Quality assessment for individual studies to be used for the review of the resuscitation science for 2010

The literature review process will result in the identification of studies that appear relevant to the question asked. For each of these studies, it is expected that a number of assessments will be made. These include:

- Level of Evidence
- Relevance to the question asked
- Methodological quality
- Outcome(s) assessed
- Magnitude of any observed effect
- Direction of support or otherwise for the (+ve) hypothesis, according to the specific outcomes that have been assessed

This document will describe the process for allocating “methodological quality” to each of the studies identified.

“Methodological quality” (internal validity) of a study can be defined as “the extent to which a study's design, conduct, and analysis has minimized selection, measurement, and confounding biases” ([http://www.ahrq.gov/clinic/epcsums/strengthsum.htm](http://www.ahrq.gov/clinic/epcsums/strengthsum.htm)). That quality is separate to “non-methodological” quality, which refers to the external validity or generalizability of the study results to other (broader) population groups.

There is no uniformly agreed way of defining methodological quality. There are a large number of published methods to assess the quality of studies. These have been summarised in two reviews: one performed by the Agency for Healthcare Research and Quality ([http://www.ahrq.gov/clinic/epcsums/strengthsum.htm](http://www.ahrq.gov/clinic/epcsums/strengthsum.htm)), and the other published by the GRADE working group (Atkins, 2004). None of the individual approaches (largely involving checklists) have been found to be appropriate in all settings, and the use of different approaches may actually end up with conflicting results.

The quality assessment tools that are to be used are listed for each of the three major categories of studies, depending on the type of question being asked: intervention, diagnosis, or prognosis. These are listed below as sections A, B and C:

- Quality assessment for studies assessing interventions (LOEs 1 to 5)
- Quality assessment for studies assessing prognosis (LOEs P1 to P5)
- Quality assessment for studies assessing diagnosis (LOEs D1 to D5)

This approach is based on the principle that within a given level of evidence, higher levels of methodological quality are allocated to studies that minimise the risk of bias, and increase the certainty that the results reported can be interpreted with confidence. Instead of a strict criterion based assessment, we ask the worksheet reviewer to allocate the quality of each study into Good, Fair or Poor:

- Good studies = have most/all of the relevant quality items
• Fair studies = have some of the relevant quality items
• Poor studies = have few of the relevant quality items (but sufficient value to include for further review).

If a study has insufficient relevant quality items to even be classified as “poor” then that study should be excluded from further review.

The allocation of a grade for methodological quality should also be supplemented by the discussion of the relevant quality factors for the individual studies in the “Reviewer’s Final Comments” section.

At the end of this document are references and a glossary.

A separate document refers to how to allocate the appropriate Level of Evidence for a given study (C2010LOEs.doc).
A. Quality assessment for studies assessing interventions

The Levels of Evidence for therapeutic interventions that we are using are shown below:

<table>
<thead>
<tr>
<th>LOE 1: Randomised Controlled Trials (or meta-analyses of RCTs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOE 2: Studies using concurrent controls without true randomisation (eg. “pseudo”-randomised)</td>
</tr>
<tr>
<td>LOE 3: Studies using retrospective controls</td>
</tr>
<tr>
<td>LOE 4: Studies without a control group (eg. case series)</td>
</tr>
<tr>
<td>LOE 5: Studies not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.)</td>
</tr>
</tbody>
</table>

Studies are allocated a rating for methodological quality (Good, Fair or Poor) according to the presence of the quality items for that Level of Evidence:

- Good studies = have most/all of the relevant quality items
- Fair studies = have some of the relevant quality items
- Poor studies = have few of the relevant quality items (but sufficient value to include for further review).

The relevant quality items for studies addressing interventions are listed below for each of these Levels of Evidence:

**LOE 1**

Quality assessment for Randomised Controlled Trials

There are a large number of published methods to assess the quality of RCTs (including Cochrane, GRADE, SIGN, CEBM and (A)NH&MRC). None of these approaches (largely involving checklists) have been found to be appropriate in all settings, and the use of different approaches may end up with apparently conflicting results. A number of independent factors have been reported to have an impact on the outcome of individual studies, these include: true random allocation, concealment of allocation, blinding, and funding or sponsorship ([http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.section.71218](http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1.section.71218)).

