The development of a tool for the critical appraisal of studies reporting moderators of treatment effect

R van Hoorn*; W Kievit; GJ van der Wilt; M Tummers on behalf of INTEGRATE-HTA

*PhD student, Radboud Institute for Health Sciences

Radboud umc

INTEGRATE-HTA

http://integrate-hta.eu

- INTEGRATE-HTA aims to adapt and develop concepts and methods for HTA to enable an <u>integrated assessment</u> of issues of <u>complex technologies</u>

- Complex technologies
  - Heterogeneity of patients
  - Interacting factors
  - Multiple outcomes

Moderators of treatment outcome

Factors that influence treatment effects

- To determine which treatments work best for whom
- Making medicine more personalised
- Better valuation of research outcomes
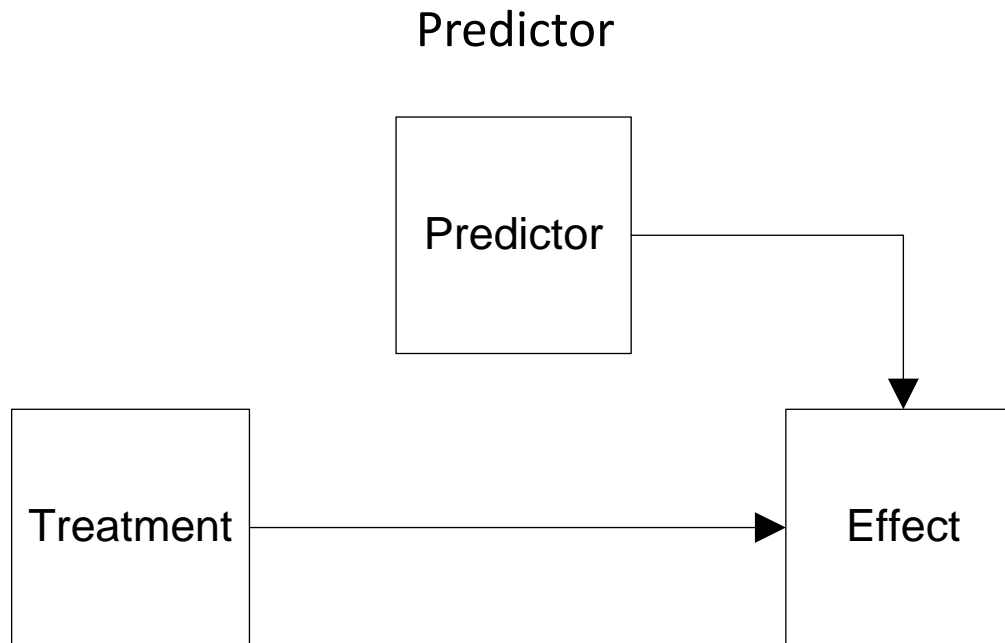

FLAW of AVERAGES
AVERAGE DEPTH 3 ft.
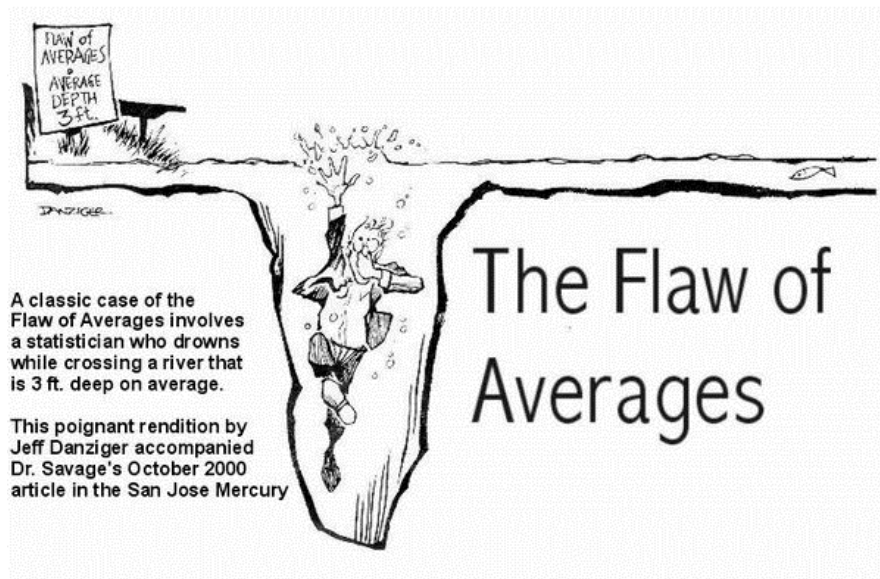
A classic case of the Flaw of Averages involves a statistician who drowns while crossing a river that is 3 ft. deep on average.

