



Canadian Journal of Health Technologies

March 2024 Volume 4 Issue 3

CADTH Health Technology Review

The Development of a Model Validation Tool to Assist in the Conduct of Economic Evaluations

Doug Coyle

Alex Haines

Karen Lee



Authors and Contributors

Authors

Doug Coyle, PhD

Professor, School of Epidemiology and Public Health, University of Ottawa

Alex Haines, BSc, MSc

Manager, Health Economics, CADTH

Karen Lee, MA

Director, Health Economics, CADTH

Contributors

CADTH would like to thank the following individuals for their input during the roundtable discussion as well as their methodological support.

Kednapa Thavorn, PhD

Associate Professor, School of Epidemiology and Public Health, University of Ottawa
Ottawa, Ontario

Scott Klarenbach, PhD

Professor, Faculty of Medicine and Dentistry, Medicine Department, University of Alberta
Edmonton, Alberta

Eldon Spackman, PhD

Associate Professor, Community Health Sciences, University of Calgary
Calgary, Alberta

Lauren E. Cipriano, PhD

Associate Professor, Ivey Business School, University of Western Ontario
London, Ontario

Cody Black, MSc

Lead, Health Economics, CADTH

Mike Paulden, PhD

Associate Professor, School of Public Health, University of Alberta
Edmonton, Alberta

Petros Pechlivanoglou, PhD

Scientist, The Hospital for Sick Children



Toronto, Ontario

Reviewers

CADTH would like to thank the following who helped review the final report and noted suggestions that have been incorporated into the final tool.

Tania Conte, RN MSc

Lead, Health Economics, CADTH

L'Institut national d'excellence en santé et en services sociaux (INESSS) — pharmacoéconomie:

Patrick Dufort, MSc

Coordonnateur scientifique en pharmacoéconomie

Loïc Gaugain, MSc

Coordonnateur scientifique en pharmacoéconomie

Laurence Giroux, MSc

Coordonnatrice scientifique en pharmacoéconomie

Ludovick Larocque-Laplante, MSc

Coordonnateur scientifique en pharmacoéconomie

Thomas Mortier, PhD

Coordonnateur scientifique en pharmacoéconomie



Table of Contents

Key Messages	6
Introduction	6
Methods	7
Development of an Initial Framework	7
Roundtable of Expert Reviewers	8
Testing of the Framework	9
Finalizing the Framework and Tool Format	9
Content of the Model Validation Tool	9
Validation of the Conceptual Model.....	10
Computer Model Validation and Verification.....	17
General Issues of Concern.....	22
Conclusion	24
References	25
Appendix 1: Model Validation Tool	26

List of Tables

Table 1: Consequences of Omissions	10
Table 2: Decision Problem Items and Consequences (Items 1 to 5)	10
Table 3: Model Specification Items and Consequences (Items 6 and 7)	12
Table 4: Clinical Effectiveness Items and Consequences (Items 8 to 25)	12
Table 5: Model Transparency Items and Consequences (Items 26 to 28).....	17
Table 6: Assessment of Model Behaviour Items and Consequences (Items 29 to 46)	18
Table 7: Validation of Model Coding Items and Consequences (Items 47 and 48)	22
Table 8: General Issues of Concern and Consequences (Items 49 to 53)	23
Table 9: Validation of Decision Problem (Items 1 to 5).....	26
Table 10: Validation of Model Specification (Items 6 and 7)	26
Table 11: Validation of Clinical Effectiveness (Items 8 to 25)	27
Table 12: Validation of Model Transparency (Items 26 to 28)	29
Table 13: Validation of Assessment of Model Behaviour (Items 29 to 46)	29
Table 14: Validation of Model Coding (Items 47 and 48).....	31
Table 15: General Issues of Concern (Items 49 to 53)	31

Key Messages

- Model validation is critical to ensure that the results from an economic model address the needs of the decision-maker, measure what it intends to measure, and are valid. There is a need for a comprehensive and accessible tool to support both the conceptual and technical validation of models. Tools to assist in this important aspect are currently lacking.
- Through a targeted literature search, and in consultation with other health economists, CADTH developed a tool that will assist researchers and decision-makers to ensure economic models are fit for purpose and function properly.
- This model validation tool was considered robust and practical to implement through internal testing and in consultation with health economists external to CADTH.
- This tool summarizes the approach taken by CADTH in model validation. Although this is not a requirement for submissions to CADTH, the tool will help ensure a more consistent and reproducible approach to model validation for economic evaluations of all health technologies.

Introduction

The 2017 *Guidelines for the Economic Evaluation of Health Technologies: Canada*¹ produced by CADTH aimed to provide clear, concise, and practical guidance of a high standard to support the conduct of economic evaluations and to help meet the needs of decision-makers. The primary focus of the guidelines is to assist researchers to produce evaluations that provide unbiased estimates of the long-term costs and outcomes associated with alternative strategies.

A key component of building an economic evaluation is conducting a robust validation of the economic model. This ensures the researcher that their model is performing as intended while ensuring that any conclusions drawn from the model can be relied upon to make decisions. The Modelling section of the current CADTH guidelines addresses the appropriateness of models for facilitating decisions with specific guideline statements:¹

- 8.1 Model conceptualization and development should address the decision problem.
- 8.2 The model should be consistent with the current understanding of the clinical or care pathway for the health condition and the interventions being compared. The scope, structure, and assumptions should be clearly described and justified.
- 8.3 Researchers should consider any existing well-conducted and validated models that appropriately capture the clinical or care pathway for the condition of interest when conceptualizing their model.
- 8.4 The choice of modelling technique should be justified. The approach should be no more complex than is necessary to address the decision problem.

8.5 Baseline natural history should be representative of the target population considered in the decision problem.

8.6 Models should be subjected to rigorous internal validation. This process should involve quality assurance for all mathematical calculations and parameter estimates, and these processes and their results should be reported.

The Guidelines in Detail section focuses primarily on the reporting of the model structure and the generated results rather than a critical appraisal of the model itself. This is aligned with other published guidelines and checklists. While checklists for economic evaluations exist,^{2,4} their intent is not to provide the level of guidance or depth required in most cases to evaluate the technical aspects relating to model validity, transparency, and flexibility. The main reason is that usually the model, especially when built in TreeAge or Excel, is not published in the academic literature, and therefore technical validation of the model does not occur during the peer-review process. The focus of these checklists is on ensuring thorough reporting of the methods adopted to populate a model and the results produced by the model rather than on the technical validity of the model itself. More information is provided in the Methods section of this report from a targeted literature review looking at published checklists that specifically focus on model validity and why these still may not be fit for purpose.

This report will provide a description of the process by which the framework of the model validation tool was developed (e.g., items identified, validated, and tested) and a detailed description of the items included in the tool; the full tool is available in Appendix 1. Although the tool is intended to be used on its own, it is recommended that users read this report and refer to the specific sections for more detailed information. The purpose of this report is to assist in the development and validation of health economic models. The information presented should be seen as good practice guidance not requirements. The purpose is to help producers of economic models build robust and transparent models for decision-making and help identify potential errors that may arise. Building economic models can be a complex process, and this report does not cover all the possible permutations that could arise in model development. The copy of the full [model validation tool](#) is available on the website.

Methods

Development of an Initial Framework

An initial draft list of items relevant for inclusion in the proposed tool was developed through a 4-stage process of gathering information on current practices regarding model appraisal and validation.

