Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence

Question	Step 1 (Level 1*)	Step 2 (Level 2*)	Step 3 (Level 3*)	Step 4 (Level 4*)	Step 5 (Level 5)
How common is the problem?	Local and current random sample surveys (or censuses)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
Is this diagnostic or monitoring test accurate? (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non-consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
What will happen if we do not add a therapy? (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
Does this intervention help? (Treatment Benefits)	Systematic review of randomized trials or *n*-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
What are the COMMON harms? (Treatment Harms)	Systematic review of randomized trials, systematic review of nested case-control studies, *n*-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
What are the RARE harms? (Treatment Harms)	Systematic review of randomized trials or *n*-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
Is this (early detection) test worthwhile? (Screening)	Systematic review of randomized trials	Randomized trial	Non -randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

*** Level may be graded down on the basis of study quality, imprecision, indirectness (study PICO does not match questions PICO), because of inconsistency between studies, or because the absolute effect size is very small; Level may be graded up if there is a large or very large effect size.**

**** As always, a systematic review is generally better than an individual study.**

How to cite the Levels of Evidence Table

OCEBM Levels of Evidence Working Group*. "The Oxford 2011 Levels of Evidence".

Oxford Centre for Evidence-Based Medicine. http://www.cebm.net/index.aspx?o=5653

** OCEBM Table of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson*

Presented in ***KT Update*** (Vol. 1, No. 5 - August 2013) [http://www.ktdrr.org/products/update/v1n5/]
An e-newsletter from the Center on Knowledge Translation for Disability and Rehabilitation Research

Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development

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

---

Dr. Marcel Dijkers, rehabilitation researcher at the Icahn School of Medicine at Mount Sinai, presents another in a series of brief articles around evidence-based research and knowledge translation topics. This article explains the GRADE process (Grading of Recommendations Assessment, Development and Evaluation) and explores its usefulness for rehabilitation and disability research.

GRADE (Grading of Recommendations Assessment, Development and Evaluation) is a well-developed formal process to rate the quality of scientific evidence in systematic reviews and to develop recommendations in guidelines that are as evidence-based as possible. GRADE was developed by an international panel, including members of some of the premier evidence-based practice centers (McMaster, Harvard, the Norwegian and German Cochrane Centres, etc.). While there were some earlier publications,[1-7] a series of papers published in the *Journal of Clinical Epidemiology* from 2011 to 2013 constitute the most complete and systematic expose.[8-22] More information can be found on the GRADE Working Group's website (http://www.gradeworkinggroup.org).

A number of panels and agencies have adopted GRADE, among others the Cochrane Collaboration (the Effective Practice and Organisation of Care group, the Public Health and other groups), World Health Organization (various guideline development groups), England's National Institute for Health and Clinical Excellence (NICE); the Canadian Task Force on Preventive Health Care, the Norwegian Knowledge Centre for the Health Services, the CDC's Advisory Committee on Immunization Practices, the Kaiser Permanente National Guideline Program, and some groups in the Campbell Collaboration. Some use it with modifications (not recommended by GRADE), and some report minor or major challenges in using the GRADE process unmodified.[23]

GRADE was designed for reviews and guidelines that examine alternative clinical management strategies or interventions, which may include no intervention or current best management. In developing GRADE, the authors considered a wide range of clinical questions, including diagnosis, screening, prevention, and therapy. For that reason, the system can also be applied to rehabilitation, public health, and health systems questions.

GRADE is much more than a rating system, such as those published by various Evidence-Based Practice (EBP) organizations. It offers a transparent and structured process for developing and presenting evidence summaries for systematic reviews and guidelines and for carrying out the steps involved in developing recommendations. GRADE specifies an approach to framing questions,[9] choosing outcomes of interest and rating their importance,[9] evaluating the evidence,[10] including making explicit the risk of various biases,[11,12] and taking into account issues of imprecision (i.e. broad confidence intervals),[13] inconsistency of results between studies,[14] and indirectness (i.e. using evidence from a similar population, e.g. stroke instead of traumatic brain injury).[15]

GRADE incorporates evidence with explicit consideration of the values and preferences of patients and society at large to arrive at recommendations. Furthermore, it provides clinicians and patients/clients with a guide to using those recommendations in clinical practice, and policy makers with a guide to their use in health policy.

