Bringing in the patient/client: The Grading of Recommendations

Assessment, Development and Evaluation (GRADE) process

Webinar Series - Part 3

Presented by Marcel Dijkers

July 2, 2014

Text version of PowerPoint™ presentation for webinar sponsored by SEDL’s KTDRR. More information about the webinar can be found here:

[https://www.ktdrr.org/training/webcasts/webcast10-13/12/index.html](http://www.ktdrr.org/training/webcasts/webcast10-13/12/index.html)

Title Slide 0:

Bringing in the patient/client: The Grading of Recommendations Assessment, Development and Evaluation (GRADE) process.

An online webinar sponsored by SEDL’s Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR). Funded by NIDRR, US Department of Education, PR# H133A120012. Copyright © 2014 by SEDL

Slide 1: 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

Slide 2: 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 (in red)
4. A discussion of future developments in methods of qualifying and synthesizing evidence that might benefit disability/rehabilitation practice

Slide 3: Questions?

Slide 4: June 18 session topics:

* Oxford CEBM hierarchies, covering about all questions an EBP practitioner might ask
* AAN hierarchies, also covering various questions, with stronger focus on creating (rather than using) a systematic review

Slide 5: Today’s topic: GRADE: more sophistication, more complication

* “Grades of Recommendation, Assessment, Development and Evaluation”
* Developed since about 2000 by an international group of EBP (mostly EBM) specialists
* Published in *Allergy* (3 papers, 2009-11: introductory), *BMJ* (6 papers, 2008: focused on user perspective), and an ongoing series in the *J Clin Epidemiol* (2011-; to date, 15 out of a planned 20 papers: directed to producers) (see references on SEDL CKTDRR website)
* Ideas being widely copied (with/without modifications) by other guideline developers (AAN, Oxford CEBM, NICE, WHO, etc.)
* Many publications using GRADE (hard to find if abbreviation is used in abstract: ‘GRADE’ abbreviation: 113,295 hits; full name: 84 hits)

Slide 6: GRADE used in disability and rehabilitation (easily found ones only)

* Horticultural therapy for schizophrenia. *Cochrane Database Syst Rev*. 2014
* Multidisciplinary rehabilitation following botulinum toxin and other focal intramuscular treatment for post-stroke spasticity. *Cochrane Database Syst Rev*. 2013
* Non pharmacological interventions for spasticity in multiple sclerosis. *Cochrane Database Syst Rev*. 2013
* A systematic review of effectiveness of complementary and adjunct therapies and interventions involving equines. *Health Psychol*. 2013
* Kinematics of the cervical adjacent segments after disc arthroplasty compared with anterior discectomy and fusion: a systematic review and meta-analysis. *Spine (Phila Pa 1976)*. 2012
* KDIGO clinical practice guideline for the care of kidney transplant recipients. *Am J Transplant*. 2009

Slide 7: Four major characteristics of GRADE

* GRADE is for creating systematic reviews and for creating guidelines, with separate steps for each: (1) do a systematic review (or use one or more existing systematic reviews); (2) create the guideline
* GRADE is outcome-focused: evidence is reviewed and summarized separately for each (important) outcome (positive and negative) of an intervention or a diagnostic process; then cross-outcome evidence is used to make a recommendation, taking importance of each in mind
* GRADE offers little in the way of standard algorithms that need to be followed; it focuses instead on reviewers/guideline developers making explicit their values, assumptions, judgments, etc. that go into qualifying the evidence and creating the end product, and documenting these: transparency (in red text) is a central element
* GRADE deals with Tx and Dx only – not Px, etc.

