Grading and the GRADE instrument

G-I-N meeting

November 2004

Regina Kunz and Cindy Farquhar

Professional good intentions and plausible theories are insufficient for selecting policies and practices for protecting, promoting and restoring health.

Iain Chalmers

How can we judge the extent of our confidence that adherence to a recommendation will do more good than harm?

GRADE

Grades of Recommendation Assessment, Development and Evaluation

Why bother about grading?

- People draw conclusions about the
  - **quality of evidence**
  - **strength of recommendations**
- Systematic and explicit approaches can help
  - protect against errors
  - resolve disagreements
  - facilitate critical appraisal
  - communicate information
- However, there is wide variation in currently used approaches

Confused?

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

Guideline development process

Quality of evidence

The extent to which one can be confident that an estimate of effect or association is correct.

It depends on the:

- **study design** (e.g. RCT, cohort study)
- **study quality/limitations** (protection against bias; e.g. concealment of allocation, blinding, follow-up)
- **consistency of results**
- **directness of the evidence** including the
  - **populations** (those of interest versus similar; for example, older, sicker or more co-morbidity)
  - **interventions** (those of interest versus similar; for example, drugs within the same class)
  - **outcomes** (important versus surrogate outcomes)
  - **comparison** (A - C versus A - B & C - B)

Quality of evidence

The quality of the evidence (i.e. our confidence) may also be REDUCED when there is:

- ↓ Sparse or imprecise data
- ↓ Reporting bias

The quality of the evidence (i.e. our confidence) may be INCREASED when there is:

- ↑ A strong association
- ↑ A dose response relationship
- ↑ All plausible confounders would have reduced the observed effect
- ↑ All plausible biases would have increased the observed lack of effect

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	**Study quality**: -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness**: -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association:** +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of a **Dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low	Any other evidence		

Categories of quality

- **High**: Further research is very unlikely to change our confidence in the estimate of effect. ⨁⨁⨁⨁
- **Moderate**: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. ⨁⨁⨁◯
- **Low**: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. ⨁⨁◯◯
- **Very low**: Any estimate of effect is very uncertain. ⨁◯◯◯

Judgements about the overall quality of evidence

- most systems just use evidence about primary benefit/outcome
- but what about other outcomes (downsides)?
- options:
  - ignore all but primary outcome
  - base it on the evidence for benefits
  - some blended approach
  - having separate grades for benefits and harms
  - weakest of any outcome

Strength of recommendation

The extent to which one can be confident that adherence to a recommendation will do more good than harm.

- **trade-offs** (the relative value attached to the expected benefits, harms and costs)
- **quality of the evidence**
- **translation of the evidence** into practice in a specific setting
- **uncertainty about baseline risk**

Judgements about the balance between benefits and harms

- Before considering cost and making a recommendation
- For a specified setting, taking into account issues of translation into practice

Clarity of the trade-offs between benefits and the harms

- the estimated size of the effect for each main outcome
- the precision of these estimates
- the relative value attached to the expected benefits and harms
- important factors that could be expected to modify the size of the expected effects in specific settings; e.g. proximity to a hospital

Balance between benefits and harm

- **Net benefits:** The intervention does more good than harm.
- **Trade-offs:** There are important trade-offs between the benefits and harms.
- **Uncertain net benefits:** It is not clear whether the intervention does more good than harm.
- **Not net benefits:** The intervention does not do more good than harm.

Judgements about recommendations

This should include considerations of costs; i.e. "Is the net gain (benefits-harms) worth the costs?"