Based on the recommendations by Johnston and Dijkers[24] for improved evidence standards, a review of a number of existing approaches to systematic reviewing of evidence and developing recommendations (Cochrane Collaboration; Campbell Collaboration, American Academy of Neurology, Centre for Evidence-Based Medicine, among others) likely comes to the conclusion that GRADE is the most flexible methodology with respect to evaluating the evidence (downgrading, upgrading, handling indirect evidence, etc.). It also goes beyond the other systems where it concerns the translation of evidence into recommendations. The special emphasis in GRADE on the values and preferences of consumers (which now is being adopted by others) fits eminently with the traditional emphasis in rehabilitation and disability studies.

The GRADE methodology is applicable whether the quality of the relevant evidence is high or low. The GRADE system was among the first to lay out a systematic way of evaluating whether evidence should be downgraded—for instance, a randomized controlled trial (RCT) executed with poor allocation concealment and high attrition should not be considered to be equivalent to a well-done RCT (see Table 1).

In addition, GRADE was the first to specify under what circumstances evidence from a study may be upgraded—for instance, when an effect size is very large, a dose-response gradient is shown, or other circumstances would suggest that what traditionally has been considered a “rather weak design” (e.g. a case-control study) may produce evidence that is of a level produced by an average RCT.[16]

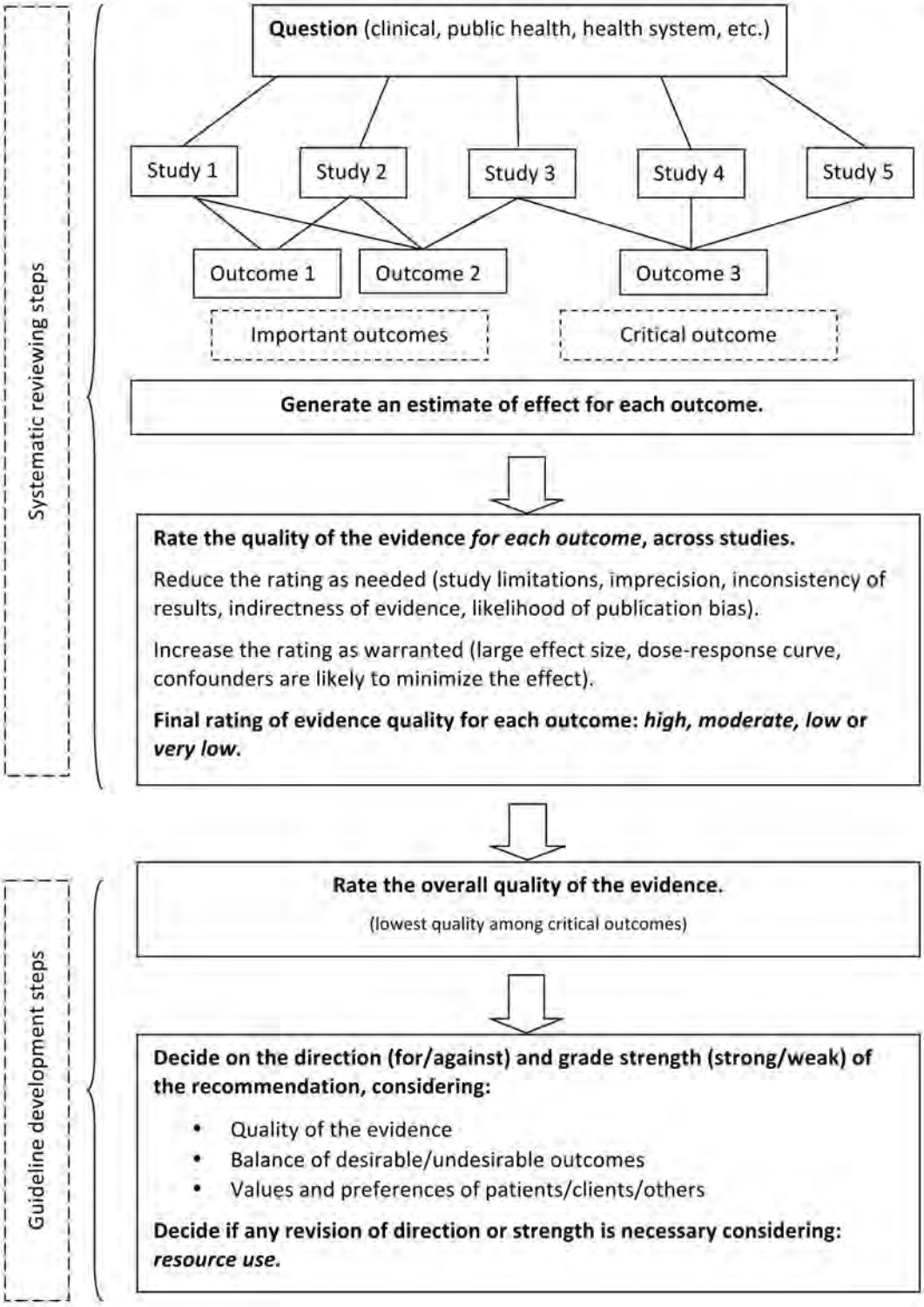
Table 1. Factors that may lead to downgrading or upgrading of evidence in the GRADE approach