Slide 8: GRADE’s fit into the guideline development process: 1

1. Prioritize problem

2. Establish review team and/or guideline panel

3. (in red text) Define questions to be addressed

4. (in red text) Find and critically appraise systematic reviews(s)

 and/or

 prepare protocol(s) for systematic review(s) and prepare systematic

 review(s): searches, selection of studies, data collection and analysis

5. ((in red text) Re)assess the relative importance of outcomes

6. (in red text) Prepare an evidence profile, including (1) an assessment of the

 quality of evidence for each outcome and (2) a summary of the findings

Slide 9: GRADE’s fit into the guideline development process: 2

7. (in red text) If developing guidelines: Assess the overall quality of evidence and decide on the direction and strength of the recommendation(s)

8. Draft the systematic review or guideline

9. Consult with stakeholders and/or external peer reviewers

10. Disseminate the review or guideline

11. Update review or guideline when needed

12. Adapt guideline, if needed

13. Prioritize recommendations for implementation

14. Implement or support implementation of the guideline

15. Evaluate the impact of the guideline and implementation strategies

16. Update the systematic review and guideline

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 10: Questions?

Slides 11 and 12 provide a schematic view of GRADE's process for developing recommendations.

- Slide 11 describes steps in the process common to systematic reviews and to guideline development.

- Slide 12 describe steps that are specific to guideline creation.

Slide 11: Schematic view of GRADE's process for developing recommendations (1). On the left side of slide 11 is a box on its side with the words, "Systematic reviewing steps." A bracket between that box and several boxes on the image indicate they relate to the Systematic reviewing steps.

At the top of the image is a yellow box with the words: **Question** (clinical, public health, health system, etc.) From the bottom of the Question box, five lines connect that box to five smaller boxes with the words Study 1, Study 2, Study 3, Study 4, and Study 5.

Below that row of five boxes are three boxes side by side, labeled Outcome 1, Outcome 2, and Outcome 3.

Lines from the bottom of Study 1 and Study 2 connect to Outcome 1 on the left and Outcome 2 in the center. A box with a dashed border is below Outcomes 1 and 2; it is labeled Important outcomes.

Lines from Study 3 connect to Outcome 2 and to Outcome 3.

Lines from Study 4 and Study 5 connect to Outcome 3 on the right. A box with a dashed border is below Outcome 3; it is labeled Critical outcome.

Below “Important” and “Critical” outcomes, a green box with a solid border contains the sentence: **Generate an estimate of effect for each outcome.**

Below the green box, a thick red arrow points to a blue box under it that contains four sentences:

**Rate the quality of the evidence *for each outcome* across studies.**

Reduce the rating as needed

Increase the rating as warranted

**Final rating of evidence quality for each outcome: *high, moderate, low or very low*.**

Under the blue box, a thick red arrow points down to the following slide (Slide 12)

Slide 12: Schematic view of GRADE's process for developing recommendations (2). On the left side, the bottom half is indicated by a box on its side with the words, "Guideline development steps." A bracket between that box and two boxes on the lower half of the image indicate they relate to the Guideline development steps.

A thick red arrow at the top center of the slide points to a blue box below it that contains the sentence:

**Rate the overall quality of the evidence** (lowest quality among critical outcomes)

Under that box is a thick red arrow points to another box, which contains two sentences:

**Decide on the direction (for/against) and grade strength (strong/weak) of the recommendation, considering:**

* Quality of the evidence
* Balance of desirable/undesirable outcomes
* Values and preferences of patients/clients/others

**Decide if any revision of direction or strength is necessary considering: *resource use*.**

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 13: Step 3. Define the questions to be addressed

* PICO(T)(S) recommended
	+ Population
	+ Intervention (incl. diagnostic intervention)
	+ Comparator
	+ Outcomes
		- Positive, negative, costs, other burden
		- Rated on a 1-9 scale of importance
			* 1-3: unimportant
			* 4-6: important
			* 7-9: critical
	+ Time point / time frame
	+ Setting: resource-rich vs resource-poor countries

Slide 14: Step 3. Rate the importance of the outcomes

* Ideally based on (systematic reviews of) patient ratings of relative importance of various outcomes
* Second best, ratings put together ad-hoc by patient groups
* Third best, proxy ratings by panel members
* 7-9: critical
* 4-6: important
* 1-3: of limited importance

