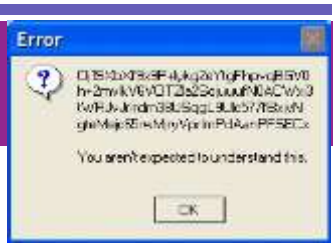
# TECH-VER: A QUALITY ASSURANCE AND VERIFICATION CHECKLIST FOR HEALTH ECONOMIC DECISION ANALYTIC MODELS OR… MODEL POLICING 101

**NASUH BÜYÜKKARAMIKLI, MAUREEN RUTTEN-VAN MÖLKEN, HANS SEVERENS AND MAIWENN AL**
ERASMUS UNIVERSITY ROTTERDAM, INSTITUTE OF HEALTH POLICY AND MANAGEMENT, THE NETHERLANDS

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

## RECAP: VERIFICATION / MODEL ERRORS




- Verification as an integrated part of validation
  - (QC, technical validation, testing, debugging, auditing...)
  - Model implemented according to its concept?
- Model errors/ credibility threats (Tappenden & Chillcott, PE 2014)
  - Type of errors: unequivocal errors, violations of best practice, matters of judgement.
  - Are they important?
- Verification > checking for errors in the model
- Model policing: necessary evil (time/budget, adversary for creativity, etc.)

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

## CHECKLISTS / GUIDELINES

- Drummond and Jefferson 1996, Philips et al. 2004, Eddy et al. 2012, AdvisHE. 2016
- Those checklists have a broader aspect
  - Some of these above: a validation type has been conducted or not.
  - Not detailed enough for technical aspects/verification purposes
- TECH-VER: global technical verification checklist.
- Type of errors/ credibility threats targeted (Tappenden & Chillcott, PE 2014)
  - Use of evidence
  - Concept → implementation
  - Model implementation
  - Reporting/evaluation/interpretation

Erasmus Universiteit Rotterdam
Institute of Health Policy & Management

## TECH-VER: OR MODEL POLICING 101

- Manuscript including the checklist is in submission, email for a copy. Will be published on www.imta.nl
- Ingredients: the model(s), technical report, (independent) reviewer(s)
- Might be helpful in identifying errors (not a guarantee).
- Ease of communication between different reviewers/users on the verification efforts conducted
- Combination of black-box testing and descriptive questions to structure white/gray-box testing.
- Can be used for training purposes (for junior health economists)
- Should be coordinated with other validation efforts
- Some of the questions might not apply in some situations. (Common sense)

Erasmus Universiteit Rotterdam
Institute of Health Policy & Management

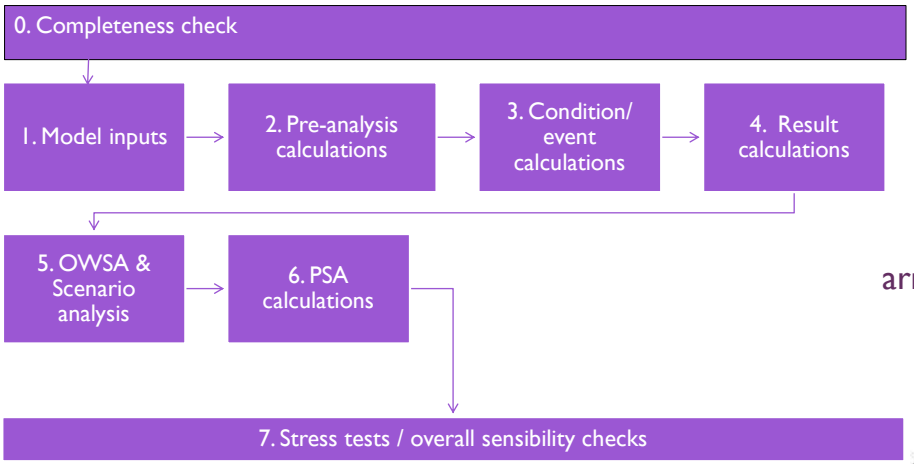
## METHODS

Step 1: Development of the 1st draft based on literature & authors' experience (NB, MRvM, HS, MA)

Step 2: Pilot studies with several health economic models/ iterative revision after each case

Step 3: Discussion of the checklist 1) in a workshop at the faculty 2) in a conference (LOLA 2016) with other health economists in the Netherlands

## STAGES OF TECH-VER




Direction of the arrows can be different!

## MODEL INPUTS / PRE-ANALYSIS CALCULATIONS

- Inputs table (Reported input ≠ input used in the model ≠ value in source reference)
- Subpopulation values make sense?
- Inputs consistent with the perspective?

**Pre-analysis calculations**: (reference inputs → ready for use model inputs)

- Identify cells/programming lines where pre-analysis calculations conducted.
- Which methods are used? Are these methods applied correctly?
- Above questions should be applied for:
  - Safety & efficacy data | QoL & utility | Costs | Other?



