Guide for Ensuring High Quality of Modelling Outputs in the BODE^3 Programme

Prepared by selected members of the BODE^3 Team, University of Otago, Wellington:

Professor Nick Wilson, Professor Tony Blakely, Dr Nhung Nghiem,
Dr Christine Cleghorn, Dr Giorgi Kvizhinadze, June Atkinson, Dr Nisha Nair,
Dr Anja Mizdrak, Frederieke Sanne van der Deen

May 2017

Contents

Introduction

This document is designed for BODE^3 staff involved in epidemiological and health economic modelling work and for those considering BODE^3 research outputs. We envisage updating this document at 1-2 yearly intervals (or more immediately if we adopt any major new approaches).

Why high quality is important for modelling work

- **Knowledge gain and societal benefit.** Models are an important way of estimating relative health gains, costs and cost-effectiveness for interventions across the health sector (and beyond). Models are also developed to help decision-makers when the questions involved are too complex to be solved or weighed up just by individual humans or groups of people discussing a question. Modelling can be an efficient approach to informing decision-making if the alternative is establishing large and expensive trials or conducting other types of experiments. Modelling can also identify the key areas of uncertainty, which can then be used to prioritise further research to reduce such uncertainty. But there is some societal-wide distrust in models (especially if these are seen as "black-box" with limited transparency). Such public distrust is also related to the specific issue being modelled eg, it is higher for climate change models than it is for models that indicate how best to eradicate measles or to guide spacecraft to accurately land on another planet. In the health sciences it is therefore critical that modellers strive to explain the value of conducting modelling, to achieve high quality through appropriate quality control processes (eg, around model quality and model validation etc), and also to be transparent about limitations and assumptions.
- **Benefit to the University.** Given the important role of universities in modern democratic societies, it is desirable that their reputation is appropriately maintained and that governments and taxpayers continue to support them. Ensuring a high quality of research outputs by university staff is part of this process.
- **Benefit to the research team and individual researchers.** Maintaining research team and researcher reputations for quality is important for ensuring the future success with research grants and for individual career prospects.

Context for BODE3 modelling compared to other quality systems

Achieving high quality is important as detailed above, but at some point further improvements in model quality may be less than the opportunity costs in terms of time and resources. Giving too much emphasis on model improvements with diminishing marginal gains in accuracy will result in reduced total research outputs and reduced effort put into distribution of knowledge and stakeholder engagement.

Of note is that:

- **Epidemiological modelling work is typically not like other activities where precision is critical.** Some domains of human activity need extremely high certainty about when, how much and other outputs, e.g. : nuclear weapons on high alert; nuclear reactors; management of explosive chemicals, biological agents and bioweapons; and air transport etc. These type of systems are often tightly coupled[1] – and so uncertainty and errors can result in major disasters. In comparison, epidemiological modelling has

unavoidable uncertainty in its predictions: there is unavoidable uncertainty in inputs and our understanding of how health is 'produced', meaning uncertainty in outputs is inevitable (eg, uncertainty intervals about the likely quality-adjusted life-years [QALYs] gained). Epidemiological and economic decision modelling is just one part of decision-making (eg, political and other considerations are often considered alongside QALYs, net costs, or ICERs etc). But it is still highly valuable for the world to have a collection of highly developed epidemiological models on important health topics to provide some quantitative prediction and guidance for decision making. For example, tobacco control models may allow for effective prevention of some of the predicted one billion plus tobacco-related deaths likely to occur this century.

- **Multiple models may be more efficient than a single extremely high quality model.** It may be best for separate research groups to develop independent models than for one research group to develop a single extremely high quality model. This is because policy-makers will probably have more trust in similar results from separate models (if these are appropriately independent models), than a single model that took much resource (which is often limited) to develop.

Standard BODE^3 Quality Processes

The following are the agreed processes for BODE^3 epidemiological and cost-effectiveness modelling work. They build on the DECC QA Document[2] (see below) and the collective experience of the BODE^3 Team which has now accumulated substantial collective experience in modelling work (see the publications on the BODE^3 website).

Quality and workplace culture

Modelling is usually complex work and so consideration is given to the following workplace variables:

- A particular emphasis by the BODE^3 Team on making maximal use on a few highly developed models (eg, a tobacco control model, a diet/physical activity [PA] model, and a cardiovascular disease [CVD] model), as opposed to a larger range of models (where staff content knowledge becomes more thinly spread).
- Striving for an appropriate balance of interesting model development work by staff with the potentially more tedious work around extensive model checking.
- Taking great care with staff and student recruitment to ensure high quality personnel with relevant skill and knowledge bases.
- Investing in ongoing staff skill and knowledge development so that staff keep up with developments in modelling (this has been happening with conference attendance and staff visits to Melbourne University etc).
- Continuing with a weekly BODE^3 Team meeting to assist with communication and sharing with modelling challenges. At the end of each of these meetings there is a routine discussion of any emerging quality issues (especially in regard to current work activities).
- Continuing to promote open and respectful communication amongst team members.
- Continuing to promote early error reporting and where appropriate "near-miss" reporting.
- Continue a "no blame" culture around errors identified – with a strong focus on system changes to prevent error repetition and to upgrade quality mechanisms.

Individual work practices

While the most important approach is to have *systems* that focus on quality, individual team member behaviours also matter. That is, it is important that individuals should:

- Raise with team leaders and senior staff any issues of relevance to workplace functioning that may need to be addressed (see above). More general quality control issues can be raised by staff at the routine Monday morning meeting (and potentially then incorporated in routine quality processes eg, this or other documents).
- Raise any issues of real or potential conflicts of interest. (This should be rare as BODE^3 work is very much focused on research for official government agencies: HRC, MBIE and other NZ Government agencies.)
- Be familiar with the contents of this document on quality issues, with the BODE^3 Protocol, and relevant technical reports (as appropriate to work role).
- Put appropriate effort into model documentation – at the same time as models are being built. This includes writing a technical report that usually is submitted to journals alongside an article as supplementary information file, but also regular updates to the logbook in the model (which each of our models contains). All changes made to the model need to be documented in this logbook to allow all users to understand why and when these changes were made and if they were signed off by the appropriate other staff.
- Discuss model challenges with other team members who may contribute to finding solutions.
- Take regular short breaks when doing complex modelling work to ensure higher quality performance.
- Avoid complex modelling work when tired or unwell. All staff should normally have a range of less-demanding work tasks that can be done when they are slightly tired (eg, for nearer the end of the working day).
- Take sick leave when appropriate and not return to working until fully functional.