Downgrading
1. Serious risk of bias
2. Serious inconsistence between studies
3. Serious indirectness
4. Serious imprecision
5. Likely publication bias
Upgrading
1. Large effect size
2. Dose-response gradient
3. All plausible confounding would reduce a demonstrated effect
4. All possible confounding would suggest a spurious effect when the actual results show no effect

Another advantage of GRADE is that it requires the systematic reviewer to make explicit his or her judgment of each factor that determines the quality of evidence for each outcome. Because alternative diagnostic or therapeutic approaches may all have a balance of positive and negative outcomes (costs, side effects, positive effects in various domains), a guideline developer needs to find a way to systematically identify these, and weigh evidence for all of them simultaneously in making recommendations; GRADE offers a systematic approach to resource use[17] and to handling multiple outcomes.[18] Finally, a computer program (GRADEpro) with its associated help file facilitates the development of evidence tables (in GRADE called evidence profiles, or EPs[19,20]) and summary of findings (SoFs) tables that are based on the EPs.

Figure 1 presents a schematic view of GRADE's process for developing recommendations; the top half describes steps in the process common to systematic reviews and to guideline development, and the lower half describe steps that are specific to guideline creation. One begins by defining the question in terms of the populations, alternative management strategies (an intervention, sometimes experimental and a comparator, sometimes standard care), and all patient-important outcomes (in this case three). The authors have provided guidance as to which clinical and other questions are suitable for answering with GRADE (or with any systematic review approach, for that matter) and for collecting evidence.

Figure 1. Overview of the GRADE process (modified from Guyatt et al.[8])



For guidelines, one classifies the outcomes as either critical (one outcome in Figure 1) or important but not critical (two outcomes). A systematic search leads to inclusion of relevant studies (in this scheme, five such studies). Systematic reviewers or guideline authors then use the data from the individual eligible studies to generate a best estimate of the effect on each patient-important outcome and an index, typically a confidence interval (CI), of the uncertainty associated with that estimate.

The Figure illustrates that evidence must be summarized for each patient-important outcome—the summaries ideally coming from optimally conducted systematic reviews. For each comparison of alternative management strategies, all outcomes should be presented together in one EP or SoFs table. It is likely that all studies relevant to a rehabilitation or disability question will not provide evidence regarding every outcome. For example, Figure 1 shows the first study providing evidence for the first and second outcome, the third study for outcomes two and three, and so on. Indeed, there may be no overlap between studies providing evidence for one outcome and those providing evidence for another. For instance, RCTs may provide the relevant evidence for benefits, and observational studies provide the evidence for rare but serious adverse effects.

In the GRADE approach, RCTs start as high-quality evidence and observational studies as low-quality evidence to support estimates of intervention effects. As described above, five factors may lead to rating down the quality of evidence and three factors may lead to rating up (see Table 1). Ultimately, the quality of evidence for each outcome falls into one of four categories from high to very low. Systematic review and guideline authors use this approach to rate the quality of evidence for each outcome across studies (i.e., for a body of evidence). This does not mean rating each study as a single unit. Rather, GRADE is “outcome centric” in that a rating is made for each outcome, and quality may differ—indeed, is likely to differ—from one outcome to another within a single study and across a body of evidence.