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 15: Importance of outcomes and their evolution

* Step 3: Preliminary classification before the evidence is reviewed, to guide evidence searching
* Step 5: Reassessment of relative importance after the evidence has been reviewed, to make sure outcomes not considered in step 3 but found in literature review are given their proper weight (e.g. rare but serious adverse effects)
* Step 7: In making a recommendation and deciding on its strength, take the (revised) importance into account in judging the balance between desirable and undesirable effects of an intervention

Slide 16: Step 4. Find and critically appraise systematic reviews(s) and/or prepare protocol(s) for systematic review(s) and create systematic review(s): searches, selection of studies, data collection and analysis

* Given the need to extract outcomes and classify quality of evidence based on information that is not likely to be in the available systematic reviews, GRADE panel may need to redo existing reviews
* Evidence needs to be rated *separately for each outcome*, as the quality of evidence may differ by outcome

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 17: Table with 4 columns and 4 rows. The columns are Study design, Quality of evidence, rate evidence quality lower if: and Rate evidence quality higher if.

Column 1. Cell 1: study design; cell 2: RCT, accompanied by an arrow pointing to the column 2 cell next to it; cell 3: blank; cell 4: observational study accompanied by an arrow pointing to the column 2 cell next to it; cell 5 blank

Column 2. Cell 1: Quality of evidence; Cell 2 High; Cell 3: Moderate; Cell 4: Low; Cell 5: Very low.

Column 3. Cell 1: rate evidence quality lower if:. Cells 2, 3, 4 and 5 are merged, and contain the text:

Risk of bias

-1 Serious

-2 Very serious

Inconsistency

-1 Serious

-2 Very serious

Indirectness

-1 Serious

-2 Very serious

Imprecision

-1 Serious

-2 Very serious

Publication bias

-1 Likely

-2 Very likely

Column 4. Cell 1 Rate evidence quality higher if:; Cells 2, 3, 4 and 5 are merged, and contain the text:

Effect size

+1 Large

+2 Very large

Dose response

+1 Evidence of a gradient

Plausible confounding

+1 All would reduce a demonstrated effect OR +1 All would suggest a spurious effect when actual results show no effect

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 18: A more expansive listing of designs (Brozek et al. 2009)

*(Unable to clear copyright in time for presentation. You may download the article for free and view the content in Table 3.)*

Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schünemann HJ; GRADE Working Group. (2009). Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions**.** *Allergy,64*(5):669-77.

<http://onlinelibrary.wiley.com/doi/10.1111/j.1398-9995.2009.01973.x/full>

See Table 3:

[http://onlinelibrary.wiley.com/doi/10.1111/j.1398-9995.2009.01973.x/full#t3](http://onlinelibrary.wiley.com/doi/10.1111/j.1398-9995.2009.01973.x/full)

Slide 19: Current GRADE descriptions of the levels of evidence

Table with two columns: (1) Rank and (2) Definition

Row 1: Rank High; Definition-We are very confident that the true effect lies close to that of the estimate of the effect

Row 2: Rank Moderate; Definition-We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Row 3: Rank low; Definition-Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Row 4: Rank very low; Definition-We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 20: Rating down: study limitations resulting in increased risk of bias: ‘small d’ issues in RCTs

* Lack of allocation concealment
* Lack of blinding (patients, clinicians, outcome assessors)
* Incomplete accounting of patients and outcome events
	+ Attrition
	+ Failure to use intent-to-treat analysis
* Selective outcome reporting bias (‘publication bias in situ’)
* Other limitations
	+ Stopping early for benefit
	+ Use of unvalidated outcome measures (e.g. PROs !!!)
	+ Carryover effects in a cross-over RCT
	+ Recruitment bias in cluster-randomized trials

Slide 21: Rating down: study limitations resulting in increased risk of bias: ‘small d’ issues in observational studies

