of subjective outcomes.

PS4. If the outcome is subjective, do the authors report on blinding of the outcomes assessors to prognostic factors?

- **Clay et al.:**
  - N/A: outcome is 'objective'

PS5. Do the authors pay attention to whether or how the primary studies measured and dealt with potential confounders?

- **Look for:**
  - A listing of possible important confounders in the area of research covered by the primary studies.
  - A checklist or other indication that the systematic reviewers scrutinized the primary studies for the presence and statistical control of these confounders.
  - Exclusion or other special treatment of studies that did not adequately deal with confounding.

PS5. Do the authors pay attention to whether or how the primary studies measured and dealt with potential confounders?

- **Rationale:**
  - Confounding occurs when variables of interest have joint associations or associations with other variables not assessed that also have associations with the outcome of interest.
  - Presence of confounding may make it impossible to determine the independent effect of a variable of interest on the outcome.

PS5. Do the authors pay attention to whether or how the primary studies measured and dealt with potential confounders?

- **Clay et al.:**
  - Evidence of adequate adjustment for common confounders is a study quality indicator
  - No exclusion or special treatment of studies with inadequate treatment of confounders
  - Comment on the fact that many primary studies do not state whether common confounders were included in multivariate models
  - No comment on confounding issues

PS6. Do the authors scrutinize the analyses in the primary studies, especially in those using multiple prognostic factors?

- **Look for:**
  - Attention to selective reporting of results.
  - Specification in an evidence table of the analytic methods used in the primary studies.
  - A judgment by the systematic reviewers that the primary studies used the appropriate statistical method in the correct way.

PS6. Do the authors scrutinize the analyses in the primary studies, especially in those using multiple prognostic factors?

- **Rationale:**
  - Studies reviewed are likely to have included a wide range of possible predictive factors in the predictive models.
  - The potential impact of these multiple predictive factors across the studies reviewed must be considered.
  - Reviewers may be tempted to focus just on prognostic factors of interest rather than considering the full predictive models developed in various studies.

PS6. Do the authors scrutinize the analyses in the primary studies, especially in those using multiple prognostic factors?

- **Clay et al.:**
  - Studies with only univariate analysis not included
  - Type of analysis in evidence table
  - Note that primary studies often fail to report non-significant predictive factors

Questions?

For the next session (April 2):

- Read manual sections/AQASR questions on
  - Studies of measurement instruments
- Read Dobson et al., with a focus on the sections/ information on measurement instruments

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