This poignant rendition by Jeff Danziger accompanied Dr. Savage's October 2000 article in the San Jose Mercury
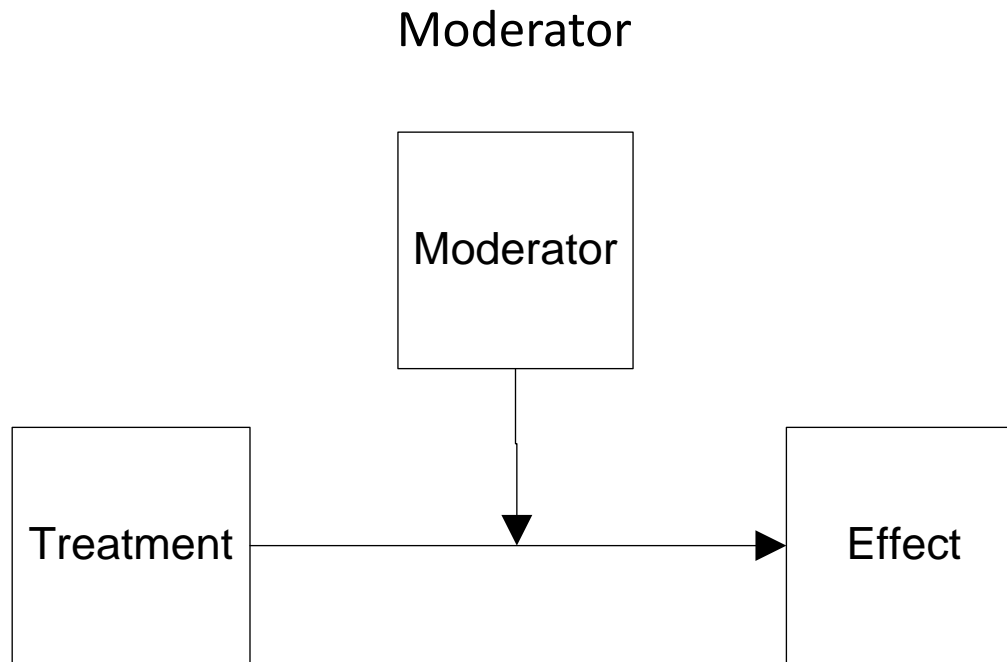
The Flaw of Averages

Definitions

Predictor

Predictor

| Treatment | → | Effect |

Definitions

Moderator

Moderator/predictor analysis

- Usually secondary analysis (subgroup or regression analysis)
    - Common in Randomized Controlled Trials

- Often part of a larger collection of hypothesised factors
- Despite accessible guidance, quality of analyses is variable

"While it is generally recognised that subgroup analyses can produce spurious results, the extent of the problem is almost certainly underestimated" (Brookes et al., 2001)

Deciding on analysis after looking at the data is "dangerous, useful, and often done" (I.J. Good 1983)

Aim

- Critical appraisal is needed
- Current appraisal tools aim to appraise main effects
- Appraisal of moderators is important
  - Post hoc versus a priori

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, and its value and relevance in a particular context (A Burls, 2009)

Methods

- Literature search
  - Theoretical pieces, guidelines
  - Existing critical appraisal tools
  - Identification of relevant researchers

- Delphi panel
  - RAND/UCLA appropriateness method
  - Relevant participants
  - Two rounds

- Testing of the tool: palliative care case study
- Revision
  - Third Delphi panel round
- Finalisation

Design choices

- Considerations on tool design based on existing tools
  - Checklist
  - Multiple choice (Yes, Partially, No,
                    Not (clearly) described, Don't know)
  - No summarising score
  - No overall study design appraisal

Radboud umc

First results

- 48 items retrieved from literature → 31 items included in test version
  - Study design (4)
  - Population selection and measurements (4)
  - Analysis (3)
  - (presentation of) results (9)
  - Meta-analysis (8)
  - Body of evidence (3)

Study design

	Yes	Partially	Not (clearly) described	No	Don't know
1. Was the selection of moderators to consider based on expert knowledge and/or theoretical considerations?	☐	☐	☐	☐	☐
2. Is there (in)direct evidence that supports the hypothesized interaction (e.g. biological rationale)?	☐	☐	☐	☐	☐
3. In the case of hypothesis-testing, are moderator hypotheses and their analyses defined a priori?	☐	☐	☐	☐	☐
4. In the case of hypothesis-testing, was the selection of moderators to consider pre-specified?	☐	☐	☐	☐	☐

First results

- Strong points
  - Added value over existing checklists
  - Helps systematically appraising studies
  - Common problems in literature (e.g. power analysis)

- Problems
  - [Identification of moderator versus appraisal]
  - Reporting bias
  - Items discerning hypothesis testing and generating
  - Non-randomised studies
  - Non-intervention studies

What is next?

- Revision of the checklist
- Final adjustments using
  - Delphi panel
  - External expert panel
- Finalisation

http://integrate-hta.eu

Universität Bremen

Radboud umc

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

ECORYS

UiO : University of Oslo

UNIVERSITÀ CATTOLICA del Sacro Cuore

The University Of Sheffield.