

EVIDENCE AND RECOMMENDATION GRADING IN GUIDELINES

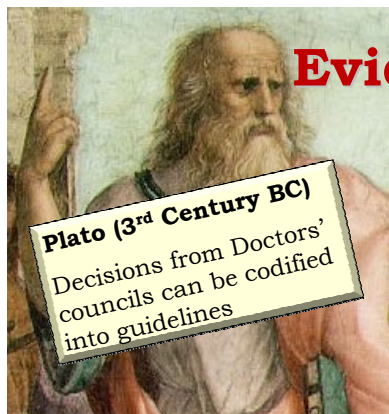


A short history

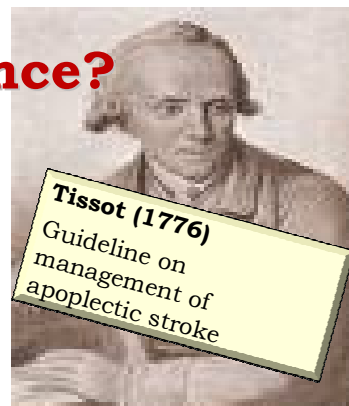
Françoise Cluzeau
Senior Advisor
NICE International

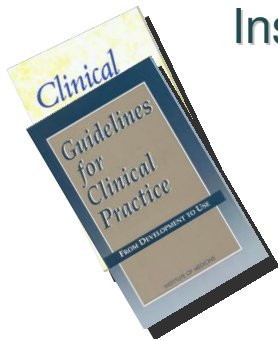
**G-I-N, Lisbon
2 November 2009**

Long ago.... Before grading



Evidence?



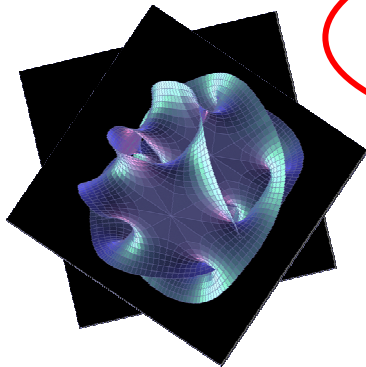


Institute of Medicine, 1990

"Guidelines are valid if, when followed, they lead to the predicted outcomes in terms of health gain and costs. The prospective assessment of validity will examine:

- the type of evidence used
- the methods for evaluating the evidence
- the consistency between the evidence & recommendations
- The strength of the evidence and expert judgement used in the process."

Validity: two dimensions



- **Internal validity** – the degree to which the results of a study are likely to approximate to the 'truth'; lack of bias
- **External validity** – the extent to which the effects observed in a study are applicable outside of the study (generalisability and clinical applicability)

Why grade evidence?

- People draw conclusions about
 - quality of evidence (internal validity)
 - strength of recommendations (external validity)
- Busy clinicians need summaries
- Systematic and explicit approaches can help
 - facilitate critical appraisal
 - protect against errors
 - resolve disagreements
 - communicate information

First attempt



- The Canadian task Force on the Periodic Health Examination published levels of evidence and strength of recommendations
- Both the evidence statements and recommendations were graded to communicate the quality of evidence available.
- Based on study design alone

Canadian Task Force on the Periodic Health Examination: **The periodic health examination.** *Can Med Assoc J* 1979, **121**:1193-254

Example

Condition	Effectiveness of prevention & Treatment	Evidence quality	Manoeuvre	Recommendation
Tuberculosis	BCG vaccine effective treatment; Chemoprophylaxis prevents development in those affected	I	BCG immunisation & chemoprophylaxis	A (for high risk group)

Evidence quality	I	Evidence obtained from at least one properly randomized controlled trial.
------------------	---	---

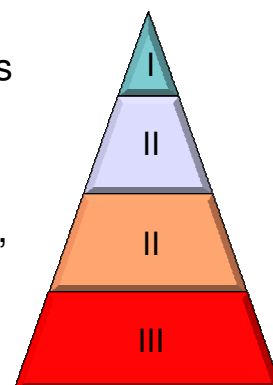
Recommendation classification	A	There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination.
-------------------------------	---	--

Hierarchy of evidence

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies & Case Control Studies
- Case Reports & Case Series, non-systematic observations
- Opinions

BIAS



About this first attempt...



Method Simple, easy to use



Too simple; many implicit judgments

- Should a small, poorly designed RCT be considered level I evidence?
- Should RCTs with conflicting results still be considered level I evidence?
- Should observational studies always be considered level II evidence, regardless of how convincing they are?

So many grading systems!

- 2002 survey identified 40 grading systems
- 2006 study found 20 more

- Australian NHMRC
- Scottish Intercollegiate Guidelines (SIGN)
- Oxford Center for EBM
- US Preventative Services Task Force
- Professional organizations (AHA/ACC, ACCP, AAP, Endocrine society, etc....)



1. Agency for Healthcare Research & quality, AHRQ,2002
2. Schunemann et al, Health Res Policy Syst 2006



Which grading system?