First, a review of the grey literature was conducted in November 2020 to identify currently published checklists for model validation relating to health technology assessment agencies and other organizations with the responsibility of reviewing economic models to facilitate reimbursement decisions. The agencies reviewed included the Institute of Health Economics (IHE), Institut national d'excellence en santé et en services sociaux (INESSS), Institute for Quality and Efficiency in Healthcare (IQWiG), National Institute for

Health and Care Excellence (NICE), National Centre for Pharmacoeconomics (NCPE), the Pharmaceutical Benefits Advisory Committee (PBAC), and Scottish Medicines Consortium (SMC). From this review, 3 potential documents were identified and examined for relevance (from PBAC, INESSS, and NCPE).⁵⁻⁷ The documents from PBAC and INESSS related primarily to reporting of model structure and results; the NCPE document was relevant but only available as an abstract.

Second, a targeted review of the published literature was conducted to determine the availability of any published checklists either relating to professional societies (such as the International Society for Health Economics and Outcomes Research [ISPOR] and Society for Medical Decision Making [SMDM]) or published independently by academic authors with experience with model review and validation. The review was conducted through a web search of relevant academic societies' websites and through a targeted literature search through MEDLINE from 2005 onward. This was conducted in April 2021, and a range of potential documents were identified.⁸⁻¹¹ The identified papers gave useful insight into the important components required for a thorough validation of a model. Although these papers provided a thorough assessment of how specific models were validated, many of the publications were highly technical in nature and required intimate knowledge of how the model was built. A gap was identified for a more generalized tool that could be completed without the need for free-flowing text. Such a tool could provide an immediate overview of how robust a model is and, if judged robust, could lead to further, more detailed validation.

Third, consultations were conducted in March and April 2021 with 6 health economists who assist with reviews of economic evaluations for different organizations (SMC, NICE [England and Wales], the NCPE [Ireland], and the PBAC [Australia]) to determine their current process for model validation.

Finally, reviewers of pharmacoeconomic submissions for CADTH were sent a survey in May 2021 that included 11 questions relating to their own processes for model validation. The survey was informed by the 3 steps previously mentioned. Questions related to conceptual model validation (3 questions about the specification and application of a decision problem and the incorporation of clinical effectiveness) and computer model validation (8 questions about their own process of model validation, including error identification and specific features of models the reviewers felt limited transparency).

From this process, an initial list of topics to consider when validating models was developed. A draft list covered items related to conceptual model validation, computer model verification and validation, and general issues of concern.

Roundtable of Expert Reviewers

To ensure the list of items reflected the experience of Canadian health economists, a roundtable was held with experienced health economists (CADTH staff and health economists outside of CADTH). All members of the roundtable were provided with information on the project and the draft list of items to review in advance of the meeting. The members were asked to provide their feedback on the items as well as any potential missing items. The roundtable involved reviewing the draft framework section by section, with focus on assessing the viability of each component of the framework, revising the wording of the framework to ensure clarity if necessary, identifying any omissions within the framework, and determining how the

framework can be implemented within the CADTH process. The draft framework was developed from this process.

Testing of the Framework

As an additional validation check, the framework was trialled alongside several CADTH reviews for economic reviewers to note what aspects of the tool were not clear and/or transparent. These reviews did not alter the number of items considered for the tool but helped refine the language to ensure that the consequences of not checking items was clear to the reader. The tool was also sent to INESSS for comment and review. Based on feedback from INESSS, several aspects of the tool were also updated to provide additional context and clarity.

Finalizing the Framework and Tool Format

The model validation tool was finalized by consolidating feedback from multiple sources both internal and external to CADTH. Based on discussions from the roundtable, the most suitable format for a tool that researchers could easily use is a list of questions with binary choice answers (yes/no). However, other tools in this space, identified through the targeted literature searches, have used free text and/or the users of the tool would likely need to be involved in the development of the model.⁸⁻¹⁰ And although these other tools are robust for the purposes for which they are intended, completing them does not guarantee that the model is valid.

Given the variability of modelling techniques, no tool can be comprehensive enough to cover all the intricacies of how a model could be conducted. Therefore, this model validation tool encourages good modelling practices but does not necessarily determine good models. If a single item in the list is not present, then the validity of the model is in question. If all items are present, this increases the likelihood – but does not guarantee – the model is valid. It was felt a tool was needed that focused on the fundamental aspects of model appraisal and could indicate to the user whether the model was valid through a series of binary choice questions. Therefore, items that required free-flowing text were removed. The tool also should be used by someone who was not involved in the building of the model, as this will provide a separate validation, and confirm the transparency and ease of use of the model.

Content of the Model Validation Tool

The full tool can be found in [Appendix 1](#). The tool consists of 3 main sections:

- validation of the conceptual model
- computer model validation and verification
- general issues of concern.

Within each section of the tool are a series of statements. If the model adheres to the guidance, the user should select “yes.” If the guidance has not been adhered to, the user should select “no” and then refer to the relevant table to determine the consequences associated with not adhering to the guidance. There are

3 broad consequences associated with nonadherence to the tool, which are summarized in [Table 1](#). This information enables users to determine the risk of not adhering to the specified guidance. The magnitude of the risk will be dependent on the decision problem.

Table 1: Consequences of Omissions

Description	Risk
A potential error has been identified.	Unless this is corrected, the results of the model will be inaccurate and potentially misleading.
The model is not sufficiently transparent.	Models that lack transparency may be challenging to validate, therefore it cannot be ensured that the model is free from errors. The results from the model may also be difficult for decision-makers to interpret if it is not transparent how they were derived.
The model is not accurately specified.	All models should address the needs of the decision-maker. If the model is not correctly specified, the results may not be useful for decision-making however accurate they are.

Validation of the Conceptual Model

Validation of the conceptual model relates to the process by which a researcher determines whether the modelling framework is appropriate for the context of the review. This requires consideration of 3 components of the conceptual model:

- the decision problem
- the specification of the model
- the incorporation of clinical effectiveness information.

Decision Problem

As part of the validation of the model, it is necessary to ensure that the decision problem addressed by the model is consistent with the relevant decision problem from the perspective of the decision-maker.

Table 2: Decision Problem Items and Consequences (Items 1 to 5)

Item	Why it is assessed	Consequence if not checked
1	The model built is reflective of the population that the decision problem applies to. The model should reflect the same population that the decision-maker will be making decisions on.	The model is not accurately specified. If the model is not reflective of the population for which the decisions are being made, the results of the model may not be valid for decision-making.
2	The model can examine key subgroups within the population of interest. Key subgroups identified a priori by clinical experts can be explored in the model. This may be relevant to inform decisions.	The model is not accurately specified. Absence of subgroup analysis restricts the amount of information that can be obtained from the model. This may limit decision-makers' ability to make informed choices about subgroups.
3	The model assesses all comparators used to currently treat the stated population. The cost-effectiveness of a technology is contingent on what it is being compared to. To understand whether a technology is cost-effective, it is important to know what alternatives it could	The model is not accurately specified. Excluding relevant comparators from the model may prohibit an assessment of cost-effectiveness from being made. If a decision-maker needs to decide between 2 technologies, the additional benefit for

Item	Why it is assessed	Consequence if not checked
	displace and, therefore, what the additional costs and benefits are relative to these.	patients and the additional costs must be known. Without this information, decision-makers cannot make robust decisions.
4	<p>The model incorporates costs that are consistent with the specified perspective of the analysis.</p> <p>The costs included in an economic evaluation will be determined by the researcher through the choice of the perspective(s). Once this decision is made, the model should be consistent in analyzing the costs relevant to the stated perspective(s).</p>	<p>The model is not sufficiently transparent OR the model is not accurately specified.</p> <p>If costs assessed in the model do not align with what is stated, then there is a disconnect. Until this is resolved, results from the analysis may not be valid for decision-makers. For transparency, identification and reporting of relevant costs is a key aspect of validation .</p>
5	<p>The model assesses all outcomes deemed important by clinicians and patients.</p> <p>The model reflects all outcomes deemed important and relevant to capture all the potential benefits and harms associated with each considered technology.</p>	<p>The model is not accurately specified.</p> <p>Excluding relevant outcomes may bias the results. If an excluded outcome is deemed to occur equally among technologies, its exclusion may not have a noticeable impact on the conclusions and its exclusion may be justified. A thorough description of the relevant included and excluded outcomes is also required to ensure no double counting has occurred.</p>

Model Specification

Before appraising the technical aspects of the model, it is important to consider what would be an appropriate specification of the model. Researchers must consider the optimal specification of the model, which should be informed by our current understanding of the clinical pathway and validated by clinicians and patients familiar with the condition. Two distinct processes are recommended:

- a review of existing models in the area
- a formal consideration of the disease process and clinical pathway.