* Failure to develop and apply appropriate eligibility criteria, e.g.
	+ Under- and overmatching in case-control studies
	+ Exposed and unexposed cases in cohort studies selected from different populations
* Flawed measurement of both exposure and outcome, e.g.
	+ Differences in measurement of exposure (e.g. recall bias)
	+ Differential surveillance of the exposed and unexposed in cohorts
* Failure to adequately control confounding, e.g.
	+ Failure to accurately measure all prognostic factors
	+ Failure to match for/statistically adjust for prognostic factors
* Incomplete follow-up

Slide 22: Rating down: Imprecision

* Defined as: the confidence interval around the estimated effect size is unacceptably wide
* (in blue text) There is no definition of / cut-off for ‘too wide’
* Seriousness of the problem depends on the balance of benefits and costs (side effects, burdens) of a treatment
	+ If there are serious downsides to a treatment, there is a great need to know what exactly the effect size is, within a very narrow confidence interval
	+ If a treatment is cheap and has no or very minor side effects, we do not worry about a wide confidence interval (even if it may include 0?)

Slide 23: Rating down: Imprecision

* GRADE suggests taking into account:
	+ Patient value and preference judgments: importance of outcomes
	+ Adverse effects
	+ Burden on patient
	+ Resource use
	+ Sample size (individual and pooled studies) and event rate (number of events per 1,000 clients)

in setting criteria for ‘too wide’ and in deciding whether to rate down the evidence, one or even two levels

Slide 24: Rating down: inconsistency in magnitude of results (effect sizes) of studies

* In GRADE discussed as relevant to intervention studies using dichotomous outcomes only, not for Dx studies and not for treatment studies using continuous outcome measures
* Focus is on relative risk (risk ratio, hazard ratio, odds ratio) – relative risk often is stable from study to study even if absolute risk reduction varies based on population characteristics and other issues
* Relative risk may vary based on the following:
	+ Population
	+ Intervention (dose, cointerventions, comparator)
	+ Outcomes (e.g. duration of follow-up)
	+ Study methods (e.g. low-risk vs high-risk RCT)

Slide 25: Rating down: inconsistency in magnitude of results (effect sizes) of studies

* Effect-size differences due to population, intervention, outcome:
	+ should be expected
	+ can be formally tested based on a-priori hypotheses
	+ may result in different recommendations for different subgroups, outcomes, interventions
* When these factors do not explain inconsistency, reviewers should consider rating down evidence when:
	+ Point estimates vary widely across studies
	+ Confidence intervals show minimal or no overlap
	+ Statistical test for heterogeneity shows a low p value
	+ I2 (a measure of heterogeneity not taking sampling into account) is large
	+ Inconsistency reduces confidence in results in relation to a particular decision

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 26: Inconsistency that is immaterial to the decision to recommend the treatment

Plot graph with a range of .3 to 10. The Y axis is at the 1 point mark. Four data points are marked all less than 1. The data points have a confidence interval around them. The confidence interval for the data point at about .9 contains a vertical line drawn through 1, indicating that treatment and comparator are equally effective.

Slide 27: Inconsistency that should lead to a decision to rate down the evidence

Plot graph with a range of .3-10. The Y axis is at the 1 point mark. 5 data points are marked. 3 points are less than 1. 2 points are greater than 1. The data points have a confidence interval around them; none of the confidence intervals overlap with the vertical line.

Slide 28: Rating down: indirectness of evidence

* When the studies providing evidence match the clinical question [PICO(T)(S)], evidence is considered direct
* When there is no 100% match, the evidence is ‘indirect’, and reviewers should consider rating down the evidence
* Rating down should also be considered if two treatments of interest are not compared directly in the studies available, but through a third intervention, most likely a placebo (AKA network meta-analysis)

Slide 29: Three types of indirectness

Table with two columns indirectness and example.

