Laboratory Medicine Best Practices
Guide to Rating Study Quality
August 27, 2008

Individual study quality ratings are based on four dimensions of study quality:

- Study
- Practice
- Outcome Measure(s)
- Findings/Result(s)

The principles and guidelines for making judgments along these four dimensions are outlined below.

Our main goal is to judge whether sufficient evidence is available concerning a practice's effectiveness to support a recommendation of "best practice" (that is, a practice likely to be effective in improving one or more outcomes of interest in comparison to other commonly used practices).

The system is designed to be inclusive, so we can make best use of the limited data available. Our methods for rating study quality do not penalize studies for not using a randomized design or for not being published in a peer reviewed journal. This approach acknowledges that many practices in laboratory medicine do not lend themselves to evaluation by traditional research designs, and much useful evidence may be obtained though our network affiliates whose priority on service delivery takes precedence over publishing.

Evaluating Study Quality

The four study quality dimensions are rated separately, with a rating score assigned up to the maximum for a given dimension. The rating scores for each dimension are added to reach a single summary score reflecting overall study quality. A total of 10 points are available to each study, with points subtracted from the maximum point total according to the guidance below. In this scheme, a rating of zero in any one of the four categories is sufficient to exclude a study from further consideration as evidence for a "best practice" recommendation.

Dimension 1. Study (3 points maximum)

Assess the likely generalizability of the results by evaluating:

- Study setting
- Sample characteristics (representativeness sufficient for practice)
- Potential study biases (study design, time period/duration and sample selection methods)

Criteria for point deduction

- Facility description
  - Deduct 1 point if the study location is sufficiently distinctive that the results obtained through that setting may not be generalizable to other settings.
  - Deduct 2 points if the study location is sufficiently distinctive that the results obtained through that setting are unlikely to be generalizable to other settings
  - Score 0 points if it is clear that the setting or situation is unique such that the results cannot generalize to other settings.
- Study design/ study period/ patient population
  - Deduct 1 point if the sample (either subjects or tests) may not be representative of the likely results of the practice with respect to how the sample was obtained or identified
  - Deduct 2 points if the sample (either subjects or tests) are probably unrepresentative of the results of the practice with respect to how the sample was obtained or identified.
  - Score 0 points if the sample is sufficiently unrepresentative based on how it was obtained/identified to clearly nullify the generalizability of the results.
- Potential study bias:

- Deduct 1 point if the study design, time period and sample selection methods may introduce a study bias that would substantially affect results (i.e., may produce study results interpreted as inconsistent with the true results)
- Deduct 2 points if the study design, time period and sample selection methods are likely to introduce a study bias that would substantially affect results (i.e., would likely produce study results interpreted as inconsistent with the true results)
- Score 0 points if there is reason to believe that the study characteristics can not produce results representative of the practice

Dimension 2. Practice (2 points maximum)

Assess the description of the practice and its adequacy.

Criteria for point deduction

- Description of the practice
  The practice should be well enough described to meaningfully distinguish it from alternative practices and provide a clear understanding of its requirements and characteristics (does not require that the description be exhaustive or support exact replication)..
  - Deduct 1 point if the practice and its basic characteristics are not sufficiently identified.
  - Score 0 points if the practice and its basic characteristics can not be clearly identified.

- Adequacy of practice description
  Ideally, seven components of practice description would be addressed: a) content, b) implementation, c) population / setting, d) training, e) requirements, f) cost, and g) staff responsible and implementing. However, detailed information on all components is not necessary to evaluate the effectiveness and feasibility of implementing a practice, and is typically not provided (e.g., cost and training).
  - Deduct 1 point if an important aspect/component that is likely to critically affect implementation of the practice is not well described.

Dimension 3. Outcome Measure (2 points maximum)

Outcome measures capture the result of implementing a practice. Rating scores reflect their face validity for capturing the outcome(s) of interest, whether the methods used to record results provide an incomplete or inaccurate record of the impact of a practice.

Most studies use multiple outcome measures. Raters should concentrate on measures that are directly related to the review question, which relates to health care quality (Institute of Medicine domains: safe, timely, effective, patient-centered, efficient, and equitable), and may ignore secondary measures, especially those gauging implementation feasibility.

- Face validity: The measure should capture the outcome being estimated.
  - Deduct 1 point if:
    - measure does not capture well the outcome being estimated

    OR
    - the ‘best’ measure from a study estimates an outcome that is only modestly related to the evidence review question (e.g., provider satisfaction, compared with change in an error rate)
  - Score 0 points if:
    - The ‘best’ measure from a study is confounded by: the practice itself (that is, the outcome is a direct result of the practice which was not available or applicable to the comparison)

    OR
    - The ‘best’ measure from a study is confounded by: the context in which the practice has been implemented (that is, the outcome is unlikely to be clearly attributed to the practice).

    OR

        - the 'best' measure from a study is not directly related to the evidence review question.

- Recording method: The method for recording or documenting practice results should be reliable and accurate.
    - Deduct 1 point if the method(s) of recording:
        - is not described
        - does not accurately capture all instances of the outcome
    - Score 0 if the method of recording the outcome is unreliable.

Dimension 4. Results/Findings (3 points maximum)

Results are affected by each of the dimensions of quality previously discussed. With this dimension, a narrow set of quality factors relating to (1) sample sufficiency, (2) appropriateness of statistical analysis and,(3) uncontrolled deviations along with results/conclusions bias.