The seven factors that we have included as the relevant quality items for RCTs are:

- Was the assignment of patients to treatment randomised?
- Was the randomisation list concealed?
- Were all patients who entered the trial accounted for at its conclusion?
- Were the patients analysed in the groups to which they were randomised?
- Were patients and clinicians "blinded" to which treatment was being received?
- Aside from the experimental treatment, were the groups treated equally?
- Were the groups similar at the start of the trial?
Quality assessment for meta-analyses of RCTs

A meta-analysis of RCTs is also allocated a LOE = 1.

The six factors that we have included as the relevant quality items for meta-analyses are:

- Were specific objectives of the review stated (based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)?
- Was study design defined?
- Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
- Were inclusive searches undertaken (using appropriately crafted search strategies)?
- Were characteristics and methodological quality of each trial identified?
- Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?

As you can see from this list, the worksheets being prepared for C2010 are expected to conform to all of these criteria.

LOE 2

Quality assessment for studies using concurrent controls without true randomisation

The four factors that we have included as the relevant quality items for these studies (both experimental and observational) are:

- Were comparison groups clearly defined?
- Were outcomes measured in the same (preferably blinded), objective way in both groups?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete?

For these studies it would be reasonable to consider the presence of all 4 factors = Good, only 3 factors = Fair, and only 2 factors = Poor. A study with only one would be considered of insufficient quality to include in the next step of the review.

Quality assessment for meta-analyses of studies using concurrent controls without true randomisation

A meta-analysis of these types of studies is also allocated a LOE = 2.

The six factors that we have included as the relevant quality items for meta-analyses are:

- Were specific objectives of the review stated (based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)?
- Was study design defined?
- Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
- Were inclusive searches undertaken (using appropriately crafted search strategies)?
- Were characteristics and methodological quality of each trial identified?
- Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?

**LOE 3**

**Quality assessment for studies using retrospective controls:**
The four factors that we have included as the relevant quality items for these studies are:
- Were comparison groups clearly defined?
- Were outcomes measured in the same (preferably blinded), objective way in both groups?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete?
For these studies it would be reasonable to consider the presence of all 4 factors = Good, only 3 factors = Fair, and only 2 factors = Poor. A study with only one would be considered of insufficient quality to include in the next step of the review.

**LOE 4**

**Quality assessment for case series**
The three factors that we have included as the relevant quality items for these studies (both experimental and observational) are:
- Were outcomes measured in an objective way?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete?
For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

**LOE 5**

**Quality assessment for studies that are not directly related to the specific patient/population**
LOE 5 studies are those not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.), and should have their methodological quality allocated to the methodology of the study. The suggested relevant quality criteria here are:
- Good = randomised controlled trials (equivalent of LOE 1)
- Fair = studies without randomised controls (equivalent of LOE 2-3)
- Poor = studies without controls (equivalent of LOE 4)
This would mean that a randomised controlled trial performed in a related population (eg. stroke patients or animals), would be categorised as Good quality LOE 5 study.
B. Quality assessment for studies assessing prognosis

The Levels of Evidence that we are using for studies assessing prognosis are shown below:

<table>
<thead>
<tr>
<th>C2010 Levels of Evidence for Prognostic Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOE P1: Inception (prospective) cohort studies (or meta-analyses of inception cohort studies), or validation of Clinical Decision Rule (CDR)</td>
</tr>
<tr>
<td>LOE P2: Follow up of untreated control groups in RCTs (or meta-analyses of follow-up studies), or derivation of CDR, or validated on split-sample only</td>
</tr>
<tr>
<td>LOE P3: Retrospective cohort studies</td>
</tr>
<tr>
<td>LOE P4: Case series</td>
</tr>
<tr>
<td>LOE P5: Studies not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.)</td>
</tr>
</tbody>
</table>

Studies are allocated a rating for methodological quality (Good, Fair or Poor) according to the presence of the relevant quality items for that Level of Evidence:

- Good studies = have most/all of the relevant quality items
- Fair studies = have some of the relevant quality items
- Poor studies = have few of the relevant quality items (but sufficient value to include for further review).