Row 1- Indirectness- population Example- Cognitive rehabilitation was studied in patients with stroke, not in the group of interest, patients with TBI

Row 2- Indirectness- intervention Example- High-intensity exercise as part of cardiac rehabilitation is the intervention of interest, but most existing studies have investigated low-intensity exercise

Row 3- Outcome- intervention Example- Existing intervention studies have addressed increasing the ability of persons with SCI to perform various office tasks (ICF Activities), but none has studied actual return to work in an office setting (ICF Participation)

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 30: And the fourth type of indirectness

Diagram. Box on left reads: Indirect comparison: treatment B is 1.75/1.57=1.11 times better than treatment A?

Four boxes in the middle read: 1. Treatment A: 55% success, 2. Placebo: 35% success, 3. Placebo: 40% success, 4. Treatment B: 70% success

Middle boxes 1 and 2 are grouped together and linked to a box reading: Study 1: treatment A is 55/35=1.57 times better than placebo

Middle boxes 3 and 4 are grouped together and linked to a box reading: Study 2: treatment B is 70/40=1.75 times better than placebo

Slide 31: Rating down: publication bias

* Non-publication of ‘negative’ studies because of lack of interest by (1) author; (2) peer reviewer; (3) editor
* Non-publication of negative studies by big pharma
* Late publication because of ‘bouncing’ of the manuscript toward low-quality journals (time-lag bias)
* Only ‘positive’ studies are published, unless well-done large negative studies with sufficient power manage to ‘break through’
* Consequence: effect sizes in systematic reviews are artificially inflated toward the treatments studied

Slide 32: Rating down: publication bias funnel plot

Slide shows a equilateral triangle (the upside-down funnel) located in X-Y coordinates. The Y axis is labeled sample size with low on the bottom and high on the top. The X axis is Effect Size, with a vertical line drawn through the top angle of the triangle; the Left is Comparator is better. Middle is 0. Right is Treatment is better. Multiple date points are plotted. There is a horizontal red line in middle of the plot points, and approximately equidistant between the top angle of the triangle and its base. The vertical and horizontal lines divide the triangle in four quadrats. Quadrant 1 (left upper) has 3 plot points. Quadrant 2 (right upper) has 2 plot points. Quadrant 3 (left lower) has no plot points. Quadrant 4 (right lower) has 6 plot points; in addition, there is a plot point just to the right of quadrant 4. There is a box next to quadrants 1 and 2 with the text: Average effect size of large studies: -0.05. There is a box next to quadrants 3 and 4 with the text: Average effect size of small studies: 1.03.

Slide 33: Rating down: publication bias

* Detection of publication bias:
	+ Funnel plot (not very reliable – uses eye-balling)
	+ Fill and trim tests
	+ Tests based on the statistical significance reported
	+ Meta-analysis for each of successive years of publication (detection of time-lag bias)
* Counteracting publication bias: dig up unpublished studies (much easier with the increasing registration of studies, on e.g. clinicaltrials.gov )
* More feasible action: if one (preferable more) ‘detectors’ suggest publication bias, rate down evidence
* GRADE suggests going down one level only

Slide 34: Rating up the evidence

* Three primary reasons for rating up the quality of the evidence
	+ Large effect size
	+ Dose-response gradient
	+ All plausible confounders/biases increase our confidence in the estimated effect
* Likely occurring infrequently
* Mostly in cases of observational studies (including otherwise very low quality evidence, such as before-after studies and time series studies)

Slide 35: Rating up: large effect size

* Clinical interventions that practitioners have confidence in, and that indeed have a large effect size
	+ Modeling suggests that with relative risk (RR) > 2 or RR < 0.5 confounding is ‘not likely’ to be an explanation, and quality rating can be improved by 1 level
	+ With RR > 5 or RR < 0.2, this becomes ‘very unlikely’ and a 2 level quality rating increase may be justified
* Often onset of change is almost immediate, and improvement curve is steep
* Knowledge that without in intervention there would be deterioration or at best no improvement adds to our confidence
* Often there is indirect evidence to support our believing in a large treatment effect