Table 3: Model Specification Items and Consequences (Items 6 and 7)

Item	Why it is assessed	Consequence if not checked
6	<p>The structure of the model (i.e., the process and clinical pathway) has been validated by clinical experts.</p> <p>A model must reflect what is known about the condition. This requires assumptions be made about how a disease progresses, and how this impacts patient outcomes and costs. Only those with experience and/or clinical expertise can validate these assumptions.</p>	<p>The model is not accurately specified.</p> <p>If the model has not been validated by a clinical expert and patients in the field, it is unknown whether the model structure and assumptions are appropriate. This restricts the applicability of the model's conclusions.</p>
7	<p>The model follows previous models in this clinical area or justification has been provided about why the model structure differs from previous models.</p> <p>In many cases, other models have been built in the same disease area. Understanding what models have come before and how these were appraised can help ensure the new analysis draws from learnings from the past.</p>	<p>The model is not accurately specified.</p> <p>If the model is different from previous models and no justification is provided, it is unclear why this change in model structure is warranted and may limit its applicability.</p> <p>[If the model is the first to be built in a particular disease area, NA should be selected. This does not mean the results are invalid but that additional care will be needed to ensure expert validation (as per item 6) was conducted.]</p>

Modelling of Clinical Effectiveness

How clinical effectiveness information is incorporated within a model will likely have the biggest impact on the results of the related analysis. Thus, researchers must consider how the following issues are considered within the model:

- the quality and consistency of evidence
- the assumed duration of effect
- the mechanisms by which benefit associated with the technology is generated
- the potential for double counting of benefit
- appropriate consideration of uncertainty.

If the output from the model does not match relevant data or expectations, this is due to either an error or insufficient transparency regarding why the model output is expected to deviate from the relevant data.

Table 4: Clinical Effectiveness Items and Consequences (Items 8 to 25)

Item	Why it is assessed	Consequence if not checked
8	<p>Time spent in each health state, for each technology assessed, can be extracted from the model.</p> <p>To validate the model output, time spent in each health state should be transparently stated.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model cannot transparently detail how much time is spent in each health state, this limits the usability of the model results. It is important to know what drives the model results to ensure the results can be validated as well as transparently communicated to a decision-maker.</p> <p>[If the model does not utilize health states, such as</p>

Item	Why it is assessed	Consequence if not checked
		a decision tree, NA should be selected and this item does not apply.]
9	<p>The model output matches the evidence provided to support time spent in health states among technologies.</p> <p>To assess external validity of the model, output from the model should be compared to a reliable external data source. This can be trial data, real-world evidence, or expert opinion.</p>	<p>The model is not sufficiently transparent OR a potential error has been identified.</p> <p>Using data to support findings from the model increases the robustness of the model output.</p> <p>If the output from the model does not match relevant data or expectations, there is either an error or insufficient transparency regarding why the model output is expected to deviate from the relevant data.</p> <p>[If NA was selected for item 8, this item should be skipped because it does not apply.]</p>
10	<p>If clinical events are modelled (i.e., hospitalizations, exacerbations, strokes, hip fractures), the number of events for each technology can be extracted from the model.</p> <p>To validate the model output, the number of clinical events that occur should be transparently stated.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model cannot transparently detail how many clinical events occur, this limits the usability of the model results. It is important to know what drives the model results to ensure the results can be validated as well as transparently communicated to a decision-maker.</p> <p>[If clinical events are irrelevant to the decision problem, NA should be selected because this item does not apply.]</p>
11	<p>Model output matches the evidence provided to support the number of clinical events across technologies.</p> <p>To assess external validity, the model output should be compared to a reliable data source. This can be trial data, real-world evidence, and expert opinion.</p>	<p>The model is not sufficiently transparent OR a potential error has been identified.</p> <p>Using data to support findings from the model increases the robustness of the model output.</p> <p>If the output from the model does not match relevant data or expectations, this is due to either an error or insufficient transparency regarding why the model output is expected to deviate from the relevant data.</p> <p>[If NA was selected for item 10, this item should be skipped because it does not apply.]</p>
12	<p>The impact of adverse events on health outcomes and costs can be extracted from the model.</p> <p>To validate the model output, the impact of adverse events should be transparently stated.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model cannot transparently detail the impact of adverse events, this limits the usability of the model results. It is important to know what drives the model results to ensure the results can be validated as well as transparently communicated to a decision-maker.</p> <p>[If adverse events are irrelevant to the decision problem, NA should be selected because this item does not apply.]</p>
13	<p>The model output matches evidence provided for adverse event type and frequency from the evidence.</p> <p>To assess the external validity of the model, output from the model should be compared to a reliable data source. This can be trial data, real-world evidence, and expert opinion.</p>	<p>The model is not sufficiently transparent OR a potential error has been identified.</p> <p>Using data to support findings from the model increases the robustness of the model output.</p> <p>If the output from the model does not match relevant data or expectations, this is due to either an error or</p>

Item	Why it is assessed	Consequence if not checked
		<p>insufficient transparency regarding why the model output is expected to deviate from the relevant data.</p> <p>[If NA was selected for item 12, this item should be skipped because it does not apply.]</p>
14	<p>Life-years are reported as a result within the model.</p> <p>To validate the model output, life-years should be transparently stated.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model cannot transparently detail how long patients live in the model, this limits the usability of the model results. It is important to know what drives the model results to ensure the results can be validated as well as transparently communicated to a decision-maker.</p>
15	<p>The impact the evaluated technologies has on mortality is clear</p> <p>Impact on mortality is a key component in influencing the results of a cost-effectiveness analysis. It should be clear whether there is an expectation that the proposed technology will influence mortality rates.</p>	<p>The model is not sufficiently transparent.</p> <p>It should be transparent whether a technology has the potential to influence mortality. If this is not known or it is not transparent, this should be resolved before a thorough validation of the model is undertaken.</p> <p>[If it is clear and transparent that a technology will not influence mortality, NA should be selected because this item does not apply.]</p>
16	<p>If differences in mortality are noted in Item 15, select the reasons for differing mortality in the model (more than 1 reason can be selected):</p> <p>It should also be transparent how a technology is expected to influence mortality. Based on the majority of model structures, this can be modelled in 4 ways:</p> <p>a) Duration of time spent in health states with higher mortality risk</p> <p>If a technology influences movement between health states with differing probabilities of death, this should be clear and transparent.</p> <p>b) There is a difference in the frequency of fatal clinical events</p> <p>If a technology reduces the rate of fatal clinical events (e.g., cardiac failure), it should be clear and transparent whether the technology influences the rate of these events.</p> <p>c) There is a difference in the frequency of fatal adverse events</p> <p>Rarely, a technology may have fatal adverse events. It should be transparent if these are included and, if so, what impact the technology has on them.</p> <p>d) Direct impact on risk of death, not stated previously, has been modelled (i.e., direct modelling of overall survival from the trial)</p> <p>Mortality can be modelled directly and not based on state occupancy in a model. For example, overall survival curves are frequently reported in oncology trials. Some models (such as partition survival models) directly model these curves.</p>	<p>The model is not sufficiently transparent.</p> <p>All the mechanisms by which each technology impacts mortality should be transparent in the model. If these mechanisms are not transparent in the model, the results of the analysis may be invalid.</p> <p>[If NA was selected for item 15, this item should be skipped because it does not apply.]</p>