Specific quality processes around the phases of building epidemiological models within BODE3

Following the proposed leadership roles for modelling in the DECC QA Document[2], the designated roles in BODE^3 are shown in Table 1 for the main MSLT models for the period 2016-21. Other models used in BODE^3 (eg, for AAA screening, diabetes), will be managed as they arise.

Table 1: Designated roles for major BODE^3 Models as per 2017

Role (using DECC terminology) and descriptor	**Tobacco Multi-state Life-table (MSLT) Model**	**DIET/PA MSLT Model**	**CVD MSLT Model**	**Active Trans-port MSLT Model**	**Tobacco Fore-casting Model**
Senior Responsible Owner (SRO) • Holds overall accountability for the success of the model • Ensures correct evidence and data assembled • Maintains task (or project	Cleghorn (probably van der Deen later in 2017)	Cleghorn	Nghiem	Mizdrak	van der Deen

Role (using DECC terminology) and descriptor	Tobacco Multi-state Life-table (MSLT) Model	DIET/PA MSLT Model	CVD MSLT Model	Active Trans-port MSLT Model	Tobacco Fore-casting Model
management) lists • Delegates and/or coordinates work on the model • Responsible for the technical documentation around the model (includes a list of all the non-trivial equations used in the model)					
Deputy SRO (to facilitate management if SRO is unavailable or changes job and logic checks to model tasks performed by SRO and vice versa)	van der Deen	Mizdrak	Kvizhinadze	Cobiac*	Cobiac*
Senior Analyst • Approves, or modifies, analysis plans for use of model • Assists or leads analyses using the model	Cleghorn (probably van der Deen later in 2017)	Cleghorn	Nghiem	Mizdrak	van der Deen
Approving Body (AB) • Signs off (and checks where necessary) that all QA tasks are completed – both those identified in this document, and the project-specific lists held and managed by SRO • Signs off all key aspects of model design and content of research outputs • Signs off all publications arising from the model (ie, prior to submission)	Blakely, Wilson	Blakely, Cobiac*	Wilson, Blakely	Blakely, Cleghorn	Wilson, Blakely

* Still being confirmed as of May 2017.

Table 2 outlines quality control and assurance processes.

Table 2: Specific quality control processes used for BODE^3 MSLT Models

Domain	Summary details
Designated roles	For each major model and major new version of an existing model – precise roles are agreed (eg, as per Table 1).
Model conceptualisation	For each model there is a clear direction on the approach to take (eg, is a MSLT built in Excel still optimal), and the most important disease states to include. Particular emphasis is given to precise descriptions of the interventions, which can ideally span what is potentially fairly feasible in a modern democratic society to the more hypothetical (the latter being a quality check as well on the plausible envelope for maximal impact). Comparator specification can be challenging and needs to be carefully determined by the modelling team. In some domains there might be a need to check with external advisors to ensure that the assumptions are not too simplistic (to maximise subsequent credibility of outputs with decision-makers).
Parameter identification	BODE^3 strives to make the best use of NZ's high quality epidemiological data and relatively high quality cost data.[3] Intervention-related parameters are sourced ideally from systematic reviews of randomised trials, and relevant NZ specific studies (eg, on intervention uptake and acceptability). Only when there are no other options, expert opinion is used for estimating input parameters and

Domain	Summary details
	associated likely uncertainty. Many of the model parameters are sourced from the Global Burden of Disease (GBD) Study as a default. Otherwise, one person in the modelling team (the default being the SRO) will maintain a regular automated search strategy (of PubMed) for new literature around key parameters. This can allow for updates right up to the point of "final" model runs (or even when an article is revised after journal article reviewer feedback). Indeed, there may need to be a re-evaluation of the value of modelling a particular intervention if the investigative work suggests a serious lack of key parameters (and the implausibility of these coming from an expert elicitation processes).
Model construction	Documentation around model building is routinely detailed in the Technical Report relating to the specific model. But modifications and extensions to the built model are detailed in a "Readme tab" in MSLT Excel models. This "Readme tab" and/or a separate "checking tab" is now being used for logging QA checks and issues as per the DECC "QA log" terminology. The latter is a summary of the checking being carried out, the results of this checking, any associated action points and a place for the AB to sign off each check. Further details can be documented in the Technical Report associated with the model if deemed necessary. Once a version of the model has been signed off by the AB then additional processes will take place for every change that takes place that impacts the results of the model. Models will be run before and after the change and results by age, sex and ethnicity will be pasted into the "Readme tab". Differences in the results before and after the change will be examined and any unexpected changes will be followed up with the appropriate checks. A summary of the process will be included in the "checking tab". All model builders should be aware of the quality details in the DECC Quality Document.
Scenario and sensitivity analyses & extreme value testing	BODE^3 experience indicates that having a very extensive range of scenarios and sensitivity analyses is a good way to identify any residual quality issues (and to improve the potential usefulness of the modelling work for policy-makers). Even if some of these analyses might not be published in the final work – it can be worthwhile to still run them as part of the checking process. It can be more intuitively easy to check results using a 0% discount rate (though a 3% discount rate can also be used as this allows comparisons with other BODE^3 results). Sensitivity analyses can overlap with extreme value testing – but not always. Hence it is useful to test the model with mathematical extremes (eg, disease incidence drops to zero, intervention has zero effect). Such testing should be routine and be summarised in the "Readme tab" (with potentially further details in the Technical Report).
Version control	The SRO should be responsible for version control and for appropriate documentation in the "Readme tab" or Technical Report. Thought needs to go into generating explanatory filenames for model variants – eg, "Special vers CVD MSLT for MPH". Similarly for where they get stored on the shared drive. Of note is that aspects of final models can have components "locked" to prevent accidental modifications.
Model calibration and validation (further details follow below)	This is routine in BODE^3 with the framework used being based on ISPOR guidance.[4] Examples of relevant details and text used in BODE^3 documentation are detailed in subsequent sections in this document. In particular, BODE^3 is in a very good position to perform comparisons between models built or being built by collaborators in Melbourne, Brisbane, Oxford, Cambridge and Erasmus (see the section on validation below). But also in NZ there are occasional opportunities for model comparison (eg, the BODE^3 colorectal cancer screening model has been usefully compared with one developed by others [MoDCONZ]).
Final model checking	At key time-points such as the end of one batch of model development or before a batch of publications, the model should be checked and these checks are signed off by the AB. The SRO for the model requests that other BODE^3 team members help conduct checks of key aspects of the model (including