Slide 36: Rating up: large effect size examples

* Instant improvement of mobility in patients with new onset SCI provided with a wheelchair
* <1% of 500 patients with SCI developing, over a year’s time, a pressure ulcer with a new model seat cushion, vs known incidence of >10% with various existing models
* Deteriorating mobility in women with arthritic knees, instantly reversed by knee replacement

Slide 37: Rating up: dose response

* If a (steep) dose-response gradient is found within studies or between studies, we are more likely to belief that the effect of the intervention is not a result of chance, or of poor study design and implementation

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 38: Dose-response relationship

There is a clustered column chart. The columns are Discharge Functional Independence Measure and Inpatient rehabilitation length of stay. There are five countries along the Y axis, USA, Canada, Australia, Germany, and England. There is a general trend that as the discharge Functional Independence Measure increases from one country to the next, so does the inpatient rehabilitation length of stay.

Slide 39: Does longer length of stay result in improved functional status?

There is a clustered column chart. The columns are Discharge Functional Independence Measure and Inpatient rehabilitation length of stay. There are five countries along the Y axis, USA, Canada, Australia, Germany, and England. Although the discharge functional independence measure increases from one country to the next, the inpatient rehabilitation length of stay is shows no clear trend line at all.

Slide 40: Rating up: confounders

* Because in observational research there is no randomization that will equally distribute all confounders (factors with an impact on the outcome) between the two groups that are being compared (at least when N is > 200 per group: law of large numbers), these need to be taken into account statistically, e.g. by means of
	+ ANCOVA
	+ Propensity scores
* Only factors that were actually measured, and measured well, can be statistically controlled for
* What remains is ‘residual confounding’ or ‘residual biases’

Slide 41: Rating up: confounders

* In unusual circumstances, ALL plausible confounders unaccounted for in the adjusted analysis would result in an UNDER estimate of the treatment effect (rather than an overestimate, the usual problem), e.g.
	+ Only the sicker, less able patients only get the treatment, yet they have better outcomes than the untreated patients with more ability at baseline
	+ (Regression discontinuity design)

Slide 42: Regression discontinuity design

Line graph. Y axis is Discharge score on XYZ test with a range of 0 to 60. X axis is Admission score on XYZ test with a range for 10 to 60. There is a line in the middle of the X axis (at value 35) which an arrow designates as representing the cut point. To the left of this vertical line is the Experimental intervention group and to the right is the No intervention group. Two regression lines are plotted, separately for the Experimental intervention group and the No intervention group. The regression line for the latter coincides with the main diagonal, suggesting that admission and discharge XYZ scores are very similar for this group. The extension of the regression line of the No Intervention group suggests where the experimental intervention group would have been without the intervention. In actuality, the regression line for this group is displaced upward, such that on the Discharge XYX test the average “Experimental intervention group” member scores as high as the average “No intervention group” member.

Slide 43: Step 5. (Re)assess the relative importance of outcomes (guideline developers only)

* The careful review of studies may suggest outcomes of the intervention that were not considered initially, but are important if not critical
* Especially (rare) adverse events with catastrophic consequences – e.g. permanent disability

Slide 44: Step 6. Prepare an evidence profile, including (1) an assessment of the quality of evidence for each outcome and (2) a summary of the findings

* Evidence profiles (EPs)
	+ Include explicit judgment of each factor that determines the quality of evidence for each outcome (limitations, inconsistency, indirectness, imprecision & large effect size, dose-response, plausible confounding)
	+ Include data – e.g. N of patients in pooled group(s), pooled relative and absolute risk, etc.
	+ Are prepared for review authors, and for anyone who questions a quality assessment

Slide 45: Table 1. GRADE evidence profile: antibiotics for children with acute otitis media.

This is a table of 11 columns and 5 rows.