Item	Why it is assessed	Consequence if not checked
17	<p>If there are mortality differences between technologies, the model can extract which of the reasons from item 16 has the largest impact on incremental life-years.</p> <p>The model should be transparently programmed such that the different mechanisms that influence mortality can be turned off. For example, if a technology reduces fatal events and reduces the chance of disease progression, it should be clear how each of those factors influence life-years separately. In many cases, evidence informing 1 aspect of mortality may be robust, and evidence for another aspect may be highly uncertain. The model should be programmed adequately to explore these impacts separately.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear which mechanisms influence overall survival, it cannot be determined whether the results of the model are valid.</p> <p>[If mortality is irrelevant to the decision problem, NA should have been selected for item 15 and this item should be skipped because it does not apply.]</p>
18	<p>Model output matches evidence regarding mortality rates between different technologies.</p> <p>To assess external validity, model output should be compared to a reliable external data source. This can be trial data, real-world evidence, and expert opinion.</p>	<p>The model is not sufficiently transparent OR a potential error has been identified.</p> <p>Using data to support findings from the model increases the robustness of the model output.</p> <p>If the output from the model does not match relevant data or expectations, this is due to either an error or insufficient transparency regarding why the model output is expected to deviate from the relevant data.</p> <p>[If mortality is irrelevant to the decision problem, NA should have been selected for item 15 and this item should be skipped because it does not apply.]</p>
19	<p>It is clear that the model does not utilize technology-specific utilities.</p> <p>One purpose of a model is to consider all known impacts of a technology on quality of life. This is best handled by modelling different probabilities of being in states combined with state-specific utility values. State-specific utilities that vary by technology are problematic because the reason for them occurring is not transparently provided. Therefore, it is uncertain whether a technology-specific utility is being double counted elsewhere in the model or is the result of a chance finding in the data. A technology may have a higher rate of adverse events and result in lower utility. If a model utilizes technology-specific utilities and models adverse events, the impact of adverse events will be double counted.</p>	<p>The model is not sufficiently transparent OR a potential error has been identified.</p> <p>If the model includes technology-specific utilities and does not have an option to exclude them, the results of the analysis may be invalid because it is unclear whether the modelled estimates of QALY gains with a technology are double counted.</p>
20	<p>Based on the results from the submitted model, it can be determined which of the following has the largest impact on cost-effectiveness conclusions: time spent in health states, number of clinical events occurring, adverse events, and mortality.</p> <p>A technology can generate additional QALYs in many ways. It is important to disaggregate the impact a technology has to be transparent about from where benefit is derived.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear how a technology generates additional QALYs, it is unclear whether the analysis is valid. Likewise, results will not be transparent for decision-makers to review.</p>
21	<p>The model distinguishes data that are based on extrapolation methods.</p> <p>Models require a sufficient time horizon to assess all the short- and long-term impacts a technology may have (often a patient's</p>	<p>The model is not sufficiently transparent.</p> <p>If a model does not delineate what data are based on extrapolation techniques, it is uncertain what output of the model can be validated against known data for the</p>

Item	Why it is assessed	Consequence if not checked
	lifetime). Extrapolation requires the model to make predictions about future health outcomes. It is important to delineate what aspects of the model can be verified by data vs. those which are based on predictions.	purpose of external validity. [If a model does not require extrapolation (i.e., a within-trial economic evaluation) then NA should be selected.]
22	<p>The model time horizon can be adjusted for just the period for which there are clinical data available.</p> <p>The period for which data are available represents the time period when inputs and outputs are most robust. Running the model for just the period data that are available allows a direct comparison to be made to the trial data.</p>	<p>The model is not sufficiently transparent.</p> <p>The model should be able to assess the impact of the technology at different time horizons. If the model cannot do this, it is challenging to validate when benefits are accrued in the model time horizon. Likewise, decision-makers may wish to know when the majority of benefits and costs are accrued.</p>
23	<p>If the model incorporates both direct and indirect effects, it is clear how double counting has been avoided.</p> <p>When assessing the benefit of a technology in a model, it is necessary to consider whether treatment effects are direct, indirect, or both. Direct effects are applied directly to a parameter of interest; for example, if assessing mortality, a hazard ratio can be applied to an overall survival curve. Indirect effects are those which indirectly impact the parameter of interest; for example, a technology may reduce heart attacks, which in turn impacts mortality. It is important that a model is clear about what direct and indirect effects are used so no double counting occurs.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear how double counting has been controlled for, the model results may be biased because the impact of a technology on a parameter of interest may be counted twice. For example, it may be noted in a trial that a drug reduces the risk of death by reporting a hazard ratio < 1 for overall survival and that the risk of hospitalization is reduced again with a hazard ratio < 1. A model may directly model overall survival and apply the overall survival hazard ratio to this. The model may also model a risk of death associated with hospitalizations. If the model also reduces the rate of hospitalizations, the impact of the drug on death will be double counted. To avoid double counting, the model will need to either only model death using overall survival curves or model all the individual mortality risks separately and ensure these align with overall survival.</p> <p>[If the model only models direct or indirect effects (not both), NA should be selected because this item does not apply.]</p>
24	<p>Does the modelled relationship between surrogate outcomes and final outcomes (quality of life and mortality) match the evidence presented?</p> <p>Surrogate outcomes are used in trials in place of outcomes that directly assess a patient's quality of life or mortality. It is important that the link between surrogate outcomes and quality or length of life is transparently described and programmed in the model to ensure the robustness of this link can be validated.</p>	<p>The model is not sufficiently transparent.</p> <p>If a model does not transparently detail how a surrogate outcome links to quality and/or length of life, the results of the analysis may be invalid.</p> <p>[If the model does not use surrogate outcomes, NA should be selected because this item does not apply.]</p>
25	<p>The model allows flexibility to explore waning of treatment effects.</p> <p>Over time, the efficacy of a technology can wane, meaning it may become less effective as time goes on. It is important that the impact of treatment waning be explored or justified in the analysis because the assumption of permanent and enduring treatment effects does not always hold.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model is not flexible to assess waning of treatment effect, and there is no justification to support enduring and permanent treatment effects, the results of the analysis may be biased in favour of the most effective technology.</p> <p>[If no extrapolation of a technology's effect is required, then NA should be selected because this item does</p>

Item	Why it is assessed	Consequence if not checked
		not apply. For example, data from the trial captures the full time horizon.]

Validation of the conceptual model should be conducted before the verification of the technical aspects of the model coding. An appropriate conceptualization of the model is required because a technically correct model may still be invalid if either the decision problem or the clinical pathway is misspecified.

Computer Model Validation and Verification

Within the computer model validation and verification section, the focus is on whether the model is technically correct in the process of determining outcomes of interest from the data input. This is separated into 2 distinct processes: assessment of model behaviour (black box testing) then the scrutinization of the coding of the model (white box testing). It is important that these are done in the correct order.

