www.nhmrc.gov.au

Australian Government
National Health and Medical Research Council

INVESTING IN AUSTRALIA'S HEALTH

RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

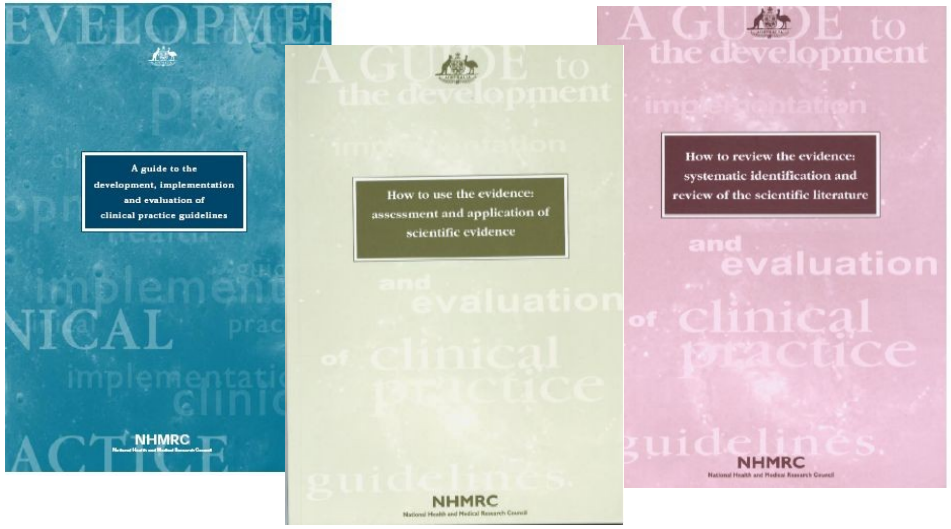
RE-ASSESSING AND REVISING 'LEVELS OF EVIDENCE' IN THE CRITICAL APPRAISAL PROCESS

NHMRC Guideline Assessment Register (GAR) working party -

Kristina Coleman, Karen Grimmer, Susan Hillier, Tracy Merlin, Philippa Middleton, Sarah Norris, Janet Salisbury, Rebecca Tooher, Adele Weston

National Health and Medical Research Council (NHMRC)

- Develops health publications
  - Health advisories
  - Evidence-based clinical practice guidelines
  - Methodology publications



NHMRC levels of evidence

Level of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomised controlled trials
II	Evidence obtained from at least one properly-designed randomised controlled trial
III-1	Evidence obtained from well-designed pseudorandomised controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test/post-test

Source: NHMRC (1999) 'A guide to the development, implementation and evaluation of clinical practice guidelines'

Flaws of 'levels'. I.

- Limited applicability for certain research questions
  - Effectiveness ✓
  - Diagnostic accuracy ✗ ✗
  - Prognosis ✗
  - Aetiology ✗
- Framed in terms of a body of evidence, but used to describe individual studies

Flaws of 'levels'. II.

- Includes 'quality' judgements e.g. 'well designed', 'properly designed' → ignored in practice.
  - Critical appraisal done separately and level of evidence applied to study design alone

- Does not include systematic reviews of all types of study designs

Objectives of Working Party

- Revise levels of evidence
  - Create a framework that aligns as closely as possible with the current levels of evidence – to minimise confusion for current users/ interpreters - but which also addresses other research questions appropriately
- Develop grading system for body of evidence

'Levels' working party

- The Development Process
  - Identify a suitable framework upon which to model the revised levels → CEBM
  - Maintain levels I-IV
  - Ensure it reflects individual studies rather than body of evidence
  - Ensure consistency in hierarchy across all research questions
  - Utilise empirical evidence supporting relationships between study design and bias wherever possible

'Levels' working party

- Outcomes
  - 'Levels of evidence' framework
  - Explanatory notes
  - Glossary of study designs/terminology

Table 1. Designation of levels of evidence according to type of research question