Domain	Summary details
	design and layout, correct links to data, and formulae used). Doing this collectively over a designated week can potentially lead to useful synergies and efficiencies. The extensiveness of this checking process depends on time and resources, but for models informing critical government decisions the following checks, comparing relevant results to expected values, could be considered: • Set all relative risks to 1.0 • Set a variety of different time-lags including extreme values • Run and compare the range of interventions • Set extreme values for change in risk factors • Set extreme values for population distribution of risk factors • Change the targeting of the intervention to specific population groups • Changing all the scenario switches • Randomly check back to source, 100 values of input: (Costs, Disability Rates) • Change pYLDs to: 1.0 and then 0.0 • Change costs to 0. Compare baseline costs in different models • Compare baseline QALYs in different models (when diseases are the same) • Look at risk factor by disease for various interventions • Randomly check 5-10 formulae in each BLOCK of each Exposure_risk factor spreadsheet • Check results for various interventions over the next 10, 20 etc model years • Check all the RRs match the RRs in the source paper • Check that the RRs feed into the exposure sheet(s) (where the interventions feeds through the model) correctly • Change starting age (eg, set 4 different ages) and check resulting prevalence for some diseases • Compare the disease specific deaths in model inputs to MOH/BDS data • Check links between disease sheets of diseases just added to the model, new disease data and the life-table • Run model for life-time and compare disease mortality rates to the original mortality rates used as DISMOD inputs. Start at ages 32, 42, 62 & 72 years. For a scattering of 5 year age-groups for each sex and ethnic group. Mortality risk, convert to rate. Do for a selection of diseases. • When the models are run with Ersatz consider un-ticking the option of 'no screen updates while running' and check sheet by sheet if things are running as they are set up while running the model (this process has worked well with modelling tobacco endgame strategies). • Go through and check follow-up on all comments in the Readme tab since the last check • Check documentation: DISMOD process and graphs and check the readme files for new diseases A full list of all the non-trivial equations used in the model should be in the Technical Documentation – and each of these should be checked with what is in the actual model software.
Paper and report writing	To maximise quality, the writing team should be writing the early drafts of the relevant journal article before the model building occurs as well as during the model building process. That is, writing a draft of the *Introduction* forces the team to be clear on the research questions and to be familiar with key existing literature on the intervention domain. Writing the draft *Methods* forces the team to be clear on the design features along with the parameter values and intervention and comparator specifications. Past publications provide a good template for the typical approach to analysis, but it is often useful to produce additional non-critical additional analyses for checking purposes (potentially for

Domain	Summary details
	online Appendices). Approving the initial paper draft is a Senior Analyst role – see Table 1 above – as well as team-wide.
Internal peer-review of outputs	All team members need to carefully critique drafts of the research products from modelling work (presentations, reports, journal articles). Named authors need to meet Vancouver Criteria for being named in journal articles (otherwise their contribution can be acknowledged as appropriate).
External peer-review of outputs from colleagues	Peer-review processes have many limitations eg, busy reviewers may not have time to focus on details or may not have the relevant depth of topic expertise or relevant modelling expertise. But it can still be very helpful and BODE^3 sometimes sends draft articles to colleagues in the Department of Public Health.
Feedback from journal reviewers	Reviewers for journals will also provide a valuable additional check on the quality of the research outputs. Responding to journal reviewers is also an opportunity to: • Carefully re-read the whole manuscript and supporting documentation. • If model re-runs for additional analyses are suggested by reviewers, then this might also be a time for a quick PubMed/Google Scholar search for systematic reviews or new trials around key parameters. This can then maximise value of the final outputs. • At the final stage of journal article production, it is optimal to read the hard copy of proofs as this improves error identification.
Stakeholder feedback	An important quality check is to establish if the modelling results produced were intelligible to relevant policy-makers, were considered by them, and if they were actually used in a decision-making process that resulted in an intervention or other change.
Impact assessment	The SRO should check annually for citations of their published modelling work (eg, in Google Scholar) and inform co-authors of any issues arising. This can help capture subsequent uses of the modelling in government policy papers. It can also identify published critiques of that type of modelling work. Citation levels can give some approximate indication of the international usefulness of the modelling work.
When past modelling work becomes outdated	Scientific knowledge and technical progress is rapidly advancing, especially in the health sciences. The modelling team therefore needs to keep up-to-date with topic areas and with key guidance (eg, see the ISPOR guidance in Appendix 1). If this new knowledge suggests the need for revisions to past BODE^3 modelling work – then this should be considered (resources permitting). If not, then it may be desirable to: (i) write a letter to a journal or publish a blog that explains how our past work might now be partly out-dated; (ii) communicate directly with any NZ policy-makers that may have been using past BODE^3 work for decision-making. Nevertheless, all policy-makers should be regularly made aware of the rapid speed of progress in scientific knowledge and how published results can become outdated, even within months.