Recommendation for use of oral anticoagulation in patients with atrial fibrillation and rheumatic mitral valve disease

Evidence	Recommendation	Organization
• B	Class I	➤ AHA
• C+	1	➤ ACCP
• IV	C	➤ SIGN

So what is wrong with grading systems?

- Confuse quality of evidence with strength of recommendations
- Lack well-articulated conceptual framework
- Criteria not comprehensive or transparent
- Do not accommodate novel research designs
- Ignore that different types of questions need different types of evidence

Atkins et al. (2004) Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches The GRADE Working Group; *BMC Health Serv Res* 4:38

Glasziou P, Vandenbroucke J, Chalmers I, Assessing the quality of research; *BMJ* 2004;328:39-41



And more problems!

- RCTs are poor at demonstrating external validity
- Rarely answer questions about diagnostic accuracy or less common adverse effects of treatments
- Sometimes are unnecessary or impossible to undertake
- Tend to downgrade the validity, and relevance of observational studies
- Use of hierarchies of evidence does not ensure consistency

Example

HRT & women at risk of breast cancer



Are women taking HRT at increased risk of developing breast cancer?

NICE Clinical Guideline 14, Classification and care of women at risk of familial breast cancer in primary, secondary and tertiary care, May 2004.

NICE 2004: Levels of evidence & classification of recommendations

Level of evidence	Description
I	Meta-analysis of randomised controlled trials; at least one RCT
II	At least one controlled study without randomisation; at least one other type of quasi-experimental study
III	Non-experimental descriptive studies (comparative studies, correlation studies and case-control studies)
IV	Expert opinion, formal consensus

Recommendations

- C** Directly based on category III evidence, or extrapolated recommendation from category I or II evidence
- D** Directly based on category IV evidence, or extrapolated recommendation from category I, II or III evidence

Evidence

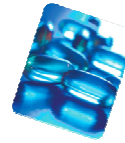
- The Million Women Study (Lancet 2003; 362)
 - Cohort study

Evidence level III
- The Collaborative Group on Hormonal Factors in Breast Cancer (Lancet, 2001; 358: 1389-99; Lancet, 2002; 360: 187-95).
 - Collaborative re-analysis of individual patient data
 - Studies identified by consultation with collaboration members only
 - Systematic quality assessment of studies not undertaken

Evidence level III



Recommendations



- Women with a family history of breast cancer who are considering taking or already taking HRT should be informed of the increase in breast cancer risk with type and duration of HRT. (C)



- HRT usage in a woman at familial risk should be restricted to as short a duration and as low a dose as possible. (D)

In fact....

- **Collaborative Group Study:**
 - Systematic review of cohort studies with re-analysis of individual patient data
 - Cover 90% of worldwide epidemiological data
 - Redefines outcomes
 - The Group re-analysed over 80% of the worldwide data and continues to provide definitive evidence on risk factors for breast cancer



High quality evidence to answer the question

The Collaborative Group on Hormonal Factors in Breast Cancer (Lancet, 2001; 358: 1389-99; Lancet, 2002; 360: 187-95).

And from the guideline authors...

“Despite attracting a ‘low’ level of evidence it is nonetheless the best and methodologically appropriate evidence. This should be seen as a shortcoming of the grading approach rather than a reflection of the evidence available.”

“A ‘D’ recommendation does not mean it is not an important recommendation. It only reflects the level of evidence that underpins the recommendation and is not a reflection of its clinical or policy relevance or importance.”

NICE CG41 Familial breast cancer: full guideline
<http://guidance.nice.org.uk/CG41/Guidance/pdf/English>

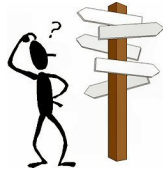


Example: Stopping smoking

Principle: Stopping smoking reduces risk of cancer and heart disease



- Consistent case-control evidence in 1950s
- Size of association compelling
- Causation established beyond doubt
- Time trends of smoking and cancer support this
- Randomising people to smoking in a trial is unethical.



Option 1: Extend and improve existing hierarchies

GRADE



- Focus on all important outcomes related to a specific question
- Clear separation between quality of evidence and strength of recommendations
- Breadth, intensity of development process
- Conceptual framework
- Comprehensive, transparent criteria
- Wide endorsement and use

Option 2: Abolish hierarchies

"Hierarchies attempt to replace judgement with an oversimplistic, pseudo-quantitative, assessment of the quality of the available evidence."

Teach practitioners principles of research instead



Michael Rawlins, Harveian Lecture Royal College of Physicians, London, 2008



Which option? Need to consider Both!



- Need to make explicit the type of evidence on which recommendations are based
- Need an informative and transparent framework (e.g. GRADE)

But

- This framework needs discussion and possible revisions
- There should be ongoing inquiry about how best to rate, review and use evidence

Ultimately

- Is it possible to collapse multiple dimensions of quality ?
- Should we grade guideline recommendations?
- What proof is there that a grading system improves outcomes?