Levels of Evidence used for the review of the resuscitation science for 2010

The Levels of Evidence to be used by the worksheet authors for the review of resuscitation science have been updated based on a review of the available evidence.

The Levels that are to be used are divided into three major categories, depending on the type of question being asked: intervention, diagnosis, or prognosis. These are listed below as sections A, B and C:

- Levels of Evidence for studies assessing interventions (LOEs 1 to 5)
- Levels of Evidence for studies assessing prognosis (LOEs P1 to P5)
- Levels of Evidence for studies assessing diagnosis (LOEs D1 to D5)

These are all based on the principle that higher levels of evidence are allocated to studies that minimise the risk of bias, and all offer the opportunity to include studies that are not directly related to the question being asked (allowing extrapolation of information from different populations, animal studies etc).

At the end of this document are references, a glossary, and two appendices:

- Appendix 1: Comparison with other interventional LOEs
- Appendix 2: Algorithm for classifying study design for questions of effectiveness (National Institute for Health and Clinical Excellence)

A separate document refers to how quality should be assessed for each of the Levels of Evidence for each of these categories of evidence.
A. Levels of Evidence for studies assessing interventions

There are a number of published, and relatively widely used Levels of Evidence for therapeutic interventions (see Appendix 1).

The Levels of Evidence for therapeutic interventions that we are using are easy to apply and are shown below:

<table>
<thead>
<tr>
<th>C2010 Levels of Evidence for Studies of Therapeutic Interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOE 1: Randomised Controlled Trials (or meta-analyses of RCTs)</td>
</tr>
<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>

Further explanatory information is listed below for each of these levels (see also Appendix 2):

**LOE 1**

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

**LOE 2**

**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

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

**LOE 3**

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

**LOE 4**

**Case series:**
A single group of people exposed to the intervention (factor under study), but without a control group.
LOE 5

As with other categories of Levels of Evidence, we have used LOE 5 to refer to studies that are not directly related to the specific patient/population. These could be different patients/population, or animal models, and could include high quality studies (including RCTs).

Example using levels of evidence for intervention studies

If we start with the following hypothetical question (in the PICO format):

*In cardiac arrest patients (P), does the use of the drug “Bettermycin” (I) when compared with usual care (C), improve intact neurological survival (O)?*

An example of the types of supportive studies that may be identified for this question are:

**LOE 5:** In 200 patients with thrombo-embolic stroke, neurologically intact survival at 1 month was higher in the 100 patients who were randomised to receive “Bettermycin”, than in the 100 patients who were randomised to placebo.

**LOE 4:** In 25 cases of Out-Of-Hospital Cardiac Arrest due to VF who received “Bettermycin”, intact neurological survival at 1 month was observed in 50% of patients.

**LOE 3:** Neurologically intact survival was observed more frequently in 50 cases of Out-Of-Hospital Cardiac Arrest due to VF who received “Bettermycin” than in the 75 cases of Out-Of-Hospital Cardiac Arrest due to VF in the previous 2 years (who did not receive “Bettermycin”).

**LOE 2:** Neurologically intact survival was observed more frequently in the 100 cases of Out-Of-Hospital Cardiac Arrest due to VF who received “Bettermycin” in one county, than in the 115 cases of Out-Of-Hospital Cardiac Arrest due to VF in the neighbouring county (who did not receive “Bettermycin”).

**LOE 1:** In the 200 cases of Out-Of-Hospital Cardiac Arrest due to VF, neurologically intact survival was observed more frequently in the 100 cases that were randomly allocated to receive “Bettermycin”, than in the 100 cases that were randomised to placebo.
**B. Levels of Evidence for studies assessing prognosis**

Prognostic questions share three elements:
- a qualitative aspect (which outcomes could happen?)
- a quantitative aspect (how likely are they to happen?), and
- a temporal aspect (over what time period?)