Calibration

Calibration has been described as ensuring that "inputs and outputs are consistent with available data". This can be achieved by adjusting model structure and input parameters.[4 5] Below is a specific example of how calibration can be documented for the BODE^3 MSLT DIET/PA Model (a multi-application model, for studying preventive interventions):

> "We conducted a number of tests on a MSLT Model (upon which the Tobacco Model and DIET/PA model is based on), to ensure that disease estimates corresponded with external data sources, with a focus on the main disease drivers in the model (ie, CHD

> and stroke). Specifically, we verified our model's estimated mortality rates for 2011 for four different non-Māori age groups by sex for CHD and stroke compared to those of the Ministry of Health Burden of Disease Study (updated from 2006 to 2011 estimates) and HealthTracker. We also verified our model's estimate of the proportion of deaths in New Zealand due to CHD and stroke in 2011 for six to eight different age groups by sex with those of the Ministry of Health's Mortality and Demographic data for 2010 (http://www.health.govt.nz/publication/mortality-and-demographic-data-2010).
>
> For CHD mortality rates, trends were very similar between our model estimates, the NZBDS and HealthTracker data for both males and females and across ages. For stroke mortality rates, the NZBDS had lower estimates than HealthTracker and our model estimates at younger ages. Our model estimates were slightly below other sources for older ages. For stroke deaths as a percentage of total deaths, model estimates for both females and males were within 1% of Ministry of Health estimates. Last, for CHD deaths as a percentage of total deaths, model estimates were within 3% or 4% of Ministry of Health estimates for females and males, respectively, where model estimates were higher."

More specifically, the following calibration options will typically be used in BODE^3 modelling work:

- Outputs from the model in 2011 (base-year) concur with external data (which may sometimes be used, in part at least, for inputs):
  - Disease incidence rates
  - Prevalence rates
  - Case-fatality rates
  - Mortality rates
  - Morbidity rates
- Outputs (see list above) from the model in 2021 and 2031 and 2051 concur with expected trends (or are at least plausible).
- Total costs and total QALYs in 2011 (and trends to 2021, 2031 and 2051) concur between BODE^3 MSLT models in the baseline.

Validation

This section is organised using the headings from an International Society for Pharmacoeconomics and Outcomes Research (ISPOR) *Good Practices in Modelling Task Force* consensus paper[4]: face validity, verification (or internal validity), cross validity, external validity, and predictive validity." Examples from past BODE3 work are used to illustrate the approach taken.

Face Validity

"Face validity is the extent to which a model, its assumptions, and applications correspond to current science and evidence, as judged by people who have expertise in the problem."[4]

In BODE^3 we strive to achieve face validity and one way that this is achieved is when research outputs are peer-reviewed prior to publication (eg, presented to research colleagues and peer-reviewed by journal reviewers). Furthermore, the models are often related to models which

have been published elsewhere (eg, work by Gartner et al on a tobacco forecasting model[6] which was related the BODE^3 one; and work by Cobiac et al on a sodium reduction model[7] that was related to the BODE^3 one[8]). Furthermore, the BODE^3 DIET/PA Model follows the form and structure of a MSLT, and more specifically the ACE-Prevention models[9-18] (including dietary and PA models) and the BODE^3 Tobacco Model.[19 20] These all lend some face validity.

Prospectively, the conceptualisation of models is tested with colleagues. For example, the DIET/PA model conceptual structure is used to orientate which variables are linked to one another.

Verification (or Internal Validity)

"Verification addresses whether the model's parts behave as intended and the model has been implemented correctly."[4]

A regular process of verification is used in building, modifying and extending the BODE^3 MSLT models, namely:

- All model changes are undertaken by the appropriate team member (see Table 1), checked and signed off by a second team member, and signed off by one of the BODE^3 Programme Directors (AB). This process accords with a Accountability for Quality Assurance process outlined by UK Department of Energy and Climate Change in their guidance for quality assurance of Excel-based models.[2] The model checking should be done before the final results used for a research publication are ready to be produced.
- All model modifications and extensions are 'logged' in a 'readme' tab in the model.
- Beneath this high-level quality assurance process, the following checks are routinely implemented:
    - A second team member – independently – randomly checks formulas and links in models.
    - A second team member – independently – works through each process from beginning to end (eg, risk factor A distribution, merged with risk factor A relative risks, to population impact fractions and their connection with disease incidence, then all-cause mortality, etc). (See also Table 2 for additional checking details.)
- In addition to routine scenario analyses and sensitivity analyses, an additional series of (extreme) sensitivity analyses are undertaken to logic (stress) test the model. For example, trends in disease incidence rates are turned off, and compared against expectations. And for stress testing, selected input parameters are changed to extreme values (eg, turning disease incidences to zero, one by one) to ensure changes in model outputs are consistent with theoretical expectation. Given the potential relevance of epidemiological and clinical knowledge on occasions – such results can be discussed with the wider team. (See Table 2 for additional checking details.)

Cross Validity

"Cross-validation involves comparing a model with others and determining the extent to which they calculate similar results."[4]

Model comparisons within the BODE^3 Programme have occurred, and are proposed with other international groups (for during 2017 and beyond). For example, within the BODE^3 programme identical dietary salt reduction interventions were run through a MSLT Model and a CVD Model built in TreeAge that had previously been developed by BODE^3.[8 21] That is, when an intervention generating a decrease in sodium of 22.8 mmol/day was run through both models, the overall QALYs gained were 110,000 in the TreeAge Model and 103,000 in the DIET model (3% discounting). As there are a number of differences between the models, generating results within 20% of each other was regarded as satisfactory, and the difference seen was closer to 5%. From our investigations it seemed that the differences seen between the two models were due to a combination of different baseline incidence rates, baseline case fatality rates and differing disability rates/weights between the two models. We therefore concluded that there was "reasonable cross-validity" – recognising some shared assumptions, and shared underlying data.

Model comparisons are also underway with the Nuffield Department of Population Health, Oxford University (Adam Briggs, Peter Scarborough and colleagues) who are working on similar types of models with similar food taxes and subsidy interventions (eg, publications involve Briggs and colleagues[22-24]). Other planned comparisons include:

- For the BODE^3 CVD Model with a microsimulation CVD model being developed by colleagues at Melbourne University (Prof Philip Clarke et al).
- For the BODE^3 E-cigarette Model with a planned Australian version of this model (University of Queensland).

As a general approach, our future cross-validation studies will (at a minimum) include 'stripping back' to the same population demography and epidemiology to allow a head-to-head comparison of any differences in model structure, then sequential addition of varying population epidemiology (eg, disease incidence rates, case fatality and trends), and population demography (eg, varying age structures). Nevertheless, our experience is that caution is also required with model comparisons – in that some models can have significant limitations in design and that there are notable differences in disease epidemiology (relative to New Zealand).