Model Transparency

Models should have specific features relating to transparency to facilitate both black box and white box testing. If the model is not sufficiently transparent, it may not be feasible to conduct the necessary testing, and the model cannot be considered as demonstrating validity. The following features have been identified:

Table 5: Model Transparency Items and Consequences (Items 26 to 28)

Item	Why it is assessed	Consequence if not checked
26	<p>The model can access the deterministic result and the results from single Monte Carlo simulations.</p> <p>A model can be run deterministically (assuming fixed parameters) or probabilistically (assuming variability in parameter inputs based on specified probability distributions). It is important that a model can assess results from the deterministic results and each run of the probabilistic analysis. This transparency improves validation checking.</p>	<p>The model is not sufficiently transparent.</p> <p>If a model cannot produce this output, the results cannot be robustly validated.</p>
27	<p>A clear trace can be identified that links all input parameters to final outcomes (i.e., only input parameters are hard coded).</p> <p>To ensure the model is free from error and doing what the researcher claims it is doing, a clear trace is important for validation checking. A clear trace means every input in the model can be traced to the final model output.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear how an input relates to the model output, the results cannot be robustly validated.</p>
28	<p>Macros are exclusively related to first- or second-order simulation and model navigation (exclusive to models built in Microsoft Excel).</p> <p>Macros are a useful tool for automating some functions of an Excel model. However, macros can obfuscate the link between model inputs and model outputs. Therefore, their use should be limited to functions that would otherwise be impossible or impractical without them, such as running the model probabilistically.</p>	<p>The model is not sufficiently transparent.</p> <p>If a model relies heavily on macros, clear user guides and transparent code should be used to navigate through them. Overreliance on macros may make creating a clear trace impractical (refer to item 27), therefore the model cannot be robustly validated.</p>

Assessment of Model Behaviour (Black Box Verification)

The first process of model validation involves ascertaining whether changing the inputs of the model leads to results which meet the general expectations of the researcher. This is referred to as black box testing because it does not require the researcher to know the inner workings of the model. If during black box testing the model fails to provide results that are explainable, the researcher should conduct detailed white box testing to determine why the results are not as expected. If the reviewer was not involved in building the model, the model should be returned to the original researcher for explanations about why invalid results are generated. The following is not an exhaustive list of black box testing that can be conducted but serves as a baseline of minimal tests that should be conducted.

Table 6: Assessment of Model Behaviour Items and Consequences (Items 29 to 46)

Item	Why it is assessed	Consequence if not checked
29	<p>You can set the effectiveness of different technologies such that QALY estimates are equal.</p> <p>Parameters that influence incremental QALYs between evaluated technologies should be transparently laid out within the model. Therefore, it should be simple to change these parameters to ensure all QALYs are equal. This task will ensure there are no unknown or unspecified effects of a technology influencing QALYs.</p>	<p>A potential error has been identified.</p> <p>If any test cannot be performed, the model programming may be inappropriately complex, and the results may be difficult to validate. It is recommended that the model be sent back to the original researcher for this functionality to be added.</p> <p>If the results from each test are not as expected, there is either an error in the code or an unspecified impact of the technology that may be invalid.</p> <p>If a specific test cannot be conducted, or it produces unexpected results, white box testing will be needed to resolve or identify the issue. If left unresolved, the model's results may be invalid.</p> <p>If the time horizon of the model is less than 1 year, discounting is not needed and NA should be selected for any tests involving changing discount rates in these scenarios (refer to items 39 and 40).</p>
30	<p>When you set effectiveness values to be extremely in favour of or against 1 technology, this leads to substantially greater or reduced QALY estimates.</p> <p>There should be a clear link between the effectiveness of each technology and QALYs. Substantially improving the efficacy of a single technology should have a large effect on the incremental QALYs relative to the base case.</p>	
31	<p>When you set effectiveness values for 1 technology to be slightly improved or reduced, this leads to greater or reduced QALY estimates.</p> <p>There should be a clear link between the effectiveness of each technology and QALYs. Slightly improving the efficacy of a single technology should have a small effect on the incremental QALYs relative to the base case.</p>	

Item	Why it is assessed	Consequence if not checked
32	<p>When you increase mortality risk for each health state or event, this leads to lower QALYs and life-years for all technologies.</p> <p>Increasing mortality risk for a health state or event should impact ALL technologies. If life-years and QALYs do not decrease for a given technology, this may indicate that a different model structure is being applied and further investigation is warranted.</p>	
33	<p>When you reduce mortality risk for each health state or event, this leads to greater QALYs and life-years for all technologies.</p> <p>Decreasing mortality risk for a health state or event should impact ALL technologies. If life-years and QALYs do not increase for a given technology, this may identify that a different model structure is being applied and further investigation is warranted.</p>	
34	<p>When you increase baseline risks of events, this leads to lower QALYs for all technologies.</p> <p>There should be a clear link between baseline risks and QALYs. Increasing baseline risk should reduce health among all technologies assessed in the analysis.</p>	
35	<p>When you reduce baseline risks of events, this leads to higher QALYs for all technologies.</p> <p>There should be a clear link between baseline risks and QALYs. Decreasing baseline risk should improve health among all technologies assessed in the analysis.</p>	
36	<p>When you set mortality to be zero (i.e., patients do not enter the death state), life-years are identical across technologies.</p> <p>How mortality is programmed in the model should be clear. This test ensures all sources of mortality have been identified and accounted for.</p>	
37	<p>When you increase the cost of a technology, the only output impacted is the total lifetime costs for strategies that include that technology; likewise, there is no effect on QALYs or life-years.</p> <p>How the cost of each technology is programmed in the model should be clear. This test helps to understand the effect the cost associated with each technology has on model output.</p>	
38	<p>When you set all utilities to 1 and all disutilities to zero, the estimated QALYs are equivalent to life-years.</p> <p>How QALYs are estimated should be clear. This test helps to ensure QALYs have been accurately programmed.</p>	
39	<p>For evaluations with a time horizon greater than 1 year, when you set the discount rate to 0%, the costs and QALYs for all technologies increase.</p> <p>There is a clear link between discounting and model outputs. By removing discounting, future costs and health gains are not valued less and therefore QALYs and costs should increase</p>	

Item	Why it is assessed	Consequence if not checked
	<p>when discounting is removed. This test helps to identify whether discounting has been programmed correctly.</p>	
40	<p>For evaluations with a time horizon greater than 1 year, if you increase the discount rate, the costs and QALYs for all technologies decrease.</p> <p>There is a clear link between discounting and model outputs. By increasing the discount rate, future costs and health gains are valued less and therefore QALYs and costs should decrease when discounting is increased. This test helps to identify whether discounting has been programmed correctly.</p>	
41	<p>When you reduce the time horizon of the evaluation (the period costs and QALYs are estimated) this leads to lower estimated costs and QALYs for all technologies.</p> <p>There is a clear link between the time horizon and model outputs, and as the time horizon is reduced, fewer costs and benefits are valued and therefore should decrease. This test helps to ensure accurate model programming.</p>	
42	<p>It is possible to switch the inputs for 2 technologies and get the same results as before, meaning by changing the inputs (effectiveness, costs, QALYs), the model structure for any decision alternative can be used to model any other decision alternative.</p> <p>It should be transparent how a model captures the treatment effect associated with a technology. Once this is understood, it should be simple to replicate the impact a technology has elsewhere in the model. For example, if comparing technology A and B and technology A improves quality of life and survival, it should be clear how that is implemented in the model. To test the logic, you should be able to change inputs for technology B such that the outputs (costs per QALYs) match technology A. This may not always be practical if the model structure is different for different technologies. For example, if technologies have different induction periods or one is a surgery and the other is a chronic medication. In these scenarios, this test should still be explored but it may be limited.</p>	
43	<p>You can calculate the correlation between both the costs and QALYs for different technologies across the Monte Carlo simulation replications.</p> <p>When the model is run probabilistically, the output per simulation should be transparently laid out. From this, it should be relatively simple to conduct tests of correlation between costs and QALYs across these simulations. From the PSA output, you should be able to determine the total QALYs for each simulation. Once you have this output, you should be able to conduct a correlation test to determine if there is correlation across simulations. If this cannot be done, this means the probabilistic sensitivity analysis output is not transparent which prohibits a robust validation.</p>	