The quality items for studies addressing interventions are listed below for each of these Levels of Evidence:

### LOE P1

**Quality assessment for inception/prospective cohort studies, or for studies validating a Clinical Decision Rule**

The four factors that we have included as the relevant quality items for studies of LOE P1, P2, and P3 are:

- Were comparison groups clearly defined?
- Were outcomes measured in the same (preferably blinded), objective way in both groups?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete (eg. >80%)?

For these studies it would be reasonable to consider the presence of all 4 factors = Good, only 3 factors = Fair, and only 2 factors = Poor. A study with only one would be considered of insufficient quality to include in the next step of the review.

**Quality assessment for meta-analyses of inception/prospective cohort studies**

A meta-analysis of these types of studies is also allocated a LOE = P1.

The six factors that we have included as relevant quality items for meta-analyses are:
• Were specific objectives of the review stated (based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)
• Was study design defined?
• Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
• Were inclusive searches undertaken (using appropriately crafted search strategies)?
• Were characteristics and methodological quality of each trial identified?
• Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?

LOE P2

Quality assessment for studies involving follow up of untreated control groups in RCTs, studies deriving a Clinical Decision Rule or studies validating a Clinical Decision Rule using a split sample

The four factors that we have included as the relevant quality items for studies of LOE P1, P2, and P3 are:
• Were comparison groups clearly defined?
• Were outcomes measured in the same (preferably blinded), objective way in both groups?
• Were known confounders identified and appropriately controlled for?
• Was follow-up of patients sufficiently long and complete (eg. >80%)?

For these studies it would be reasonable to consider the presence of all 4 factors = Good, only 3 factors = Fair, and only 2 factors = Poor. A study with only one would be considered of insufficient quality to include in the next step of the review.

Quality assessment for meta-analyses of follow-up studies

A meta-analysis of these types of studies is also allocated a LOE = P2.

The six factors that we have included as the relevant quality items for meta-analyses are:
• Were specific objectives of the review stated (based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)
• Was study design defined?
• Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
• Were inclusive searches undertaken (using appropriately crafted search strategies)?
• Were characteristics and methodological quality of each trial identified?
• Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?
LOE P3

Quality assessment for retrospective cohort studies
The four factors that we have included as the relevant quality items for studies of LOE P1, P2, and P3 are:

- Were comparison groups clearly defined?
- Were outcomes measured in the same (preferably blinded), objective way in both groups?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete (eg. >80%)?

For these studies it would be reasonable to consider the presence of all 4 factors = Good, only 3 factors = Fair, and only 2 factors = Poor. A study with only one would be considered of insufficient quality to include in the next step of the review.

LOE P4

Quality assessment of case series
The three factors that we have included as the relevant quality items for case series are:

- Were outcomes measured in an objective way?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete (eg. >80%)?

For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

LOE P5

Quality assessment for prognostic studies that are not directly related to the specific patient/population
LOE P5 studies are those prognostic studies that are not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.), and should have their methodological quality allocated to the methodology of the individual study. The suggested relevant quality criteria here are:

- Good = Inception/prospective cohort studies or studies involving validation of a Clinical Decision Rule (CDR) (equivalent of LOE P1)
- Fair = CDR derived from a population, or validated using a split-sample, or retrospective cohort studies (equivalent of LOE P2-3)
- Poor = studies without controls (eg. case series) (equivalent of LOE P4)

This would mean that an inception cohort study performed in a related population (eg. stroke patients or animals), would be categorised as Good quality LOE P5 study.
C. Quality assessment for studies assessing diagnosis

The Levels of Evidence that we are using for studies assessing diagnosis are shown below:

<table>
<thead>
<tr>
<th>C2010 Levels of Evidence for Diagnostic Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOE D1: Validating cohort studies (or meta-analyses of validating cohort studies), or validation of Clinical Decision Rule (CDR)</td>
</tr>
<tr>
<td>LOE D2: Exploratory cohort study (or meta-analyses of follow-up studies), or derivation of CDR, or a CDR validated on a split-sample only</td>
</tr>
<tr>
<td>LOE D3: Diagnostic case control study</td>
</tr>
<tr>
<td>LOE D4: Study of diagnostic yield (no reference standard)</td>
</tr>
<tr>
<td>LOE D5: Studies not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.)</td>
</tr>
</tbody>
</table>

Studies are allocated a rating for methodological quality (Good, Fair or Poor) according to the presence of the relevant quality items for that Level of Evidence:

- Good studies = have most/all of the relevant quality items
- Fair studies = have some of the relevant quality items
- Poor studies = have few of the relevant quality items (but sufficient value to include for further review).