Criteria for point deduction

- Sample sufficiency

  Many of the outcomes of interest are rare events. If too few observations are obtained or if the measurement period is insufficient to capture these events the measure may provide an inaccurate representation of the effect of the practice. Even among more common events, there may also be considerable variation in the number or rate of events over time. The period of measurement should be sufficiently long to allow robust estimates of the impact of the practice.

    - Deduct 1 point if:
        - the measurement period may be insufficient to allow a robust estimate of the impact of a practice

        OR
        - statistical power is not discussed AND the sample may be too small to allow a robust estimate of the impact of a practice
    - Deduct 2 points if
        - The number of subjects is not reported

        OR
        - The measurement period is not reported

        OR
        - the measurement period is likely insufficient to allow a robust estimate of the impact of a practice

        OR
        - statistical power is not discussed AND the sample is likely too small to allow a robust estimate of the impact of a practice

- Appropriateness of statistical analysis
    - Deduct 1 point if the study:
        - compares two practices and their estimates are based on data collected during notably different time periods (e.g., baseline or standard care in 2001 and intervention in 2005)

        OR
        - does not provide data sufficient to allow calculation of an effect size

    - Deduct 2 points if:
        - different measures or different recording practices are used when comparing the results of two practices

        OR

        - an inappropriate statistical analysis *and* insufficient data to allow calculation of an effect size

- Uncontrolled deviations and results/conclusions bias

    - Deduct 2 points if
        - Results/effect size reported are not clearly attributable to the practice being evaluated, but instead are likely related to significantly different practice(s) (e.g., major changes in staffing, technology, process improvement separate from the practice)

        OR

        - There is unexplained attrition > 70% OR the study uses a randomized design and there is differential attrition not controlled by analysis.

        OR

        - Results reported and/or conclusions are not representative of the work that was done (e.g., additional relevant findings are mentioned, but not reported and/or not incorporated in the conclusions)

Quality Dimensions	Maximum Points	Rating Criteria	Deduct 1 Point if:	Deduct 2 Points if:	Score Zero if:
Study	**3** -	Facility Description	The study location is sufficiently distinctive that the results obtained through that setting may not be generalizable to other settings.	The study location is sufficiently distinctive that the results obtained through that setting are unlikely to be generalizable to other settings	It is clear that the setting or situation is unique such that the results cannot generalize to other settings
		Study design/study time period/ patient population	The sample (either subjects or tests) may not be representative of the results of the practice with respect to how the sample was obtained or identified AND the non-representativeness suggests that the results may not be generalizable.	The sample (either subjects or tests) are unlikely to be representative of the results of the practice with respect to how the sample was obtained or identified AND the non-representativeness suggests that the results are unlikely to be generalizable.	The sample is sufficiently unrepresentative based on how it was obtained/identified to clearly nullify the generalizability of the results.
		Potential study bias	The study design, time period and sample selection methods may introduce a study bias that would substantially affect results (i.e., may produce study results interpreted as inconsistent with the true results)	The study design, time period and sample selection methods are likely to introduce a study bias that would substantially affect results (i.e., would likely produce study results interpreted as inconsistent with the true results)	There is reason to believe that the study characteristics can not produce results representative of the practice
Practice	**2** -	Description of practice	The practice and its basic characteristics are not sufficiently identified.	N/A	The practice and its basic characteristics can not be clearly identified.
		Adequacy of practice description	An important aspect/component that is likely to critically affect implementation of the practice is not well described.	N/A	N/A
Outcome Measures	**2**	Face validity	The 'best' measure from a study: - Does not capture well the	N/A	The 'best' study measure: - Is confounded by the practice itself (outcome is a

Quality Dimensions	Maximum Points	Rating Criteria	Deduct 1 Point if:	Deduct 2 Points if:	Score Zero if:
			outcome being estimated **OR** - Estimates an outcome that is only modestly related to health care quality or patient safety (e.g., provider satisfaction, compared with change in an error rate)		direct result of the practice; not available to the comparison) **OR** - Is confounded by the context in which the practice has been implemented (outcome is unlikely to be clearly attributed to the practice). **OR** - Does not have the potential to contribute to health care quality or patient safety
		Recording Method	Method(s) of recording: - Not described **OR** - Does not accurately capture all instances of the outcome	N/A	Method of recording the outcome is unreliable.
Results / Findings	**3**	Sample Sufficiency	The measurement period may be insufficient to allow a robust estimate of the impact of a practice. **OR** Statistical power is not discussed AND the sample may be too small to allow a robust estimate of the impact of a practice	Number of subjects not reported **OR** Measurement period not reported **OR** Measurement period likely insufficient for a robust estimate of the impact of a practice **OR** Statistical power is not discussed AND the sample is likely too small for a robust estimate of the impact of a practice	N/A
		Appropriateness of statistical	Compares two practices and their estimates are based on data collected during notably	Different measures or different recording practices are used when comparing the results of	N/A

Quality Dimensions	Maximum Points	Rating Criteria	Deduct 1 Point if:	Deduct 2 Points if:	Score Zero if:
		analysis	different time periods **OR** Does not provide data sufficient to allow calculation of an effect size	two practices **OR** An inappropriate statistical analysis *and* insufficient data to allow calculation of an effect size	
		Uncontrolled Deviations and Results/conclusion bias	N/A	Results/effect size reported not clearly attributable to practice being evaluated, but instead are likely related to significantly different practice(s) **OR** There is unexplained attrition > 70% or the study uses a randomized design and there is differential attrition not controlled by analysis **OR** Results reported and/or conclusions are not representative of the work that was done (e.g., additional relevant findings are mentioned, but not reported and/or not incorporated in the conclusions)	N/A
Overall Study Quality	**10**				