External Validity

"In external validation, a model is used to simulate a real scenario, such as a clinical trial, and the predicted outcomes are compared with the real world ones."[4]

Randomised trials through to disease incidence for the interventions proposed to be modelled with the BODE^3 DIET/PA Model are rare. We will consider the relevance of one of these for such validation work: a major sodium reduction trial on health outcomes,[25] but we note this might not prove to be worthwhile given the decline in CVD incidence over the 20 years of this trial.

Meta-analyses of trials (where available) are however used for parameterising intervention effect sizes in the model. 'Natural experiments' – as they accrue (eg, Danish food taxes[24 26] and Mexican SSB taxes[27]) – will also provide comparison points.

The BODE^3 CVD model, though, is more able to be compared to trials, eg, the simulated changes in mortality rates from pharmaceutical treatments compared to trials.

Calibration is sometimes positioned under external validation; we use calibration extensively in BODE^3 – please see the relevant section above.

Predictive Validity

"Predictive validity involves using a model to forecast events and, after sometime, comparing the forecasted outcomes with the actual ones."[4]

In a predictive validation exercise, a model's predicted outcomes are stored and then compared to the outcomes of processes happening in the real world as they unfold. As such, the latter type of validation exercise is typically only feasible for models that predict short-term outcomes.[4] As such, these types of exercises are not commonly performed within the BODE^3 models, which typically project long-term health and cost outcomes. In addition, for most BODE^3 modelling work it is usually not possible to compare forecast incidence and mortality rates in New Zealand from various modelled *interventions* with real world outcomes as the interventions are not being introduced in "real world New Zealand". An exception here is annual increases in tobacco tax for the next few years (with a possible extension after that). Also for some CVD treatment interventions, it is moderately plausible that some real world trends could be compared with model outputs (but this possibility needs further consideration).

Sharing of BODE^3 Models with External Parties: Quality Processes

Model sharing is underway (e-cigarette model with the University of Queensland) and is planned (colorectal cancer screening model with the University of New South Wales). BODE^3 staff will consider the following issues before engaging in further model sharing.

Advantages of model sharing are:

- The process is likely to involve additional checking (by external parties) of the BODE^3 model documentation and technical aspects of the model.
- Additional research outputs can be produced and so better inform policy-makers.
- Comparing model results between populations and countries (although dependent) can assist with cross validation (see above).
- Greater policy reach and influencing of informed decision-making may be achieved.

Disadvantages of model sharing are:

- Time involved for BODE^3 staff (eg, training) and around email requests and meetings.
- Risks of suboptimal quality processes and errors in results if the other team does not have adequate experience with modelling or with the BODE^3 processes around quality assurance.

Suggested BODE^3 quality control steps for model sharing where the receiving party is substantially amending or developing the model:

1. Try to work with groups with adequate modelling experience (published model outputs) and with appropriate staff dedicated to the project.
2. To aid model comparisons, it is worth carefully identifying parameter decisions that were made on the basis of the NZ population (and hence would typically need to be

changed for a different country), and decisions made based on international epidemiological evidence. (Eg, in the DIET/PA model, the size/number of categories of nutrition risk factors were chosen on the basis of NZ dietary distributions so this is something that might need to be changed if the model was to be used somewhere with different intakes.)

3. Checking of the proposed modelling study aims and key assumptions (by the Designated Liaison Person [DLP] from the BODE^3 Team). If evidence of poor conceptualisation – then consider reviewing/abandoning BODE^3 involvement. To aid model checking by the DLP, it is desirable that the research group that uses the model keeps track of changes made to the model in the logbook on the "readme" tab.
4. Checking of initial model results (with large numbers of scenario analyses) –by DLP. If there is evidence of poor quality control processes by the other team, then there is a need to review BODE^3 involvement and permissions to use the BODE^3 model.
5. DLP and/or other BODE^3 analyst spends at least a full day on random checks of the modified components of the model and its associated documentation. If evidence of poor quality control – then there is a need to review BODE^3 involvement.
6. All BODE^3 co-authors involved conduct checks and sign-off on the final manuscript. If any still have doubts around quality, then they should first raise this with the DLP and a BODE^3 Team review is required.

There are likely to be instances where collaborators are 'just' using the same model structure and changing a fairly limited number of input variables (eg, a different effect size for a counselling intervention in a different context, with slightly altered costs). The above quality control steps can be simplified in these circumstances.

Adapting Other Models for use by BODE^3: Quality Processes

BODE^3 has successfully modified an Australian tobacco forecasting model[6] for use in the New Zealand setting (eg, see:[28]). It also used many of the design features of a sodium reduction model[7] for a BODE^3 model around dietary sodium reduction.[8] These adaptive processes were efficient ways to speed model development for NZ. If other models are used in the future, then the general BODE^3 approaches to QA will be applied, after an initial rigorous interrogation of the supplied model and its associated documentation and publications. Where possible there will be on-going communication with the original model suppliers. We note that the DECC QA Document has a useful section on "Third Party Engagement" and "in-housing of external models".

Appendix 1: Key Documents of Relevance to Modelling & Quality Processes

The documents listed in the table below are the key ones that BODE^3 staff involved in modelling should be familiar with – albeit with some being appropriate for only specific types of modelling work. All new staff involved in modelling should read these documents, in conjunction with this current document. Key details from some of these documents are discussed elsewhere in this document.