The relevant quality items for studies addressing interventions are listed below for each of these Levels of Evidence:

**LOE D1**

**Quality assessment of validating cohort (prospective, observational) studies, or studies involving validation of a Clinical Decision Rule**

The three factors that we have included as the relevant quality items for studies of LOE D1, D2, and D3 are:

- Was the diagnostic test evaluated in an appropriate spectrum of patients (eg. in those in whom it would be used in practice)? (Minimising “spectrum bias”)?
- Was there an independent, blind comparison with a reference ("gold") standard of diagnosis? (Minimising “review bias”)
- Was the reference standard applied regardless of the test result? (Minimising “verification bias”)

For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

**Quality assessment for meta-analyses of validating cohort studies**

A meta-analysis of these types of studies is also allocated a LOE = D1.

The six factors that we have included as the relevant quality items for meta-analyses are:
• Were specific objectives of the review stated? (Based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)
• Was study design defined?
• Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
• Were inclusive searches undertaken (using appropriately crafted search strategies)?
• Were characteristics and methodological quality of each trial identified?
• Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?

LOE D2

Quality assessment for exploratory cohort studies, studies deriving a Clinical Decision Rule or studies validating a Clinical Decision Rule using a split sample
The three factors that we have included as the relevant quality items for studies of LOE D1, D2, and D3 are:
• Was the diagnostic test evaluated in an appropriate spectrum of patients (eg, in those in whom it would be used in practice)? (Minimising “spectrum bias”)?
• Was there an independent, blind comparison with a reference (“gold”) standard of diagnosis? (Minimising “review bias”)
• Was the reference standard applied regardless of the test result? (Minimising “verification bias”)
For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

Quality assessment for meta-analyses of follow-up studies
A meta-analysis of these types of studies is also allocated a LOE = D2.
The six factors that we have included as the relevant quality items for meta-analyses are:
• Were specific objectives of the review stated? (Based on a specific clinical question in which patient, intervention, comparator, outcome (PICO) were specified)
• Was study design defined?
• Were selection criteria stated for studies to be included (based on trial design and methodological quality)?
• Were inclusive searches undertaken (using appropriately crafted search strategies)?
• Were characteristics and methodological quality of each trial identified?
• Were selection criteria applied and a log of excluded studies with reasons for exclusion reported?
LOE D3

Quality assessment for diagnostic case-control studies
The three factors that we have included as the relevant quality items for studies of LOE D1, D2, and D3 are:

- Was the diagnostic test evaluated in an appropriate spectrum of patients (eg. in those in whom it would be used in practice)? (Minimising “spectrum bias”)?
- Was there an independent, blind comparison with a reference (“gold”) standard of diagnosis? (Minimising “review bias”)
- Was the reference standard applied regardless of the test result? (Minimising “verification bias”)

For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

LOE D4

Quality assessment of studies of diagnostic yield
The three factors that we have included as the relevant quality items for studies of diagnostic yield are:

- Were outcomes measured in an objective way?
- Were known confounders identified and appropriately controlled for?
- Was follow-up of patients sufficiently long and complete?

For these studies it would be reasonable to consider the presence of all 3 factors = Good, only 2 factors = Fair, and only 1 factor = either Poor or of insufficient quality to include in the next step of the review.

