RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

# Assessing the body of evidence and grading recommendations in evidence-based clinical practice guidelines

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 with acknowledgements to Janine Keough

RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

# Background

2003/4: Recognition that hierarchy of evidence ≠ grades for recommendation; and

need to grade the 'body' of evidence

2004: NHMRC commissioned Frameworks document\* which described 9 grading systems

\*The utilisation of established frameworks in assessing and applying non-intervention/non-randomised evidence (prepared by HTAnalysts)

## Timetable & background (2)

- 2004/5: Development of a system to grade a body of evidence (based on SIGN and its considered judgments)
- 2005/6: piloting of grades
- Then revision of NHMRC 'Guidelines for guidelines' suite of publications





RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

## Five components

#### 1. Volume of evidence

(number of studies, quality and relevance)

- 2. Consistency
- 3. Clinical impact
- 4. Generalisability (target popn)
  - 5. Applicability

(to local healthcare system)

RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

#### 1. Volume of evidence

#### A: Excellent

(several level I or II studies with low risk of bias)

#### **B:** Good

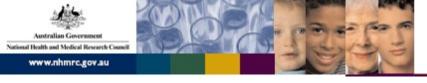
(1 or 2 level II/multiple level III - low risk of bias)

#### C: Satisfactory

(level III with low risk of bias/I or II with moderate risk of bias)

#### D: Poor

(level IV studies or level I to III studies with high risk of bias)



## 2. Consistency

A: Excellent

(all studies consistent)

**B**: Good

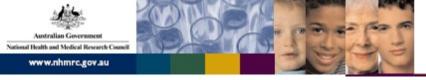
(most studies consistent)

C: Satisfactory

(some inconsistency, can be explained)

D: Poor

(evidence is inconsistent)



## 3. Clinical impact

A: Excellent

(very large)

**B**: Good

(substantial)

C: Satisfactory

(moderate)

D: Poor

(slight or restricted)



## 4. Generalisability

#### A: Excellent

(populations studied are same as target population)

**B**: Good

(similar)

C: Satisfactory

(different but clinically sensible to apply)

D: Poor

(different, hard to judge whether sensible to apply)



## 5. Applicability

#### A: Excellent

(directly applicable to [Australian] healthcare context)

#### **B**: Good

(applicable with few caveats)

#### C: Satisfactory

(probably applicable with few caveats)

#### D: Poor

(not applicable)

RESEARCH • ETHICS • ADVICE • COMPLIANCE & EVALUATION

# Grading

- Write recommendation wording (if not already done)
- Overall grade is the sum of the grades for the 5 individual components
- Recommendation cannot be graded A or B unless the volume and consistency are both either A or B



## **Body of evidence:**

A: can be trusted to guide practice

B: can be trusted to guide practice in most situations

C: provides some support for recommendation(s) but care should be taken in its application

D: is weak and recommendation must be applied with caution



# An example

### **Prognosis**

Several prospective cohort studies show that volume and stage of tumour should be considered when assessing treatment options

**OLD GRADING: IV** 



# New grading

A: Can be trusted to guide practice

Volume: A

Consistency: A

Clinical importance: B

Generalisability: A

Applicability: A



## Feedback so far

What's wrong with levels?

Why not use GRADE?

- Clinical importance should not be part of grading recommendations
- Workshop on Wednesday!