- Do it
- Probably do it

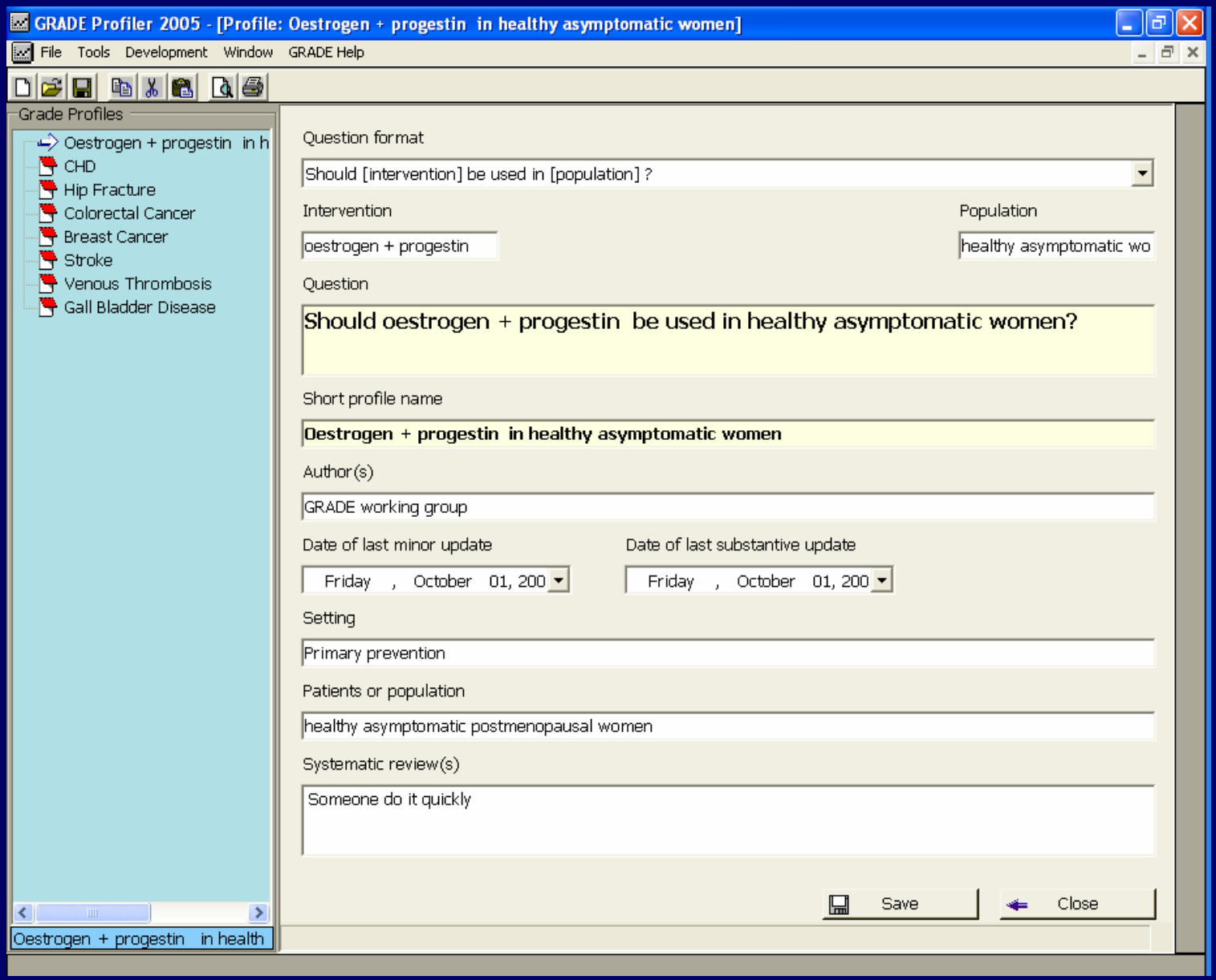
No recommendation

- Probably don't do it
- Don't do it

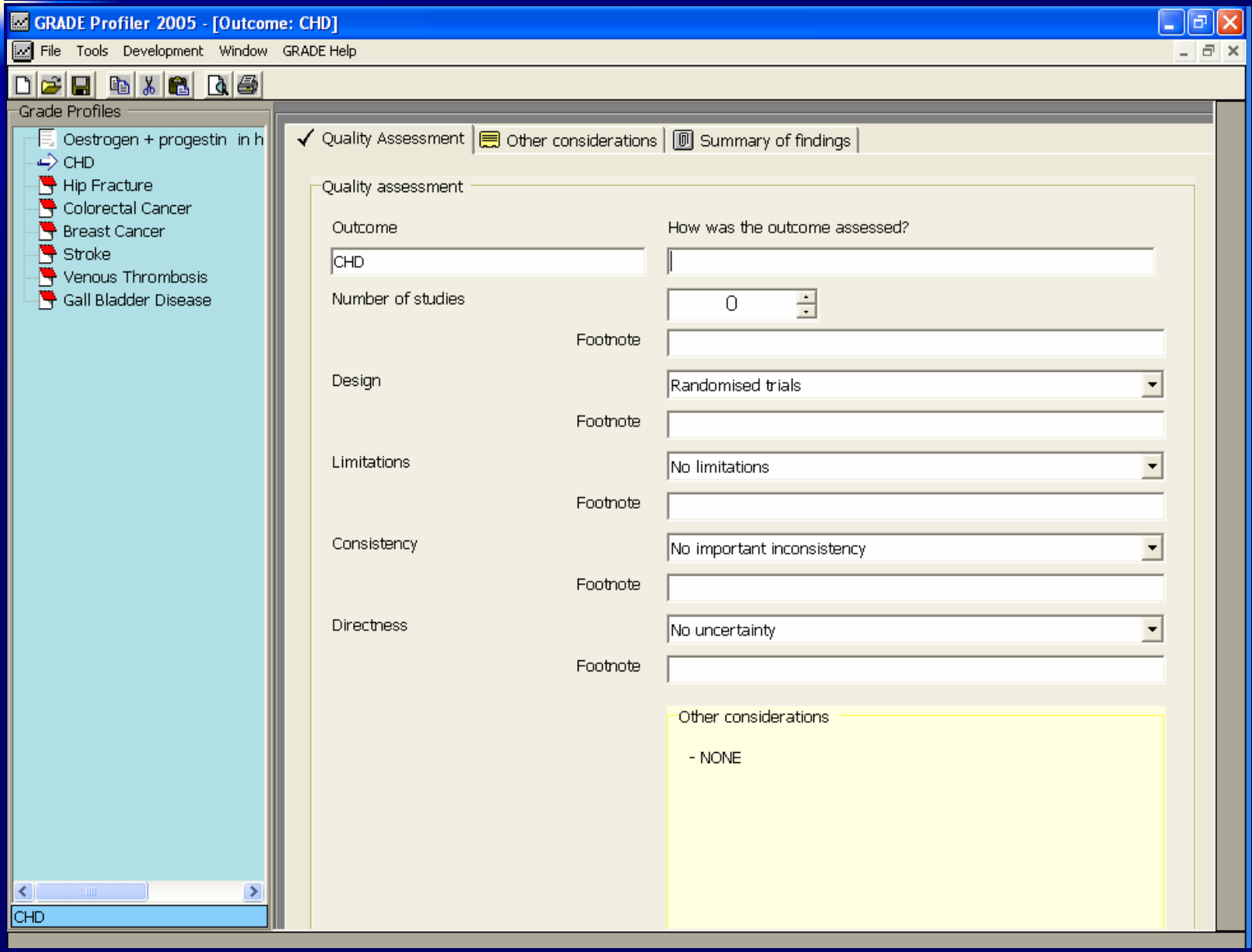
GRADE for diagnostic tests

Quality of evidence	Study design	Lower if *
High	Cross-sectional (or cohort) studies of patients with diagnostic uncertainty with direct comparison	**Study limitations** (including representativeness of population, choice of gold standard, incomplete performance of tests, independence of test interpretation) -1 Serious limitations -2 Very serious limitations -1 **Important inconsistency** **Directness** -1-Some uncertainty -2-Major uncertainty -1 **Sparse or imprecise data** -1 **High probability of reporting bias**
Moderate		
Low	Anything else	
Very low		

GRADE profiler (GRADEpro)



Separation by outcomes



Work in groups of two

- take a pencil (and paper)
- write down the most important issues/questions you have about GRADE

Small group sessions

- find a group
- select spokes person
- take 30 minutes to complete the task
- be prepared to criticise

Summary

What is good about GRADE?

-
-
-

What is most challenging?

-
-
-

What do we need to do next?

-
-
-