Item	Why it is assessed	Consequence if not checked
44	<p>Based on the results of the Monte Carlo simulation, there is a strong correlation between the estimates of costs (i.e., the estimated costs from each replication) for different technologies.</p> <p>When a model is run probabilistically, many variables apply to all technologies considered in the model, for example baseline rates, background mortality, and so on. When these parameters are varied, the impact is applied to all technologies in the model. This will create a correlation between absolute estimates of costs in the model across different simulations. This means when costs are high for 1 technology, there is a greater chance costs will be higher for other technologies (on average). If there was no correlation, this would indicate that there are no or very few common parameters across any of the technologies in the model. This is unlikely to be the case for most models.</p>	
45	<p>Based on the results of the Monte Carlo simulation, there is a strong correlation between the estimates of QALYs (i.e., the estimated QALYs from each replication) for different technologies.</p> <p>When a model is run probabilistically, many variables apply to all technologies considered in the model, for example baseline rates, probability of death, and so on. When these parameters are varied, the impact is applied to all technologies in the model. This will create a correlation between absolute estimates of QALYs in the model across different simulations. This means when QALYs are high for 1 technology, there is a greater chance QALYs will be higher for other technologies (on average). If there was no correlation, this would indicate that there are no or very few common parameters across any of the technologies in the model. This is unlikely to be the case for most models.</p>	
46	<p>The results of the deterministic analysis are broadly in line with the results of the probabilistic analysis. Justification is provided about why deterministic and probabilistic results are different.</p> <p>When we sample enough times, the average input value used for each parameter converges to the mean. In many cases, this means the probabilistic and deterministic results will generate similar results. In cases in which the results from the probabilistic results are vastly different from the deterministic, a clear explanation why should be provided, such as nonlinearities being present in the model. This will help ensure the difference is legitimate and not due to a programming error.</p>	

Scrutinization of Model Coding (White Box Verification)

The second process involves scrutinizing the code within models and establishing whether the links between the inputs and outputs are appropriate. This involves checking the detailed model calculations (white box testing). In an Excel-based model, white box testing requires scrutinizing the formulas in a spreadsheet that link the input parameters and the outcomes obtained. White box testing can be used to identify the root of possible issues raised by black box testing because black box testing alone cannot confirm whether the

model is providing correct results. White box testing should not be limited to areas of concern raised by black box testing but should focus on appraising the coding of the model with respect to all parameters that are determined important (either a priori or ex ante).

White box testing should be conducted both through a forward (assessing the link between input parameters and outcomes) and backward (tracing back from outcomes to input parameters) process. Focus should be on those input parameters that appear to lead to significant changes in QALYs and costs results when altered.

To facilitate white box testing, it would be useful for researchers to produce detailed manuals providing an example of how input parameters flowed through the model to produce outcomes.

Table 7: Validation of Model Coding Items and Consequences (Items 47 and 48)

Item	Why it is assessed	Consequence if not checked
47	<p>You can work backward from the results of the model to the location where inputs are entered.</p> <p>If you start with the final costs and QALYs from a model, it should be clear how those results were generated. This requires starting from the results and working backward to identify how they were calculated, eventually ending up at the inputs used to inform the analysis.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear how an input relates to the model output, the results cannot be robustly validated.</p>
48	<p>You can work forward from the location where inputs are entered to the results of a single Monte Carlo simulation.</p> <p>If you select any input in the model, it should be clear how that input feeds into the final results. There should be a clear trace to follow ensuring the input is programmed as intended.</p>	

Replication

An additional process – replication – could be considered, although it is unlikely to be possible in many cases. Therefore, replication is not included in the MVT-53 but it is included here because some researchers have expressed that it would be their preferred technique for model verification. Replication may be considered more relevant when significant unexplained concerns are raised after white box testing. Replication requires the reviewer to recode large parts (or, in some instances, all) of the model workings, typically when white box testing was unable to identify the root cause of the model behaviour. However, given the time constraints imposed by this method, no guidance on model replication is provided.¹²

General Issues of Concern

This final section contains a list of common issues that reviewers have noted should be avoided as part of good modelling practices to ensure transparency. The modelling practices identified can be appropriate in certain instances but frequently reduce the transparency required for model validation, particularly with respect to white box testing, and are almost always unnecessary. It would be pertinent for researchers to try to reduce or eliminate these practices within their models.

Table 8: General Issues of Concern and Consequences (Items 49 to 53)

Item	Why it is assessed	Consequence if not checked
49	<p>The use of the following functions limit model transparency, are inefficient, and are not required:</p> <ul style="list-style-type: none"> • IFERROR, IFNA, ISERROR, ISERR, or ISNA • CHOOSE, INDIRECT, OFFSET, and INDEX <p>The model makes no or limited use of these statements.</p> <p>There are many statements in Excel that force the model to ignore an error by replacing the error with a different value. Sometimes the model produces an error for a reason that is of no concern; for example, when the entire cohort has moved to the absorbing state, any errors beyond this time horizon will not impact the final results. However, when varying parameters probabilistically, for example, it can be difficult to separate out cases in which the error is legitimate from cases in which it is not. For example, an IFERROR statement may set QALYs to zero. If this happens infrequently over many probabilistic runs, then the error may be missed. It is important to ensure the model uses these statements sparingly, otherwise a robust validation is not possible.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model relies heavily on these statements, the model will be challenging to validate.</p>
50	<p>The model has no hidden sheets, rows, or columns.</p> <p>Microsoft Excel gives the user the ability to hide rows, columns, and worksheets. For the process of validation, the full model in its entirety should be available for validation.</p>	<p>The model is not sufficiently transparent.</p> <p>The researcher should ensure all rows, columns, and worksheets are unhidden before validating.</p>
51	<p>The model is free of user-created formulas embedded within VBA macros.</p> <p>Microsoft Excel gives the user the ability to create their own functions. These can be created in ways that are not transparent, meaning that they use data in a way that cannot be validated.</p>	<p>The model is not sufficiently transparent.</p> <p>If a model relies heavily on macros, clear user guides and transparent code should be used to navigate through them. Overreliance on macros may make white box testing impractical, which will limit the ability to robustly validate the model.</p>
52	<p>Parameters are not reset to default values after macros (e.g., for a Monte Carlo simulation) are run.</p> <p>When reviewing a model, a researcher may wish to make changes to certain parameters. The model may reset all parameters to default values after running certain macros. The user will then need to make all required changes again to the model, introducing the likelihood of user error. Therefore, it is important that the model removes this functionality to ensure the analysis being run is the one programmed by the user.</p>	<p>The model is not sufficiently transparent.</p> <p>If the model reverts all parameters to default after using a macro, this functionality should be removed. The original version of the model should be preserved without any changes in a separate file to ensure default values can be reverted to if required.</p>
53	<p>All input parameters that influence model results are provided in a transparent manner, preferably in a single worksheet.</p> <p>When coding, a model input should be entered once, and this should feed into all relevant parts of the model. If the same parameter is entered multiple times across different sheets, this not only introduces the likelihood of human error but also makes editing parameters challenging because this process must occur in multiple different places.</p>	<p>The model is not sufficiently transparent.</p> <p>If it is unclear how to implement a change to a certain parameter, this needs to be communicated and resolved by the model builder. If this is not possible, then a robust validation cannot be performed.</p>

Conclusion

CADTH has developed a framework that can be applied to all models with the aim to help ensure model validity in terms of both the conceptual and technical aspects of model design. The purpose of the framework is not related to concerns of any direction of bias from technical issues within economic models, but rather to ensure that the process of model appraisal is consistent and reproducible.