Guideline developers (but not systematic reviewers) then review all the information to make a decision about which outcomes are critical and which are important, and come to a final decision regarding the rating of the overall quality of evidence. They next need to consider the direction and strength of recommendation. The balance between desirable and undesirable outcomes and the application of patients’ values and preferences determine the direction of the recommendation; these same factors, along with the quality of the evidence, determine the strength of the recommendation. Both direction and strength may be modified after taking into account the resource use implications of the alternative management strategies.[22]

Because most existing systematic reviews do not adequately address all relevant outcomes in a single document, the GRADE process may require relying on more than one systematic review. Systematic reviews often address more than one comparison. They may evaluate an intervention in two disparate populations or examine the effects of a number of interventions. Such reviews are likely to require more than one SoFs table. For example, a review of cognitive remediation may evaluate the effectiveness of training of executive function for different populations, such as those with mild versus moderate traumatic brain injury.

GRADE has been used for a number of systematic reviews and guidelines, including quite a few that would be of interest to professionals with an interest in rehabilitation and disability. Examples: nonpharmacologic interventions for osteoarthritis[25] and for spasticity in Multiple Sclerosis (MS),[26] rehabilitation interventions for nonambulatory MS patients;[27] music interventions for psychological and physical outcomes in cancer;[28] art therapy for psychosomatic disorders;[29] exercises[30] and manipulation[31] for chronic neck pain; interventions for depression in the workplace;[32] parent interventions for children with intellectual disabilities;[33] childhood adversity as a cause of schizophrenia;[34] family-based cognitive-behavioral therapy for children and adolescents with obsessive-compulsive disorder;[35] the benefits of physical activity for youth with developmental disability[36] and interventions to enhance return to work by cancer patients[37]—quite a variety. It may be worth your while to investigate whether GRADE is making the grade.

References

1. Brozek JL, Akl EA, Alonso-Coello P, et al. 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*. 2009;64(5):669-677. doi: 10.1111/j.1398-9995.2009.01973.x
2. Brozek JL, Akl EA, Jaeschke R, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy*. 2009;64(8):1109-1116. doi: 10.1111/j.1398-9995.2009.02083.x
3. Brozek JL, Akl EA, Compalati E, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 3 of 3. The GRADE approach to developing recommendations. *Allergy*. 2011;66(5):588-595. doi: 10.1111/j.1398-9995.2010.02530.x

4. Terracciano L, Brozek J, Compalati E, Schunemann H. GRADE system: New paradigm. *Curr Opin Allergy Clin Immunol*. 2010;10(4):377-383. doi: 10.1097/ACI.0b013e32833c148b
5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926. doi: 10.1136/bmj.39489.470347.AD
6. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490. doi: 10.1136/bmj.328.7454.1490
7. Guyatt GH, Oxman AD, Kunz R, et al. What is "quality of evidence" and why is it important to clinicians? *BMJ*. 2008;336(7651):995-998. doi: 10.1136/bmj.39490.551019.BE
8. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-394. doi: 10.1016/j.jclinepi.2010.04.026
9. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol*. 2011;64(4):395-400. doi: 10.1016/j.jclinepi.2010.09.012
10. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406. doi: 10.1016/j.jclinepi.2010.07.015
11. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-415. doi: 10.1016/j.jclinepi.2010.07.017
12. Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*. 2011;64(12):1277-1282. doi: 10.1016/j.jclinepi.2011.01.011
13. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol*. 2011;64(12):1283-1293. doi: 10.1016/j.jclinepi.2011.01.012
14. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. *J Clin Epidemiol*. 2011;64(12):1294-1302. doi: 10.1016/j.jclinepi.2011.03.017
15. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol*. 2011;64(12):1303-1310. doi: 10.1016/j.jclinepi.2011.04.014
16. Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol*. 2011;64(12):1311-1316. doi: 10.1016/j.jclinepi.2011.06.004