Table A1: Key documents of relevance to modelling and quality processes in BODE^3

Topic area	Title and selected comments	Reference
BODE^3 Protocol (and related BODE^3 publications addressing methods)	The BODE^3 Protocol (version available on the BODE^3 website). ***Comment:*** Use of this Protocol helps insure standardised approaches – which facilitates comparisons between outputs and league table production. When this Protocol is next updated (probably in 2017) – it will be made completely consistent with this document on Quality Issues. An additional strength of BODE^3 approaches is careful consideration of equity (particularly Māori vs non-Māori). In this regard we frequently perform what we describe as an "equity analysis" as described in one of our publications.[29] Indeed, other aspects of heterogeneity are often dealt with by BODE^3 in particular detail (eg, this study of Herceptin[30]). Another somewhat special feature of BODE^3 is also its use of disability weights from the GBD.[31] At some point BODE^3 will be using the updated values from the more recent GBD work. Relevant Technical Appendices for each model will also be upgraded in an ongoing basis.	Blakely et al 2012[32]
The UK DECC Quality Assurance (QA) Document	"Quality Assurance: Guidance for Models". ***Comment:*** This document has a large amount of valuable information of relevance to BODE^3 workers. Please see specific points in the next section.	DECC 2015[2]
Guidelines on CEAs in health	"Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine" ***Comment:*** BODE^3 produces analyses designed to assist policy-makers; but it also examines underlying knowledge and academic issues. That said, much of this document by Sanders et al is consistent with current BODE^3 practices – in terms of the "health care sector perspective". But we consider the case for all CEAs/CAUs to also involve a "societal perspective" in additional to a "health care sector perspective" to be over-demanding in the NZ context. This is because decision-making in the NZ health sector is still very focused on just health (and health inequalities) and to a much lesser extent the wider societal impacts. Nevertheless, BODE^3 is currently working on including productivity costs and greenhouse gas impacts associated with particular interventions. BODE^3 staff will during 2017 discuss the following points arising in these guidelines: • That in addition to our focus on academic-orientated advances in methods and knowledge, we give consideration to describing in journal article outputs how policy recommendations might change with a range of	Sanders et al 2016[34]

Topic area	Title and selected comments	Reference
	different thresholds (for cost-effectiveness). • That aspects of the "Reporting Checklist for Cost-effectiveness Analyses" (Figure 2) be adopted for routine use. • That the recommended "structured abstract" be considered where possible for outputs focused on specific interventions (with this merged into requirements as specified by particular journals or if the structured abstract in the CHEERS Guidelines are required[33]). • We note the point about how assessments of health states might be influenced by respondents considering also productivity effects. As such we will attempt to keep up-to-date with any research on this issue. Similarly for new literature on "family spill-over effects" of health states.	
Reporting standards for health economic evaluations	"Consolidated Health Economic Evaluation Reporting Standards (CHEERS)" ***Comment:*** Much of this document is relevant to BODE^3 outputs, indeed some journals have requirements for these particular standards to be followed. It provides a 24-item checklist. The authors state that "the author team plans to review the checklist for an update in 5 years." But we note various routine BODE^3 extensions to reporting as per the "equity analysis".	Husereau et al 2013[33]
ISPOR Guidelines		
Model transparency and validation	"Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—7" ***Comment:*** This work by Eddy et al provides a valuable framework and detail. It has informed BODE^3 work, eg, see elsewhere in this current BODE^3 quality document around "validation". Some thoughtful comment and critique of the ISPOR guidance is provided by Vemer et al[35] (eg, it covers such issues as "validation as a continuous process", the concept of "valid enough", and double coding).	Eddy et al 2012[36]
Parameter estimation and uncertainty	"Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—6". ***Comment:*** The approaches used by BODE^3 to date are largely consistent with this guidance by Briggs et al. The suggested terminology around uncertainty is also in current use in BODE^3. A difference is that BODE^3 has not yet presented "Expected value of perfect information" (which may be used in work for MBIE by BODE^3).	Briggs et al 2012[37]
Dynamic transmission modelling	"Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—5". ***Comment:*** This guidance is of limited relevance at present to BODE^3 given the current focus on multi-state life-table MSLT modelling. But is included here just in case there is a future change in direction and such an approach is required.	Pitman et al 2012[38]
DES modelling	"Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—4". ***Comment:*** This guidance is of limited relevance at present to BODE^3 given the current focus on MSLT modelling (though BODE^3 has previous used DES modelling).	Karnon et al 2012[39]

Topic area	Title and selected comments	Reference
State-transition modelling	"State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—3". ***Comment:*** The guidance around "Markov model cohort simulation" is already standard practice for BODE^3.	Siebert et al 2012[40]
Model conceptualisation	"Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—2". ***Comment:*** This guidance is already standard practice for BODE^3.	Roberts et al 2012[41]
Overview of good modelling practices	"Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—1" ***Comment:*** See also thoughtful comments by a model user: Berger 2012[42]	Caro et al 2012[43]
Other		
Popular work with critiques of modelling	"Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy". ***Comment:*** This work critiques non-transparent models with algorithms that potential reproduce and even exacerbate social inequalities and injustices. The BODE^3 Team is aware of such issues (eg, see discussion of "equity analysis") and aims to maximise transparency (see elsewhere in this document).	O'Neil 2016[44]

Appendix 2: Quality Assurance Guidance for Modelling: Useful Specifics from the UK's DECC Approach

In 2015 the UK's Department of Energy and Climate Change (DECC) published a report "Quality Assurance: Guidance for Models".[2] While intended for its internal use, this document is of substantial potential value to informing QA processes for other types of modelling work. A particular benefit for BODE^3 is its focus on models built in Excel (though they also provide links to other documents eg, QA for system dynamics models). In general, however, this DECC guide is designed for a very high level of QA – which is appropriate for a government agency working in an area that can attract intense scrutiny (ie, climate change) and which is subject to auditing. The table below highlights particular areas for consideration by BODE^3 modellers, though all such BODE^3 workers should be familiar with the contents of the whole DECC document.

Table A2: Aspects of the UK's DECC QA Document[2] of Potential Relevance to Aspects of QA in the BODE^3 Programme