LOE D5

Quality assessment for diagnostic studies that are not directly related to the specific patient/population
LOE D5 studies are those diagnostic studies that are not directly related to the specific patient/population (eg. different patient/population, animal models, mechanical models etc.), and should have their methodological quality allocated to the methodology of the individual study. The suggested relevant quality criteria here are:

- Good = Validating cohort studies or studies involving validation of a Clinical Decision Rule (CDR) (equivalent of LOE D1)
- Fair = Exploratory cohort studies, CDR derived from a population, or validated using a split-sample, or diagnostic case-control studies (equivalent of LOE D2-3)
- Poor = studies without controls (eg. studies of diagnostic yield) (equivalent of LOE D4)

This would mean that a validating cohort study performed in a related population (eg. stroke patients or animals), would be categorised as Good quality LOE D5 study.
**Summary**

The assessment of the methodological quality of individual studies is an essential part of any systematic review. The specific criteria to be used for the C2010 worksheets are listed above according to the type of study. These criteria allow the reviewer to allocate a grade of methodological quality (good, fair or poor) to each included study.
References


http://www.ahrq.gov/clinic/epcsums/strengthsum.htm (Systems to Rate the Strength of Scientific Evidence; Agency for Healthcare Research and Quality (AHRQ); Evidence Report/Technology Assessment: Number 47; 2002)

http://www.cebm.net/?o=1021 (Rapid Critical Appraisal of Diagnostic Accuracy Studies (637Kb) Paul Glasziou)

http://www.cebm.net/index.aspx?q=1025 (The Levels of evidence (RTF document) sets out one approach to systematising this process for different question types.)

http://www.cebm.utoronto.ca/glossary/


Glossary

Case control study:
A case control study involves identifying patients who have the outcome of interest (cases) and patients without the same outcome (controls), and looking back to see if they had the exposure of interest (http://www.cebm.utoronto.ca/glossary/).

Diagnostic case-control study:
The index test results for a group of patients already known to have the disease (through the reference standard) are compared to the index test results with a separate group of normal/healthy people known to be free of the disease (through the use of the reference standard). In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are excluded, which can lead to exaggeration of both sensitivity and specificity. This is called spectrum bias because the spectrum of study participants will not be representative of patients seen in practice. (Note: this does not apply to well-designed population based case-control studies.)

Case series:
A single group of people exposed to the intervention (factor under study). Only outcomes after the intervention (factor under study) are recorded in the series of people, so no comparisons can be made.

Clinical Decision Rule
These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category. These can be derived, validated using a split-sample only (derived from part of population, and validated on rest of population), or validated using a separate population (single or multiple).

Cohort study
Outcomes for groups of people observed to be exposed to an intervention, or the factor under study, are compared to outcomes for groups of people not exposed.

Inception/prospective cohort studies
At study inception the cohort is either non-diseased or all at the same stage of the disease or where groups of people (cohorts) are observed at a point in time to be exposed or not exposed to an intervention (or the factor under study) and then are followed prospectively with further outcomes recorded as they happen.

Retrospective cohort studies
Where the cohorts (groups of people exposed and not exposed) are defined at a point of time in the past and information collected on subsequent outcomes (eg. the use of medical records to identify a group of women using oral contraceptives five years ago,
and a group of women not using oral contraceptives, and then contacting these women or identifying in subsequent medical records the development of deep vein thrombosis).

**Validating cohort (prospective, observational) studies:**
Test the quality of a specific diagnostic test, based on prior evidence.

**Exploratory study:**
Collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.

**Methodological quality**
“Methodological quality” (internal validity) of a study can be defined as “the extent to which a study's design, conduct, and analysis has minimized selection, measurement, and confounding biases” ([http://www.ahrq.gov/clinic/epcsums/strengthsum.htm](http://www.ahrq.gov/clinic/epcsums/strengthsum.htm)).

**Non-methodological quality**
“Non-methodological” quality refers to the external validity or generalizability of the study results to other (broader) population groups.

**Randomised Controlled Trials:**
These studies prospectively collect data, and randomly allocate the patients to intervention or control groups.

**Studies using concurrent controls without true randomisation:**
These studies can be:
- experimental - having patients that are allocated to intervention or control groups concurrently, but in a non-random fashion (including pseudo-randomisation: eg. alternate days, day of week etc), or
- observational – including cohort and case control studies

**Studies using retrospective controls:**
These studies use control patients that have been selected from a previous period in time to the intervention group.

**Study of diagnostic yield:**
These studies provide the yield of diagnosed patients, as determined by the index test, without confirmation of the accuracy of the diagnosis (ie. whether the patient is actually diseased) by a reference standard test (index test). These may be the only alternative when there is no reliable reference standard.