## CONDITION/EVENT BASED CALCULATIONS (ENGINE)

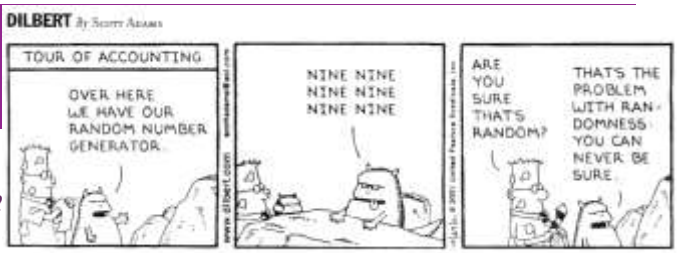
- Identify cells/programming lines
  - where conditions/events take place. How?
  - where corresponding costs/utilities were assigned to them. How?
- Implementation correct? in line with the concept (as explained in the report?)
  - Correct probabilities/rates/QALYs/costs applied for correct conditions/events in correct time?
- Trace how conditions/events/QALYs/costs change in time (e.g. *t-1*, *t*, *t+1*)
- Black-box tests

## RESULT CALCULATIONS

- Identify cells/programming lines
  - where total/incr. (un)discounted LYs/ QALYs/ Costs are calculated. How?
  - where other results (disaggregated costs/QALYs) or other relevant outcomes are calculated. How?
- Calculations correct? In line with the report? Interpretation correct?
- Other automatically generated tables/figures. (Power) pivot or power view features for spreadsheet
- Reported results = model results (base case and scenario)
- Black-box tests

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

## OWSA/ PSA/ SCEN. ANALYSES




- Uncertainty parameters
  - which parameters are investigated? If not why not?
- Appropriate distributions? 95% C.I.s?
  - Correctly applied? Model consistent with report?
- Correlation taken into account?
- Identify & check the cells/programming lines
  - Are OWSA/PSA and scenario analyses correctly applied? (Tornado diagrams, CEAC, PSA Scatterplot)
- VoI calculations (if anly)
- Consistency between reported results with model results (and PSA with deterministic)
- Regenerate some of the OWSA results (most impactful) & black-box tests

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

## STRESS TESTS/ SENSIBILITY CHECKS

- Apply different time horizons
- Reprogramming (parts of) the model
- If analytical simplification is possible
  - Compare analytic results (pen/paper) with model results in the simplified case
- Black-box tests (with extreme input value manupilations)
- A bit pushing the boundaries: unexplained implausibility of the model results
  - With clinical study results (used as an input or not)?
  - Other models?
  - Clinical expectations?

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

## PILOT STUDIES

- Different stages of life cycle / different purposes
  - Early HTA models for strategic insights / NICE (re)submission models / academic models for publication
- Different software environments
  - **Spreadsheet** / VBA macros
- Different model types
  - Decision trees/**Markov model**/micro-simulation
- Different therapeutic areas
  - Oncology/ respiratory/ rhematoulogy/ mental health
- Different reviewers
- Different levels of detail
- One controlled study (NICE TA, ERG without TECH-VER and an external member with TECH-VER)

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

# LEARNING POINTS FROM PILOT STUDIES

- Helpful in identifying model errors (unequivocal) & inconsistent inputs/reporting & deviations from the conceptual design
- Not always directly the root cause was found. Often symptoms (black-box) triggered identifying root cause
- Error correction → changes in ICER.
  - E.g. in a NICE STA: ICER increased 15% (dangerous neighbourhood). In an early HTA model: Treatment switched from being dominant to being cost effective
- Controlled study: errors identified by the ERG only ≈ errors identified by the ext. member.
- Improvement in transparency & ease of communication of verification efforts
- Required time: 1-2 days to 4-5 days
  - Model size, reviewer, clinical area etc.
  - Learning curve effect / tailored approaches for efficiency.

# CONCLUSION

- TECH-VER: a comprehensive checklist for technical verification.
- Seems to work, but impossible to estimate treatment effectiveness
  - Not a guarantee for error free models: no passing grade so beware of the false sense of security
- Never forget: model policing is a necessary evil. Should be used with common sense.
- Possible applications:
  - Verification of an externally developped model (one-time use for analysis or multiple use tool?)
  - HTA submissions (decision making) or scientific manuscript submissions
  - Guideline / Note to self while building models
- Should be always co-ordinated with other validation efforts
  - Implausible model results: due to a wrong model concept (LYs = rand()*Constant) OR **due to a copy paste error that created zombie patients in a conceptually valid, elegant model**

# DISCUSSION

- Is it effective? Is it cost-effective? (Under which threshold?)
- When should it be conducted? Which stage of the model life cycle?
  - After conceptual validation and before other model validation efforts that require model results? OR iteratively?
- What is the border between model concept and implementation?
  - Statistical methodology for pre-analysis: (different software environments)
- Future steps:
  - Improvement of model policing for other model types / softwares (e.g. Dynamic population models)
  - Increasing the awareness, sharing the best practices on verification. Short course on model policing?
  - More studies and insights on model errors are needed.
- Time to discuss: should verification be mandatory?
  - for HTA submissions?
  - for manuscript submissions?

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT