CEBM
CENTRE FOR EVIDENCE BASED MEDICINE

UNIVERSITY OF OXFORD

Levels of Evidence (March 2009)

www.cebm.net

Level 1A	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	1a SR (with homogeneity*)of RCTs SR (withhomogeneity*) of inception cohort studies; CDR† validated in different populations SR (with homogeneity*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres SR (with homogeneity*) of prospective cohort studies SR (with homogeneity*) of Level 1 economic studies
Level 1b	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	Individual RCT (with narrow Confidence Interval‡) Individual inception cohort study with > 80% follow-up; CDR† validated in asingle population Validating** cohort study with good††† reference standards; or CDR† tested within one clinical centre Prospective cohort study with good follow-up**** Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
Level 1c	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	All or none§ All or none case series Absolute SpPins and SnNouts†† All or none case-series Absolute better-value or worse-value analyses ††††
Level 2a	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	SR (with homogeneity*) of cohort studies SR (withhomogeneity*) of either retrospective cohort studies or untreated control groups in RCTs SR (with homogeneity*) of Level >2 diagnostic studies SR (with homogeneity*) of 2b and better studies SR (withhomogeneity*) of Level >2 economic studies
Level 2b	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	Individual cohort study (including low quality RCT; e.g., <80% followup) Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split sample §§§ only Exploratory** cohort study with good††† reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases Retrospective cohort study, or poor follow-up Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
Level 2c	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	"Outcomes" Research; Ecological studies "Outcomes" Research Ecological studies Audit or outcomes research
Level 3a	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	SR (with homogeneity*) of case-control studies SR (with homogeneity*) of 3b and better studies SR (with homogeneity*) of 3b and better studies SR (with homogeneity*) of 3b And better studies
Level 3b	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	Individual Case-Control Study Non-consecutive study; or without consistently applied reference standards Non-consecutive cohort study, or very limited population Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses IIncorporatingclinically sensible variations.
Level 4	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	Case-series (and poor quality cohort and casecontrol studies§§) Case-series (and poor quality prognostic cohort studies***) Case-control study, poor or nonindependent reference standard Case-series or superseded reference standards Analysis with no sensitivity analysis
Level 5	**Therapy/Prevention, Aetiology/Harm** **Prognosis** **Diagnosis** **Differential diag/symptom prevalence** **Economic and decision analyses**	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009)
(for definitions of terms used see glossary at http://www.cebm.net/?o=1116)

Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998. Updated by Jeremy Howick March 2009.

CEBM
CENTRE FOR EVIDENCE BASED MEDICINE

UNIVERSITY OF OXFORD

Levels of Evidence (March 2009)

www.cebm.net

NOTES

Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because:

EITHER a single result with a wide Confidence Interval

OR a Systematic Review with troublesome heterogeneity.

Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

*	By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level.
†	Clinical Decision Rule. (These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category.)
‡	See note above for advice on how to understand, rate and use trials or other studies with wide confidence intervals.
§	Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it.
§§	By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and nonexposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders.
§§§	Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples.
††	An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules out the diagnosis.
‡‡	Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits.
†††	Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study.
††††	Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive.
**	Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'.
***	By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors.
****	Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (for example 1-6 months acute, 1 - 5 years chronic)

Grades of Recommendation

A	consistent level 1 studies
B	consistent level 2 or 3 studies or extrapolations from level 1 studies
C	level 4 studies or extrapolations from level 2 or 3 studies
D	level 5 evidence or troublingly inconsistent or inconclusive studies of any level

"Extrapolations" are where data is used in a situation that has potentially clinically important differences than the original study situation.

Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009)
(for definitions of terms used see glossary at http://www.cebm.net/?o=1116)

Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998. Updated by Jeremy Howick March 2009.