The framework is intended to be a living document and will be updated over time to stay in line with best modelling practices.

References

1. Canadian Agency for D, Technologies in H. Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Ed.). <https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition>. Published 2017. Accessed.
2. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *International journal of technology assessment in health care*. 2022;38(1):e13. [PubMed](#)
3. Kim DD, Do LA, Synnott PG, et al. Developing criteria for health economic quality evaluation tool. *Value in Health*. 2023;26(8):1225-1234. [PubMed](#)
4. Lim KK, Koleva-Kolarova R, Fox-Rushby J. A Comparison of the Content and Consistency of Methodological Quality and Transferability Checklists for Reviewing Model-Based Economic Evaluations. *PharmacoEconomics*. 2022;40(10):989-1003. [PubMed](#)
5. Lamrock F, O'Connor J, Leahy J, Gorry C, Tilson L, Barry M. OP97 Cost-effectiveness Model Appraisal Guidelines For Health Technology Assessments In Ireland. *International Journal of Technology Assessment in Health Care*. 2019;35(S1):24-25.
6. Guidelines TPBAC. Section 3A.7 Model validation. <https://pbac.pbs.gov.au/section-3a/3a-7-model-validation.html>. Accessed September 19, 2023.
7. INESSS. CASP: PROGRAMME DE DÉVELOPPEMENT DES COMPÉTENCES EN ÉVALUATION CRITIQUE. https://www.inesss.qc.ca/fileadmin/doc/INESSS/DocuMetho/CASP_Economie_FR2013_V14012015.pdf. Published 2013. Accessed September 19, 2023.
8. Vemer P, Corro Ramos I, Van Voorn G, Al M, Feenstra T. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. *Pharmacoeconomics*. 2016;34:349-361. [PubMed](#)
9. Evelina Zimovetz SW. Reviewer's Checklist for Assessing the Quality of Decision Models. ISPOR 12th Annual European Congress; 2009; Paris, France.
10. Büyükkaramikli NC, Rutten-van Mölken MP, Severens JL, Al M. TECH-VER: a verification checklist to reduce errors in models and improve their credibility. *Pharmacoeconomics*. 2019;37:1391-1408. [PubMed](#)
11. Tappenden P, Chilcott JB. Avoiding and identifying errors and other threats to the credibility of health economic models. *Pharmacoeconomics*. 2014;32:967-979. [PubMed](#)
12. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and facilitators to model replication within health economics. *Value in Health*. 2019;22(9):1018-1025. [PubMed](#)

Appendix 1: Model Validation Tool

A copy of the full [model validation tool](#) is also available on the website.

Validation of the Conceptual Model

Decision Problem

The decision problem relates to interventions to be compared, the population(s) in which they are compared, the perspective for the evaluation, which costs and outcomes are to be considered, and the time horizon of the evaluation.

If any item is not present, the model may not reflect the decision problem and therefore its conclusions may not be valid.

Table 9: Validation of Decision Problem (Items 1 to 5)

Item	Description	Yes	No	NA
1	The model built is reflective of the stated population that the decision problem applies to.			–
2	The model can examine key subgroups within the population of interest.			–
3	The model assesses all comparators used to currently treat the stated population.			–
4	The model incorporates costs that are consistent with the specified perspective of the analysis.			–
5	The model assesses all outcomes deemed important by clinicians and patients.			–

Model Specification

Model specification relates to the choice of model type, the health states that are modelled, and, when applicable, the choice of cycle length. When building a model, it is important to consider what would be an appropriate specification of the model. To undertake this, it is worthwhile to consider what you would consider an optimal specification of the model. This can be informed by 2 distinct processes: a review of existing models in the area and a formal consideration of the disease process and clinical pathway. This is required to help ensure external validity of the model.

Table 10: Validation of Model Specification (Items 6 and 7)

Item	Description	Yes	No	NA
6	The structure of the model (i.e., the process and clinical pathway) has been validated by clinical experts.			–
7	The model follows previous models in this clinical area or justification has been provided about why the model structure differs from previous models. [Select NA if there are no previous models in this clinical area.]			

Modelling of Clinical Effectiveness

One component of the conceptual model that will have a large impact on the results of the related analysis is the process by which clinical effectiveness is incorporated within the model. It is important to consider issues such as the quality and consistency of the evidence, the assumed duration of the effect, double counting of benefit, and appropriate consideration of uncertainty.

Note the following does not consider the quality or robustness of the evidence considered. Please refer to the CADTH guidelines for guidance on evidence appraisal.

How Was Clinical Evidence Modelled?

The following items apply to state transition models. Some of the items in this section may not be applicable for decision trees or discrete event simulations. In these cases, select NA. For more complicated models, further validation steps will likely be required.

Table 11: Validation of Clinical Effectiveness (Items 8 to 25)

Item	Description	Yes	No	NA
8	Time spent in each health state, for each technology assessed, can be extracted from the model. [Select NA if the model does not utilize health states (e.g., a decision tree) and skip item 9.]			
9	The model output matches the evidence provided to support time spent in health states among technologies.			–
10	If clinical events are modelled (i.e., hospitalizations, exacerbations, strokes, hip fractures), the number of events for each technology can be extracted from the model. [Select NA if clinical events are not relevant to the decision problem and skip item 11.]			
11	Model output matches the evidence provided to support the number of clinical events across technologies.			–
12	The impact of adverse events on health outcomes and costs can be extracted from the model. [Select NA if adverse events are not relevant to the decision problem and skip item 13.]			–
13	The model output matches evidence provided for adverse event type and frequency from the evidence.			–
14	Life-years are reported as a result within the model.			–
15	The impact the evaluated technologies has on mortality is clear. [Select NA if there are no differences in mortality and skip items 16, 17, and 18.]			
16	If differences in mortality are noted in Item 15, select the reasons for differing mortality in the model (more than 1 reason can be selected): a. Duration of time spent in health states lead to higher mortality risk. b. There is a difference in the frequency of fatal clinical events. c. There is a difference in the frequency of fatal adverse events.	–	–	–

Item	Description	Yes	No	NA
	d. Direct impact on risk of death, not stated previously, has been modelled (i.e., direct modelling of overall survival from the trial).			–
17	If there are mortality differences between technologies, the model can extract which of the reasons from item 16 has the largest impact on incremental life-years.			–
18	Model output matches evidence regarding mortality rates between different technologies.			–
19	It is clear that the model does not utilize technology-specific utilities.			–
20	Based on the results from the submitted model, it can be determined which of the following has the largest impact on cost-effectiveness conclusions: time spent in health states, number of clinical events occurring, adverse events, and mortality.			–
21	The model distinguishes data that are based on extrapolation methods (i.e., using parametric survival analysis). [Select NA if no extrapolation is required.]			
22	The model time horizon can be adjusted for just the period for which there are clinical data available. [Select NA if the data cover the full period for which clinical data are available.]			
23	If the model incorporates both direct and indirect effects, it is clear how double counting has been avoided (e.g., a direct effect applies to mortality through applying a hazard ratio to overall survival and an indirect effect is applied to the probability of an event or transition that is associated with a mortality risk). [Select NA if only direct or indirect effects are included.]			
24	Does the modelled relationship between surrogate outcomes and final outcomes (quality of life and mortality) match the evidence presented? [Select NA if no surrogate outcomes are used.]			
25	The model allows flexibility to explore waning of treatment effects OR evidence and rationale is provided that suggests treatment effects are permanent and enduring. [Select NA if no extrapolation of treatment effect is required.]			