17. Brunetti M, Shemilt I, Pregno S, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol*. 2013;66(2):140-150. doi: 10.1016/j.jclinepi.2012.04.012
18. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66(2):151-157. doi: 10.1016/j.jclinepi.2012.01.006
19. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol*. 2013;66(2):158-172. doi: 10.1016/j.jclinepi.2012.01.012
20. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol*. 2013;66(2):173-183. doi: 10.1016/j.jclinepi.2012.08.001
21. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: The significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-725. doi: 10.1016/j.jclinepi.2012.03.013
22. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol*. 2013 Jul;66(7):726-35. doi: 10.1016/j.jclinepi.2013.02.003
23. Rehfuess EA, Akl EA. Current experience with applying the GRADE approach to public health interventions: An empirical study. *BMC Public Health*. 2013;13:9-2458-13-9. doi: 10.1186/1471-2458-13-9
24. Johnston MV, Dijkers MP. Toward improved evidence standards and methods for rehabilitation: Recommendations and challenges. *Arch Phys Med Rehabil*. 2012;93(8 Suppl):S185-99. doi: 10.1016/j.apmr.2011.12.011
25. Hochberg MC, Altman RD, April KT, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res*. 2012;64(4):465-474. doi: 10.1002/acr.21596
26. Amatya B, Khan F, La Mantia L, Demetrios M, Wade DT. Non pharmacological interventions for spasticity in multiple sclerosis. *Cochrane Database Syst Rev*. 2013;2:CD009974. doi: 10.1002/14651858.CD009974.pub2
27. Toomey E, Coote SB. Physical rehabilitation interventions in nonambulatory people with multiple sclerosis: A systematic review. *Int J Rehabil Res*. 2012;35(4):281-291. doi: 10.1097/MRR.0b013e32835a241a
28. Zhang JM, Wang P, Yao JX, et al. Music interventions for psychological and physical outcomes in cancer: A systematic review and meta-analysis. *Support Care Cancer*. 2012;20(12):3043-3053. doi: 10.1007/s00520-012-1606-5

29. Holmqvist G, Lundqvist Persson C. Is there evidence for the use of art therapy in treatment of psychosomatic disorders, eating disorders and crisis? A comparative study of two different systems for evaluation. *Scand J Psychol*. 2012;53(1):47-53. doi: 10.1111/j.1467-9450.2011.00923.x
30. Kay TM, Gross A, Goldsmith CH, et al. Exercises for mechanical neck disorders. *Cochrane Database Syst Rev*. 2012;8:CD004250. doi: 10.1002/14651858.CD004250.pub4
31. Lin JH, Chiu TT, Hu J. Chinese manipulation for mechanical neck pain: A systematic review. *Clin Rehabil*. 2012;26(11):963-973. doi: 10.1177/0269215512441485
32. Furlan AD, Gnam WH, Carnide N, et al. Systematic review of intervention practices for depression in the workplace. *J Occup Rehabil*. 2012;22(3):312-321. doi: 10.1007/s10926-011-9340-2
33. Einfeld SL, Stancliffe RJ, Gray KM, et al. Interventions provided by parents for children with intellectual disabilities in low and middle income countries. *J Appl Res Intellect Disabil*. 2012;25(2):135-142. doi: 10.1111/j.1468-3148.2011.00678.x
34. Matheson SL, Shepherd AM, Laurens KR, Carr VJ. A systematic meta-review grading the evidence for non-genetic risk factors and putative antecedents of schizophrenia. *Schizophr Res*. 2011;133(1-3):133-142.
35. Gomes JB, Matte BC, Vivan A, et al. Cognitive behavioral therapy with family intervention for children and adolescents with obsessive-compulsive disorder: A systematic review. *Revista de Psiquiatria do Rio Grande do Sul*. 2011;33(2):121-127.
36. Johnson CC. The benefits of physical activity for youth with developmental disabilities: A systematic review. *Am J Health Promot*. 2009;23(3):157-167.
37. de Boer AG, Taskila T, Tamminga SJ, Frings-Dresen MH, Feuerstein M, Verbeek JH. Interventions to enhance return-to-work for cancer patients. *Cochrane Database Syst Rev*. 2011 Feb 16;(2):CD007569. doi: 10.1002/14651858.CD007569.pub2

Suggested citation:
Dijkers, M. (2013). Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update (1)*5. Austin, TX: SEDL, Center on Knowledge Translation for Disability and Rehabilitation Research.
http://www.ktdrr.org/products/update/v1n5/dijkers_grade_ktupdatev1n5.html
http://www.ktdrr.org/products/update/v1n5/dijkers_grade_ktupdatev1n5.pdf

KT Update was developed by the Center on Knowledge Translation for Disability and Rehabilitation Research (KTDRR) through grant #H133A120012 to SEDL from the National Institute on Disability and Rehabilitation Research (NIDRR) in the US Department of Education's Office of Special Education and Rehabilitative Services (OSERS). However, these contents do not necessarily represent the policy of the US Department of Education, and you should not assume endorsement by the federal government.

SEDL operates the Center on KTDRR. SEDL is an Equal Employment Opportunity/Affirmative Action Employer and is committed to affording equal access to education and employment opportunities for all individuals. SEDL does not discriminate on the basis of age, sex, race, color, creed, religion, national origin, sexual orientation, marital or veteran status, or the presence of a disability.