Levels of Evidence 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>

Further explanatory information is listed below for each of these levels:

**LOE P1**

**Inception/prospective cohort studies**
In a 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. For a cohort study to be considered LOE P1, it should be an inception/prospective cohort study. In these 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.

**Validation of a Clinical Decision Rule**
Clinical Decision Rules are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category. For a study to be considered LOE P1, the CDR should be validated using a completely separate population (single or multiple) to that in which it was derived.

**LOE P2**

**Clinical Decision Rule**
Clinical Decision Rules are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category. For a study to be considered LOE P2, the CDR can be either derived from a population, or validated using a split-sample (derived from part of population, and validated on rest of population).
LOE P3

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).

LOE P4

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.

LOE P5

As with other categories of Levels of Evidence, we have used LOE P5 to refer to studies that are not directly related to the specific patient/population. These could be different patients/population, or animal models, and could include high quality studies (including inception cohort studies or validation of a CDR).

Example using prognostic levels of evidence
If we start with the following hypothetical question (in the PICO format):

In cardiac arrest patients (P), does the use of the PETER algorithm (I) allow the accurate prediction of a intact neurological survival (O)?
(PETER: Pain Eyes Temperature ‘Emodynamics Relatives)

An example of the types of studies that may be identified for the this question are:

LOE P5: In pig model of induced VF, the use of the PETER algorithm on day 3 successfully predicted intact neurological survival at 1 month
LOE P4: In 25 cases of Out-Of-Hospital Cardiac Arrest due to VF with a score of >10 on the PETER algorithm on day 3, intact neurological survival at 1 month was observed in 50% of cases.
LOE P3: Retrospective identification of 20 cases after Out-Of-Hospital Cardiac Arrest due to VF where the PETER algorithm score was >10, and 20 cases where PETER algorithm score was <3. More cases in the score >10 group survived neurologically intact.
LOE P2: The PETER algorithm was used in 100 consecutive patients after Out-Of-Hospital Cardiac Arrest due to VF, and a score of >10 was associated with >90% survival neurologically intact at 3 months, and a score of <3 was associated with no such survivors.
LOE P1: The previously proposed cut off of >10 and <3 for the PETER algorithm was tested in 1000 consecutive patients after Out-Of-Hospital Cardiac Arrest due to VF in 5
centres (throughout the world), and a score of >10 was confirmed to be associated with >95% survival neurologically intact at 3 months, and a score of <3 was associated with no such survivors.

Ideally, of course, the best evidence to support the prognostic algorithm, would be an intervention study: In a multi-centre Prospective Randomised Controlled Trial of the use of the PETER algorithm in patients supported in an ICU for up to 30 days before withdrawal, the use of the algorithm would have saved 9.5 ICU days/patient who eventually died, and saved over $US1million.
C. Levels of Evidence for studies assessing diagnosis

Studies that evaluate a diagnostic strategy, use a “test” (examination finding/investigation), and look at the “result”.

Levels of Evidence 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 followup 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>

Further explanatory information is listed below for each of these levels:

**LOE D1**

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

**Validation of a Clinical Decision Rule:**
These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category. For a study to be considered LOE D1, the CDR should be validated using a completely separate population (single or multiple) to that in which it was derived.

**LOE D2**

**An exploratory study:**
Collects information and trawls the data (e.g. using a regression analysis) to find which factors are ‘significant’.
A meta-analysis of such follow-up studies would also be considered LOE D2.

**Clinical Decision Rule:**
These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category. For a study to be considered LOE D2, the CDR can be either derived from a population, or validated using a split-sample (derived from part of population, and validated on rest of population).

**LOE D3**

**Diagnostic 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. For a case control study to be considered as LOE D3, it must
be a diagnostic case control study. In these studies, 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.

**LOE D4**

**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.

**LOE D5**
As with other categories of Levels of Evidence, we have used LOE D5 to refer to studies that are not directly related to the specific patient/population. These could be different patients/population, or animal models, and could include high quality studies (including validating cohort studies or validation of a CDR).