Computer Model Validation and Verification

The process of model verification can be separated into 2 distinct processes which are akin to those adopted in software verification: assessment of model behaviour (black box testing) then scrutinization of the coding of the model (white box testing). It is important that these processes are conducted in the correct order.

To enable black box and white box testing, there are essential features of the model that need to be in place. The following section relates to these essential features.

Model Transparency (3 Items)

Table 12: Validation of Model Transparency (Items 26 to 28)

Item	Description	Yes	No	NA
26	The model can access the deterministic result and the results from single Monte Carlo simulations.			–
27	A clear trace can be identified that links all input parameters to final outcomes (i.e., only input parameters are hard coded).			–
28	Macros are exclusively related to first- or second-order simulation and model navigation (exclusive to models built in Microsoft Excel).			–

Assessment of Model Behaviour (Black Box Verification)

The first process involves ascertaining whether changing the inputs of the model leads to results that meet the general expectations of the reviewer. This is referred to as black box testing because it does not require the reviewer to know the inner workings of the model. If, during black box testing, the model fails to provide results that are explainable, the reviewer could conduct detailed white box testing (refer to [Table 14](#)) to determine why the results are not as expected.

The following items include a range of possible black box tests.

Table 13: Validation of Assessment of Model Behaviour (Items 29 to 46)

Item	Description	Yes	No	NA
29	You can set the effectiveness of different technologies such that QALY estimates are equal.			–
30	When you set effectiveness values to be extremely in favour of or against 1 technology, this leads to substantially greater or reduced QALY estimates.			–
31	When you set effectiveness values for 1 technology to be slightly improved or reduced , this leads to greater or reduced QALY estimates.			–
32	When you increase mortality risk for each health state or event, this leads to lower QALYs and life-years for all technologies.			–
33	When you reduce mortality risk for each health state or event, this leads to greater QALYs and life-years for all technologies.			–
34	When you increase baseline risks of events, this leads to lower QALYs for all technologies.			–
35	When you reduce baseline risks of events, this leads to higher QALYs for all technologies.			–
36	When you set mortality to be zero (i.e., patients do not enter the death state), life-years are identical across technologies.			–
37	When you increase the cost of a technology, the only output impacted is the total lifetime costs for strategies that include that technology; likewise, there is no effect on QALYs or life-years.			–
38	When you set all utilities to 1 and all disutilities to zero, the estimated QALYs are equivalent to life-years.			–

Item	Description	Yes	No	NA
39	For evaluations with a time horizon greater than 1 year, when you set the discount rate to 0%, the costs and QALYs for all interventions increase. [Select NA if the time horizon is shorter than 1 year because discounting is only relevant for models with time horizons longer than 1 year.]			
40	For evaluations with a time horizon greater than 1 year, if you increase the discount rate, the costs and QALYs for all interventions decrease. [Select NA if the time horizon is shorter than 1 year because discounting is only relevant for models with time horizons longer than 1 year.]			
41	When you reduce the time horizon of the evaluation (the period costs and QALYs are estimated) this leads to lower estimated costs and QALYs for all interventions.			–
42	It is possible to switch the inputs for 2 technologies and get the same results as before, meaning by changing the inputs (effectiveness, costs, QALYs), the model structure for any decision alternative can be used to model any other decision alternative.			–
43	You can calculate the correlation between the costs and QALYs for different technologies across the Monte Carlo simulation replications.			–
44	Based on the results of the Monte Carlo simulation, there is a strong correlation between the estimates of costs (i.e., the estimated costs from each replication) for different technologies.			–
45	Based on the results of the Monte Carlo simulation, there is a strong correlation between the estimates of QALYs (i.e., the estimated QALYs from each replication) for different technologies.			–
46	The results of the deterministic analysis are broadly in line with the results of the probabilistic analysis. Justification is provided about why deterministic and probabilistic results are different.			–

Scrutinization of Model Coding (White Box Verification)

The purpose of the second process is to establish whether the links between inputs and outputs are appropriate. This involves checking the detailed model calculations (white box testing). White box testing requires scrutinizing the formulas in a spreadsheet that link input parameters and outcomes. White box testing can identify the root of possible issues raised by black box testing. Black box testing cannot identify whether the model is providing correct results. This can only be ascertained through white box testing. Thus, white box testing should not be limited to areas of concern raised by black box testing but should focus on appraising the coding of the model with respect to all parameters that are determined to be important (either a priori or ex ante).

Table 14: Validation of Model Coding (Items 47 and 48)

Item	Description	Yes	No	NA
47	You can work backward from the results of the model to the location where inputs are entered. [For example, if you take the total costs associated with an intervention, can you work back from this value to determine how it was estimated and what inputs were used to derive this value?]			–
48	You can work forward from the location where inputs are entered to the results of a single Monte Carlo simulation. [For example, if you take a random input into the model (e.g., technology cost), can you trace how this input influences costs and or QALYs in the model?]			–

General Issues of Concern

The following relates to common issues of concern expressed by the health economists consulted for this work regarding models built in Microsoft Excel. The issues are of direct relevance to white box testing. The following functions should be avoided to ensure model transparency and affect the reliability of the model analysis. Often these functions are used to override user-provided inputs.

Table 15: General Issues of Concern (Items 49 to 53)

Item	Description	Yes	No	NA
49	The use of the following functions limit model transparency, are inefficient, and are not required: <ul style="list-style-type: none"> • IFERROR, IFNA, ISERROR, ISERR, or ISNA • CHOOSE, INDIRECT, OFFSET, and INDEX The model makes no or limited use of these statements.			–
50	The model has no hidden sheets, rows, and columns.			–
51	The model is free of user-created formulas embedded within VBA macros.			–
52	Parameters are not reset to default values after macros (e.g., for a Monte Carlo simulation) are run.			–
53	All input parameters that influence model results are provided in a transparent manner, preferably in a single worksheet.			–

Authors: Doug Coyle, Alex Haines, Karen Lee

ISSN: 2563-6596

Disclaimer: The information in this document is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. While patients and others may access this document, the document is made available for informational purposes only and no representations or warranties are made with respect to its fitness for any particular purpose. The information in this document should not be used as a substitute for professional medical advice or as a substitute for the application of clinical judgment in respect of the care of a particular patient or other professional judgment in any decision-making process. The Canadian Agency for Drugs and Technologies in Health (CADTH) does not endorse any information, drugs, therapies, treatments, products, processes, or services.

While care has been taken to ensure that the information prepared by CADTH in this document is accurate, complete, and up-to-date as at the applicable date the material was first published by CADTH, CADTH does not make any guarantees to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in any third-party materials used in preparing this document. The views and opinions of third parties published in this document do not necessarily state or reflect those of CADTH.

CADTH is not responsible for any errors, omissions, injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the contents of this document or any of the source materials.

This document may contain links to third-party websites. CADTH does not have control over the content of such sites. Use of third-party sites is governed by the third-party website owners' own terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third-party sites and CADTH is not responsible for any injury, loss, or damage suffered as a result of using such third-party sites. CADTH has no responsibility for the collection, use, and disclosure of personal information by third-party sites.

Subject to the aforementioned limitations, the views expressed herein are those of CADTH and do not necessarily represent the views of Canada's federal, provincial, or territorial governments or any third-party supplier of information.

This document is prepared and intended for use in the context of the Canadian health care system. The use of this document outside of Canada is done so at the user's own risk.

This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

The copyright and other intellectual property rights in this document are owned by CADTH and its licensors. These rights are protected by the Canadian *Copyright Act* and other national and international laws and agreements. Users are permitted to make copies of this document for non-commercial purposes only, provided it is not modified when reproduced and appropriate credit is given to CADTH and its licensors.

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to Requests@CADTH.ca