Level	Intervention §	Diagnosis **	Prognosis	Aetiology †††	Screening
I *	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,§§ among consecutive patients with a defined clinical presentation ††	A prospective cohort study ***	A prospective cohort study	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,§§ among non-consecutive patients with a defined clinical presentation††	All or none §§§	All or none §§§	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparative study with concurrent controls: · Non-randomised, experimental trial † · Cohort study · Case-control study · Interrupted time series with a control group	A comparison with reference standard that does not meet the criteria required for Level II and Level III-1 evidence	Analysis of prognostic factors amongst untreated control patients in a randomised controlled trial	A retrospective cohort study	A comparative study with concurrent controls: · Non-randomised, experimental trial · Cohort study · Case-control study
III-3	A comparative study without concurrent controls: · Historical control study · Two or more single arm study ‡ · Interrupted time series without a parallel control group	Diagnostic case-control study††	A retrospective cohort study	A case-control study	A comparative study without concurrent controls: · Historical control study · Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard) ‡‡	Case series, or cohort study of patients at diff-erent stages of disease	A cross-sectional study	Case series

Diagnostic studies

Level of evidence	Study design
I	A systematic review of level II studies
II	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,§§ among consecutive patients with a defined clinical presentation ††
III-1	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,§§ among non-consecutive patients with a defined clinical presentation††
III-2	A comparison with reference standard that does not meet the criteria required for Level II and Level III-1 evidence
III-3	Diagnostic case-control study††
IV	Study of diagnostic yield (no reference standard) ‡‡

Notes for diagnostic studies

§§ The validity of the reference standard should be determined in the context of the disease under review. Criteria for determining the validity of the reference standard should be pre-specified. This can include the choice of the reference standard(s) and its timing in relation to the index test. The validity of the reference standard can be determined through quality appraisal of the study. See Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology*, 2003, 3: 25.

†† Well-designed population based case-control studies (eg population based screening studies where test accuracy is assessed on all cases, with a random sample of controls) do capture a population with a representative spectrum of disease and thus fulfil the requirements for a valid assembly of patients. However, in some cases the population assembled is not representative of the use of the test in practice. In diagnostic case-control studies a selected sample of patients already known to have the disease are compared with a separate group of normal/healthy people known to be free of the disease. 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.

‡‡ Studies of diagnostic yield provide the yield of diagnosed patients, as determined by an index test, without confirmation of the accuracy of this diagnosis by a reference standard. These may be the only alternative when there is no reliable reference standard.

Additional Important Tablenotes

* A systematic review will only be assigned a level of evidence as high as the studies it contains, excepting where those studies are of level II evidence.

††† If it is possible and/or ethical to determine a causal relationship using experimental evidence, then the "Intervention" hierarchy of evidence should be utilised. If it is only possible and/or ethical to determine a causal relationship using observational evidence (i.e. cannot allocate groups to a potential harmful exposure, such as nuclear radiation), then the "Aetiology" hierarchy of evidence should be utilised.

Note 1: Assessment of comparative harms/safety should occur according to the hierarchy presented for each of the research questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms are rare and cannot feasibly be captured within randomised controlled trials; physical harms and psychological harms may need to be addressed by different study designs; harms from diagnostic testing include the likelihood of false positive and false negative results, harms from screening include the likelihood of false alarm and false reassurance results.

Note 2: When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding research question e.g. level II intervention evidence; level IV diagnostic evidence; level III-2 prognostic evidence etc.

The consultation process

- Pilot process/trial of framework: Nov 2004 – June 2006
- Posting on website (www.nhmrc.gov.au) under 'Consultation' section → feedback form
- Disseminate information on "transitional" framework nationally and internationally
  - HTAi 2005, Rome
  - Cochrane Colloquium 2005, Melbourne

Up to here!!

- Publish in general medical journal ± methodology journal
- Report to Health Advisory Committee
- NHMRC endorsement/finalisation

Acknowledgements

- **'Levels and grades' working party**
  - Kristina Coleman, Sarah Norris, Adele Weston - Health Technology Analysts Pty Ltd
  - Karen Grimmer, Susan Hillier - Division of Health Sciences, University of South Australia
  - Tracy Merlin - Adelaide Health Technology Assessment (AHTA), Department of Public Health, University of Adelaide
  - Philippa Middleton, Rebecca Tooher - ASERNIP-S
  - Janet Salisbury – Biotext

- **Feedback** on levels and grades has been provided during the development phase from the following:
  - Paul Glasziou – Oxford University, United Kingdom
  - Brian Haynes – McMaster University, Canada
  - Andrew Oxman – Oslo, Norway (GRADE Working Group)
  - Nicki Jackson –Deakin University
  - Sally Lord and Les Irwig – University of Sydney

- **NHMRC Health Advisory Section**
  - Janine Keough
  - Chris Gonzalez