(Column 1-5 are listed under - Quality Assessment)

Column headings:

Column 1 – Limitations

Column 2 – Inconsistency

Column 3 – Indirectness

Column 4 – Imprecision

Column 5 – Publication Bias

(Column 6-11 are listed under – Summary of findings)

(Column 6 and 7 are listed under - Number of patients)

Column 6 – Placebo

Column 7 – Antibiotics

Column 8 – Relative Risk (95% CI)

Column 9 – Control risk

Column 10 – Risk difference (95% CI)

Column 11 – Quality

(Column 9-11 are listed under – Absolute risk)

Rows begin with label: No of studies (Design)

Row 1: Pain at 24h 5 (RCT):

No serious limitations; No serious inconsistency; No serious indirectness; No serious imprecision; Undetected; 241/605; 223/624; RR 0.9(0.78-1.04); 367/1,000; Not Significant; High

Row 2: Pain at 2-7 d 10 (RCT):

No serious limitations; No serious inconsistency; No serious indirectness; No serious imprecision; Undetected; 303/1,366; 228/1,425; RR 0.72 (0.062-0.83); 257/1,000; 62 fewer per 1,000 (44-98); High

Row 3: Hearing, inferred from the surrogate outcome abnormal tympanometry – 1 mo 4 (RCT):

No serious limitations; No serious inconsistency; Serious indirectness (because of indirectness of outcome); No serious imprecision; Undetected; 168/460; 153/467; RR 0.89 (0.75-1.07); 350/1,000; Not significant; Moderate.

Row 4: Hearing, inferred from the surrogate outcome abnormal tympanometry-3 mo 3 (RCT):

No serious limitations; No serious inconsistency; Serious indirectness (because of indirectness of outcome); No serious imprecision; undetected; 96/398; 96/410; RR 0.97 (0.76-1.24); 234/1,000; Not Significant; Moderate

Row 5: Vomiting, diarrhea, or rash 5 (RCT):

No serious limitations; Serious inconsistency (because of inconsistency in absolute effects); No serious indirectness; No serious imprecision; Undetected; 83/711; 110/690; RR 1.38 (1.09-1.76); 113/1,000; 43 more per 1,000 (10-86); Moderate

From Table 1. GRADE evidence profile: antibiotics for children with acute otitis media from "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), p. 387. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 46: Step 6. Prepare an evidence profile, including (1) an assessment of the quality of evidence for each outcome and (2) a summary of the findings

* Summaries of findings (SoFs)
	+ Includes assessment of the quality of the evidence, but not detailed judgments
	+ Concise summary of key information
	+ Prepared for a broader audience, including end users of the systematic review / guidelines

Slide 47: Table 2. Summary of finding: antibiotics for acute otitis media in children.

Antibiotics compared with placebo for acute otitis media in children

Patient or population: Children with acute otitis media

Setting: High – and middle-income countries

Intervention: Antibiotics

Comparison: Placebo

This is a table of 5 rows by 6 columns.

Column headings:

Columns 1-2 are listed under: Estimated risks (95% CI)

Column 1 – Control risk – Placebo

Column 2 – Intervention risk – Antibiotics

Column 3 – Relative effect (95% CI)

Column 4 – No. of Participants (studies)

Column 5 – Quality of the evidence (GRADE)

Column 6 – Comments

Rows begin with label: Outcomes

Row 1: Pain at 24h:

367 per 1,000; 330 per 1,000 (286-382); RR 0.9 (0.78-1.04); 1229 (5); High

Row 2: Pain at 2-7 d:

257 per 1,000; 185 per 1,000 (159-213); RR 0.72 (0.62-083); 2791 (10); High

Row 3: Hearing, inferred from the surrogate outcome abnormal tympanometry -1 mo:

350 per 1,000; 311 per 1,000 (262-375); RR 0.89 (0.75-1.07); 927 (4); Moderate

Row 4: Hearing, inferred from the surrogate outcome abnormal tympanometry-3 mo

234 per 1,000; 227 per 1,000 (178-290); RR 0.97 (0.76-1.24); 808 (3); Moderate

Row 5: Vomiting, diarrhea, or rash:

113 per 1,000; 156 per 1,000 (123-199); RR 1.38 (1.09-1.76); 1,401 (5); Moderate; Ideally, evidence from nonotitis trials with similar ages and doses (not obtained) might improve the quality of evidence.