Topic area	Selected comments (especially around relevance to the BODE^3 Programme)	Starting page number/s
Level of QA	The DECC QA Document does highlight the relevance of different levels of QA to apply to models. Eg, the "proportionality to the business criticality of the model is also necessary when considering the level of Quality Assurance to apply." This is relevant to BODE^3 in that if a model is being used to inform a major government decision, then QA processes will need to be particularly high (and potentially similar to high level DECC ones).	P7 in the DECC QA Document (and trade-off issues on p9)
Designated roles and accountability	Although there is existing role definition with BODE^3 around major models, we will consider adopting the formal DECC terminology for such large modelling projects (see elsewhere in this document). That is: "Senior Responsible Owner (SRO)". This could overlap with the "Senior Analyst (SA)" role. The "Approving Body (AB)" for a modelling project – can be an individual or a group. Ideally for BODE^3, the AB will comprise at least 2 senior staff, with both signing off on key aspects of model design and model research outputs.	P7-8
QA and the model cycle	The DECC suggests that QA must be embedded throughout the model cycle. There is a useful figure and checklist that has relevance to BODE^3. However, BODE^3 would typically have less emphasis around "customers and other stakeholders" given the academic nature of BODE^3 work (and the flexibility to explore related issues of wider academic value). There is reference to several potentially useful documents (eg, "Best Practice in Spreadsheet Modelling" – which BODE^3 plans to obtain from DECC). Many of the QA activities referred to are already used by BODE^3, albeit with different terminology. Eg, routine documenting of model changes in the "Readme" tab of BODE^3 models.	P11 onwards (checklist on p12).
Version control	This topic is particularly relevant to BODE^3 eg, the BODE^3 MSLT CVD Model is an adapted version of the BODE^3 MSLT Tobacco Model. Similarly, for the BODE^3 MSLT E-cigarette model. When upgrades occur for the design (or parameters) of one of these models then it may be appropriate for these upgrades to occur amongst derivative models.	P18
Formal QA scoring	The DECC suggests quality assuring an existing model by completing a QA log "with the aim that the model achieves a score of 90% or above according to the weighted criteria set out in the QA log". This more formal process may be evaluated for BODE^3 modelling during 2017 if model outputs are likely to influence a major government decision.	P19
Documentation to the level that it allows for model hand-over	The DECC states that "Good documentation allows the developer to keep track of all the QA procedures that have to be carried out. Secondly, it is also highly likely that at some point a model that one develops will be handed over to either another analyst or onto a non-technical customer and it is important that the documentation exists to allow the required knowledge transfer." In BODE^3 we will continue to strive to have a level of documentation that allows for such handover (eg, to Australian colleagues) of major models. The Appendix (p51) has a good table for documentation – that could be tested out on a BODE^3 model.	P22 (see also p29)
Scenario and sensitivity analyses	BODE^3 has a strong track record of conducting many such analyses. Nevertheless, not all such analyses have been published. Therefore BODE^3 will consider putting more such analyses into "Online Supporting Material" (associated with journal articles) or BODE^3 "Online Technical Reports".	P23
Levels of	A useful chart on differing levels of review is provided. For BODE^3 all	P26

Topic area	Selected comments (especially around relevance to the BODE^3 Programme)	Starting page number/s
review	the internal review processes apply routinely. External peer review typically comes with the submission of journal articles. It would be rare for BODE^3 to commission "external peer review", but this could be done for projects that may inform major government decisions.	
QA within time constraints	A useful table is provided on "An overview of what QA should be performed within a variety of time constraints". Consideration will be given to following this approach in BODE^3 the next time a major model is ready for such a checking/internal review process.	P27
Allowing for staff changes	The DECC QA Document recommends ensuring "that a number of staff are sufficiently well versed in a model to cope with any sudden departures or absence of key staff." This is a goal for BODE^3 to keep achieving and we are giving consideration to assigning a deputy senior responsible owner (dSRO) role for major models.	P29
Technical specifics	There are useful specific details that BODE^3 modellers should all be familiar with. Some examples follow: • "If formulae change midway through an array, row or column are there comments to state this has happened and why?" A supplementary approach is to indicate such changes with colour-coding. • "It is essential that formulae are clear to users. Formula clarity can be enhanced using "alt+enter" to break long equations over multiple lines." • "Hardcoded values within formulae must be avoided as they are difficult to understand and often introduce errors." • "You can check the Name Manager (Ctrl+F3) to ensure there are no corrupted names; external ranged names should not be used unless it is absolutely necessary, and the names should be meaningful and follow a common convention."	P36
Formulae in a model	It is desirable that all formulae in a model are reviewed. BODE^3 will do more work to consider the appropriate level that this is to be achieved for major models. (Of note is that in BODE^3 a full list of all the non-trivial equations used in the model will be assembled in the Technical Documentation).	P37
Debugging software	Debugging software can help spot potential errors. BODE^3 will give further consideration to how to routinise the use of such software.	P37
Auto-checks, Error trapping & Regression Testing	All these approaches can be considered by BODE^3 modellers as appropriate. BODE^3 will give further consideration to the extent that they should be formally routinised.	P41
Validation	The DECC QA Document has some useful points – but the framework from ISPOR is preferred for BODE^3 (see elsewhere in this document).	P42
Extreme values testing	While such testing is routine in BODE^3 there may be scope for enhanced documentation around both extreme values testing and "model breaking" tests.	P44
Re-performance testing	This involves "implementing the model methodology in a completely new model. This may seek to completely replicate the functionality of the original in a shadow model, or may be a simpler calculation of key transformations from the original." This will only be considered in BODE^3 for a model which is to inform a major government decision.	P44

Appendix 3: Additional Technical Notes

Quality checking of datasets with SAS macros

BODE^3 has obtained a set of 31 macros in SAS developed by the University of Manitoba and the lnstitute for Clinical Evaluative Sciences (ICES). These can be used on examining the quality of data sets. They identify: missing values, trends in missing values (using regression analyses); and outliers/extreme values. An automatic data dictionary is also produced. June Atkinson will be making a standard list of these macros for use with HealthTracker and IDI datasets.

References

1. Perrow C. *Normal Accidents: Living with High Risk Technologies*. Princeton: Princeton University Press, 1999.
2. DECC Modelling Integrity Team. Quality Assurance: Guidance for Models. London: Department of Energy and Climate Change, UK, 2015.
3. Blakely T, Atkinson J, Kvizhinadze G, Nghiem N, McLeod H, Davies A, et al. Updated New Zealand health system cost estimates from health events by sex, age and proximity to death: further improvements in the age of 'big data'. *N Z Med J* 2015;128(1422):13-23.
4. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Health* 2012;15:843-50.
5. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health* 2003;6:9-17.
6. Gartner CE, Barendregt JJ, Hall WD. Predicting the future prevalence of cigarette smoking in Australia: how low can we go and by when? *Tob Control* 2009;18:183-9.
7. Cobiac LJ, Vos T, Veerman JL. Cost-effectiveness of interventions to reduce dietary salt intake. *Heart* 2010;96:1920-5.
8. Nghiem N, Blakely T, Cobiac LJ, Pearson AL, Wilson N. Health and economic impacts of eight different dietary salt reduction interventions. *PLoS One* 2015;10:e0123915.
9. Vos T, Carter R, Barendregt J, Mihalopoulis C, Veerman L, Magnus A, et al. Assessing Cost-Effectiveness in the Prevention (Ace-Prevention): Final Report. Brisbane and Melbourne: University of Queensland and Deakin University, 2010.
10. Bertram M, Lim S, Barendregt J, Vos T. Assessing the cost-effectiveness of drug and lifestyle intervention following opportunistic screening for pre-diabetes in primary care. *Diabetologia* 2010;53:875-81.
11. Byrnes J, Cobiac L, Doran C, Vos T, Shakeshaft A. Cost-effectiveness of volumetric alcohol taxation in Australia. *Med J Austr* 2010;192:439-43.