**Example using diagnostic levels of evidence**

If we start with the following hypothetical question (in the PICO format):

*In cardiac arrest patients due to VF (P), does the use of VF waveforms (I) allow the diagnosis of a successfully defibrillatable rhythm (eg. shock will result in Return Of Spontaneous Circulation)?*

An example of the types of studies that may be identified for the this question are:

**LOE D5:** “VF waveform analysis” diagnoses successfully shockable rhythm in pigs with electrically induced VF

**LOE D4:** “VF waveform analysis” was able to categorise 70% of patients with Out-Of-Hospital Cardiac Arrest due to VF as “successfully shockable”, but no outcome data was available

**LOE D3:** In a collection of patients with Out-Of-Hospital Cardiac Arrest due to VF, those with ROSC after a shock had a distinctly difference appearance of VF waveform when compared with patients without ROSC after a shock.

**LOE D2:** In a group of non-consecutive patients with Out-Of-Hospital Cardiac Arrest due to VF, a specific cut off point could be determined that predicted increased likelihood (+ve Likelihood Ratio [+LR] =12) of ROSC after shock. This was determined in 50% of patients and validated in the other 50%.  

**LOE D1:** In a group of consecutive patients with VF from multiple settings, a previously determined Clinical Decision Rule was confirmed to predict increased likelihood (+LR = 12) of ROSC after shock.
Ideally, of course, the best evidence to support the diagnostic test, would be an intervention study: In a Prospective Randomised Controlled Trial, the group of patients in VF that had the Clinical Decision Rule applied, were more likely to achieve ROSC, and hospital discharge, than those who did not have the clinical decision rule applied.
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)

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

http://www.cebm.net/index.aspx?o=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/


http://www.sign.ac.uk/

**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'.

**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.
# Appendix 1: Comparison with other interventional LOEs

The following table compares the C2010 approach, with the other major published Levels of Evidence for the rating of interventional studies:

<table>
<thead>
<tr>
<th>Study type/Approach</th>
<th>C2010</th>
<th>C2005</th>
<th>Grade</th>
<th>SIGN</th>
<th>NH&amp;MRC</th>
<th>OCEBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-analyses</td>
<td>1 or 2</td>
<td>1</td>
<td>n/a</td>
<td>1++ to 2++</td>
<td>I</td>
<td>1a or 2a</td>
</tr>
<tr>
<td>RCTs</td>
<td>1</td>
<td>1 or 2</td>
<td>High/Mod/Low</td>
<td>1/1+/1++</td>
<td>II</td>
<td>1b or 2b</td>
</tr>
<tr>
<td>Concurrent controls</td>
<td>2</td>
<td>3</td>
<td>High/Mod/Low</td>
<td>2- to 2++</td>
<td>III-1 or III-2</td>
<td>2b to 4</td>
</tr>
<tr>
<td>Retrospective controls</td>
<td>3</td>
<td>4</td>
<td>Low</td>
<td>n/a</td>
<td>III-3</td>
<td>?1c!</td>
</tr>
<tr>
<td>No controls</td>
<td>4</td>
<td>5</td>
<td>Very low</td>
<td>3</td>
<td>IV</td>
<td>4</td>
</tr>
<tr>
<td>Animal/Mechanical/Model</td>
<td>5</td>
<td>6</td>
<td>? As for extrap</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Extrapolations</td>
<td>5</td>
<td>7</td>
<td>Downgraded</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Covers &quot;non-therapy&quot; studies</td>
<td>YES</td>
<td>?1to7</td>
<td>Irrelevant!</td>
<td>No</td>
<td>Detailed</td>
<td>Detailed</td>
</tr>
</tbody>
</table>
Appendix 2: Algorithm for classifying study design for questions of effectiveness (National Institute for Health and Clinical Excellence)