From Table 2. Summary of finding: antibiotics for acute otitis media in children from "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), p. 388. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 48: Step 7. If developing guidelines: Assess the overall quality of evidence and decide on the direction and strength of the recommendation(s)

* The level of evidence quality (high, moderate, low, very low) assigned to the *critical* outcome that has the *lowest* quality evidence is the level given to the intervention with respect to *all its outcomes*
* Recommendations are in favor of intervention or against
* GRADE uses two strengths of recommendations only: strong and weak= discretionary= conditional
* Direction and strength of recommendation are based on
	+ Quality of the evidence
	+ Balance of desirable vs undesirable outcomes
	+ Values and preferences of patients
	+ Resource use (cost) of the intervention

From "GRADE Guidelines: 1. Introduction-GRADE Evidence Profiles and Summary of Findings Tables" by G. Guyatt, A. D. Oxman, E. A. Akl, R. Kunz,  G. Vist, J. Brozek, J., . . . H. J. Schunemann, 2011, Journal of Clinical  Epidemiology 64(4), 383-394. Copyright © 2011 by Elsevier Inc. All rights reserved. SEDL used with fee-based permission from Copyright Clearance Center's Rightslinks service.

Slide 49: Questions?

Slide 50: Advantages of GRADE (compared to many/ all alternative evidence development approaches

1. Clear separation between quality of evidence and strength of recommendations\*
2. Explicit and comprehensive criteria for downgrading or upgrading quality of evidence
3. Explicit consideration of the relative importance of various outcomes to patients
4. Explicit acknowledgement of values and preferences that are used/assumed when making recommendations
5. Transparent process of moving from evidence to recommendations

Slide 51: Advantages of GRADE (compared to many/ all alternatives

6. Explicit advice to make recommendations about the most

 appropriate course of action, even when very little

 evidence is available

7. Grading the strength only for recommendations about the

 diagnostic or therapeutic course of action, but not about

 prognosis or etiology

8. Clear and pragmatic interpretation of ‘strong’ and ‘weak’

 recommendations

9. Balance between simplicity and methodological

 comprehensiveness

(Brozek et al., 2009, Guyatt et al., 2011)

Slide 52: Limitations of GRADE

1. GRADE does not deal with questions of risk or prognosis
2. Application of GRADE to ‘motherhood’ statements is problematic: why spend time on recommendations everyone agrees on (“rehabilitation is good”), what evidence would there be relevant to such issues
3. GRADE covers steps 3-7 out of 1-16 only
4. GRADE has not been developed well / not tested out on questions on diagnostic tests, public health, health systems.
5. GRADE does not offer guidance about designs other than (1) RCTs and (2) observational studies including a control group

Slide 53: Limitations of GRADE

6. GRADE does not eliminate disagreements between

 panelists – just makes transparent judgments about

 importance of outcomes, quality of evidence, strength of

 recommendations

7. GRADE approach in assessing risk of bias emphasizes

 simplicity and parsimony over completeness. Specifically,

 a scale to quantitatively measure ‘methodological quality’

 is not used

8. Empirical evidence underlying rules for grading down/

 grading up evidence absent or not strong

(Brozek et al., 2009, Guyatt et al., 2011)

Slide 54: Questions?

Slide 55: Wrapping Up

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@air.org

Please fill out the brief Evaluation Form:

<http://www.surveygizmo.com/s3/1697334/Evaluation-Webinar-Series-Session-3>

Slide 56: Disclaimer

This webinar was developed for grant number H133A120012 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR), Office of Special Education and Rehabilitative Services (OSERS), U.S. Department of Education. However, the contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the federal government.