12. Cobiac L, Magnus A, Barendregt J, Carter R, Vos T. Improving the cost-effectiveness of cardiovascular disease prevention in Australia: a modelling study. *BMC Public Health* 2012;12:398.
13. Cobiac L, Vos T, Doran C, Wallace A. Cost-effectiveness of interventions to prevent alcohol-related disease and injury in Australia. *Addiction* 2009;104:1646-55.
14. Cobiac LJ, Vos T, Barendregt JJ. Cost-Effectiveness of Interventions to Promote Physical Activity: A Modelling Study. *PLoS Med* 2009;6:e1000110.
15. Cobiac LJ, Vos T, Veerman JL. Cost-effectiveness of interventions to reduce dietary salt intake. *Heart* 2010;96:1920-25.
16. Cobiac LJ, Vos T, Veerman L. Cost-effectiveness of Weight Watchers and the Lighten Up to a Healthy Lifestyle program. *Aust N Z J Public Health* 2010;34:240-7.
17. van den Berg M, Smit F, Vos T, van Baal PH. Cost-effectiveness of opportunistic screening and minimal contact psychotherapy to prevent depression in primary care patients. *PLoS One* 2011;6:e22884.
18. Veerman JL, Barendregt JJ, Forster M, Vos T. Cost-Effectiveness of Pharmacotherapy to Reduce Obesity. *PLoS One* 2011;6:e26051.
19. Blakely T, Cobiac LJ, Cleghorn CL, Pearson AL, van der Deen FS, Kvizhinadze G, et al. Health, health inequality, and cost impacts of annual increases in tobacco tax: Multistate life table modeling in New Zealand. *PLoS Med* 2015;12:e1001856. [Correction at: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.2211].
20. Pearson AL, Cleghorn CL, van der Deen FS, Cobiac LJ, Kvizhinadze G, Nghiem N, et al. Tobacco retail outlet restrictions: health and cost impacts from multistate life-table modelling in a national population. *Tob Control* 2016;(E-publication 22 September).
21. Nghiem N, Blakely T, Cobiac LJ, Cleghorn CL, Wilson N. The health gains and cost savings of dietary salt reduction interventions, with equity and age distributional aspects. *BMC Public Health* 2016;16:423.
22. Briggs AD, Mytton OT, Kehlbacher A, Tiffin R, Rayner M, Scarborough P. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. *BMJ* 2013;347:f6189.
23. Briggs ADM, Wolstenholme J, Blakely T, Scarborough P. Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. *Popul Health Metr* 2016;14.
24. Smed S, Scarborough P, Rayner M, Jensen J. The effects of the Danish saturated fat tax on food and nutrient intake and modelled health outcomes: an econometric and comparative risk assessment evaluation. *Eur J Clin Nutr* 2016.
25. Cook NR, Appel LJ, Whelton PK. Sodium Intake and All-Cause Mortality Over 20 Years in the Trials of Hypertension Prevention. *J Am Coll Cardiol* 2016;68:1609-17.
26. Jensen JD, Smed S. The Danish tax on saturated fat – Short run effects on consumption, substitution patterns and consumer prices of fats. *Food Policy* 2013;42:18-31.
27. Colchero MA, Popkin BM, Rivera JA, Ng SW. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. *BMJ* 2016;352:h6704.
28. van der Deen FS, Wilson N, Blakely T. A continuation of 10% annual tobacco tax increases until 2020: Modelling results for smoking prevalence by sex and ethnicity. *N Z Med J* 2016;129(1441):94-7.
29. McLeod M, Blakely T, Kvizhinadze G, Harris R. Why equal treatment is not always equitable: the impact of existing ethnic health inequalities in cost-effectiveness modeling. *Popul Health Metr* 2014;12:15.

30. Leung W, Kvizhinadze G, Nair N, Blakely T. Adjuvant Trastuzumab in HER2-Positive Early Breast Cancer by Age and Hormone Receptor Status: A Cost-Utility Analysis. *PLoS Med* 2016;13:e1002067.
31. Salomon JA, Vos T, Hogan DR, Gagnon M, Naghavi M, Mokdad A, et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2129-43.
32. Blakely T, Foster R, Wilson N, BODE³ Team. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness (BODE3) Study Protocol. Version 2.1. Technical Report No.3.* Wellington: Department of Public Health, University of Otago, Wellington, December 2012. http://www.otago.ac.nz/wellington/otago042986.pdf, 2012.
33. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013;16:231-50.
34. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 2016;316:1093-103.
35. Vemer P, van Voom GA, Ramos IC, Krabbe PF, Al MJ, Feenstra TL. Improving model validation in health technology assessment: comments on guidelines of the ISPOR-SMDM modeling good research practices task force. *Value Health* 2013;16:1106-7.
36. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Health* 2012;15:843-50.
37. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. *Value Health* 2012;15:835-42.
38. Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, et al. Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--5. *Value Health* 2012;15:828-34.
39. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Moller J, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value Health* 2012;15:821-7.
40. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Value Health* 2012;15:812-20.
41. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2. *Value Health* 2012;15:804-11.
42. Berger ML. The view of a model user on the ISPOR-SMDM modeling good research task force report. *Value Health* 2012;15:794-5.
43. Caro JJ, Briggs AH, Siebert U, Kuntz KM, Force I-SMGRPT. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health* 2012;15:796-803.
44. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.