

Validation and modification of the Graphical Appraisal Tool for Epidemiology (GATE) for appraising studies in evidence-based guideline development.

Anita Fitzgerald

Outline

- Why a structured tool
- What is GATE
- Why did we try and validate it
- Methods
- Results
- Conclusions

Why should we consider a structured critical appraisal checklist?

- There is a tendency to look more critically at the studies whose conclusions we aren't keen on.
- 28 reviewers were asked to assess a single (fabricated) 'study' but were randomly allocated to receive either the 'positive' or 'negative' version (Mahoney 1977).
- The identical methods section of these fabricated versions was rated significantly worse by the reviewers of the 'negative' study compared with the 'positive' study.

Which one should we choose?

- There are several critical appraisal tools... many of them contain similar elements (appraisal questions/items).
- Katrak (2004) complete a systematic review of 121 published critical appraisal tools....

Persis Katrak et.al A systematic review of the content of critical appraisal tools.
BMC Medical Research Methodology 4:22 <http://www.biomedcentral.com/1471-2288/4/22>

their conclusions...

There was considerable variability in intent, components, construction and psychometric properties of published critical appraisal tools for research reports. **There is no "gold standard" critical appraisal tool for any study design, nor is there any widely accepted generic tool that can be applied equally well across study types.**

Based on the findings of this evaluation, we recommend that **consumers of research should carefully select critical appraisal tools for their needs.** The selected tools should have published evidence of the empirical basis for their construction, validity of items and reliability of interpretation, as well as guidelines for use, so that the tools can be **applied and interpreted in a standardized manner.**

What about GRADE?

- GRADE is great when there are adequate funds

But when there is less \$ for health research

- less interest in traditional guidelines
 - less reliance on big guideline groups
 - difficult to get outcomes prioritised to use GRADE effectively.
- At NZGG, face to face meetings are kept to a minimum
- Heading more towards smaller pieces of work eg. evidence updates, guideline adaptations, small carefully selected projects.

What do we need?

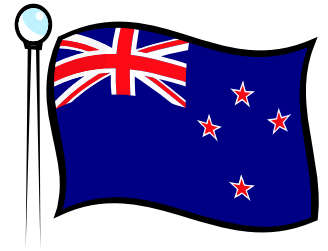
- NZGG needed a critical appraisal tool that could be used in these situations:
 - Full guideline group not available
 - Face to face meetings not funded
 - Small pieces of work with a fast turnaround time
 - Guideline adaptation/update
- In other situations GRADE is more appropriate

Our ideal tool

- Takes a reasonable time to complete
- Provides enough detail so that we can answer guideline group questions quickly and accurately.
- Relatively straightforward to complete so there is not a huge gap between experienced reviewers and new reviewers.

Graphical Appraisal Tool for Epidemiology

- New Zealand designed appraisal tool, originally used to teach EBM
- Tools for each study design
- Covers three aspects of appraisal
 - Internal validity
 - External validity/Applicability
 - Precision
- Easy to answer GDT questions without referring to the full text



GATE scores for individual items

- + ok, good: well reported and reliable
- X not ok, poor: study not reliable
- ? unclear/not reported: insufficient detail provided
- NA not applicable



Objectives

- To test inter-rater reliability for individual items on the GATE framework
- To compare judgements of summary measures of study validity between reviewers
- To document reviewers experience of using GATE.
- To adapt GATE for use within NZGG

Step one

- Researchers apply GATE appraisal checklist to 10-20 studies.
- Document ease of use while completing appraisals
 - Was it easy to understand and answer individual questions?
 - What difficulties did you encounter?
 - Do individual questions need clarification or require more explanation to answer?
 - Time taken to answer checklist - does it get shorter the more appraisals you do?

Step two

- Analyse results in terms of inter-rater reliability and ease of use.
 - percentage agreement
 - KAPPA
 - prevalence-adjusted bias-adjusted kappa (PABAK)
- Discuss adaptation for NZGG purposes among reviewers.

Step three

- Re-test the adapted version with a further 10-20 studies
- Analyse the results in terms of inter-rater reliability and ease of use.
- Discuss final adaptations.

Step four

- Make a valiant effort to update the acronym
 - **G**uideline-specific **A**ppraisal **T**ools.... how **E**ducational
 - **G**ood **A**tttempt **T**o **E**valuate studies
 - **G**reat, now I can **A**ccurately **T**ell **E**veryone
 - **G**ee, there's **A** **T**ool for **E**verything now...
 - **G**reat **A**ppraisal **T**ool... **E**nvious?!

RCT results - round one

- Percentage agreement in the first round ranged from 20% to 100%, with a median of 65%.
- Inter-rater reliability was variable ranging from a PABAK of -0.6 (poor) to 1.0 (very good).
- Inter-rater reliability appeared to be lower for items relating to internal validity and applicability than items about precision.
- Agreement on summary scores was rated poor for all categories (PABAK -0.6 to 0.2).

RCT results round two

- Agreement between reviewers for modified GATE items in the second round ranged from 0% to 100% with a mean of 70%, an improvement on the first round.
- Inter-rater reliability ranged from a PABAK of -0.2 (poor) to 1 (perfect).
- Agreement on summary scores improved in crude agreement and PABAK score for all summary measures.

Modifications to randomisation item of the RCT tool- original version

Question: Allocation to exposure and comparison groups: random or by measurement?

Explanatory notes: Was allocation to exposure and comparison randomised? If described as a randomised trial, what was the method of randomisation? If cross-over trial, was order of intervention randomised? If not randomised, was significant confounding likely?

Modifications to randomisation item of the RCT tool – following round one

Question: Was the method of randomisation adequate?

Explanatory notes:

- +
 - random number table
 - computerised random number
 - coin tossing/dice throwing
 - shuffling cards or envelopes

- x
 - odd or even date of birth
 - date (or day) of admission
 - clinic record number

- ? – not enough information

Modifications to randomisation item of RCT tool – following round two

Question: Was the method of randomisation adequate?

- +
 - random number table
 - computerised random number
 - coin tossing/dice throwing
 - shuffling cards or envelopes

- x
 - odd or even date of birth
 - date (or day) of admission
 - clinic record number

- ? – not enough information

Randomisation by stratification, blocking, or minimisation:

- if allocation concealed = +
- not concealed = x
- not stated/no enough details = ?

Conclusions

- The amended GATE checklists demonstrate improved inter-rater reliability for appraising studies.
- Critical appraisal checklists used in guideline development could be systematically improved by undertaking inter-rater reliability assessments.

Want a copy?

- Of the original version
 - <http://www.fmhs.auckland.ac.nz/soph/depts/epi/epiq/ebp.aspx>
- Of our adapted version
 - afitzgerald@nzgg.org.nz

No conflicts of interest to declare