**APPROACHES TO MODELLING IN DECISION-ANALYTIC ECONOMIC EVALUATIONS**

**APPROACHES TO MODELLING IN DECISION-ANALYTIC ECONOMIC EVALUTIONS**

**By**

**BERNICE TSOI, Hon.B.Sc., M.Sc.**

**A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree in Doctor of Philosophy**

**McMaster University © Copyright by Bernice Tsoi, 2015**



# DESCRIPTIVE NOTE

Doctor of Philosophy (2015) McMaster University. Faculty of Health Sciences, Department of Clinical Epidemiology and Biostatistics: Health Research Methodology, Health Technology Assessment Specialization

TITLE: Approaches to modelling in decision-analytic health economic evaluations: Factors to differentiate between approaches

AUTHOR: Bernice Tsoi Hons.B.Sc. (McMaster University) M.Sc. (London School of Economics and Political Sciences)

SUPERVISOR: Associate Professor Daria O’Reilly, MSc, PhD

THESIS COMMITTEE: Jean-Éric Tarride, MA, PhD

Gord Blackhouse, MBA, MSc

NUMBER OF PAGES: xvii, 178

|  |
| --- |
| ABSTRACT |

**Background & Objectives**

Decision-analytic modelling can inform healthcare resource allocation and reimbursement decision-making, with modelling approaches adapted from a variety of disciplines. The objective of this thesis was to investigate the evidence surrounding when each approach should be used when conducting health economic evaluations.

**Methods**

**Project 1:** A systematic review identified selection criteria, referred to as factors, through an evaluation of existing decision frameworks that aimed to differentiate between models.

**Project 2:** Employing the factors identified from Project 1, a systematic review explored the extent to which empirical cross-validation studies agree on the importance of each on its impact to model selection.

**Project 3:** A decision tree evaluating the cost-effectiveness of two vaccination strategies in children was reconstructed as system dynamics and agent-based models and compared. Scenario analyses assessed the situations whereby the model’s results would be sensitive to or be better handled by a particular approach.

**Results**

**Project 1:** Among the eight frameworks identified; each involved a different set of structural or practical factors. Disagreements emerged between frameworks in the classification of the structural features specific to each modelling approach.

**Project 2:** Nine exercises have been conducted, mostly focused on the criteria of interactivity (i.e., static vs. dynamic) and population resolution (i.e., aggregate vs. individual). Aggregate- and individual-level models were found to produce similar results with a practical trade-off between validity and feasibility. In the presence of large indirect effects, dynamic and static models often produced disparate results.

**Project 3:** When calibrated, all three approaches reached consistent findings. Adaptation away from base-case assumptions led to different quantitative results on which vaccination strategy would be most optimal.

**Conclusion**

Despite disagreement among the frameworks on how to recommend modelling approaches, consistent conclusions were observed in empirical cross-validation studies. More empirical evidence is therefore needed to improve one’s understanding of the impact of the remaining factors on model selection.

|  |
| --- |
| ACKNOWLEDGEMENTS |

The road to completing a PhD can be both long and arduous. Some may describe it as a rollercoaster ride, others as a journey through rough seas. These metaphors all share a commonality in that it describes the ‘highs’ and ‘lows’ characterizing the process and emotions experienced while pursuing higher education: it is unexpected, rewarding and enlightening, but it can also be mixed with times of feeling adrift. The work represented herein would not have been possible without the support of many who celebrated my achievements but, more importantly, supported me even through tough times.

This dissertation represents the successive growth in scientific inquiry and knowledge over the past four years. My deepest gratitude goes to my supervisor: Dr. Daria O’Reilly. Serendipity brought her into this role and I am sincerely grateful for her constant support and dependability that paved a smoother journey over these years. Our collaborations have truly had a transformative influence on me and I am fortunate that beyond simply being a supervisor, she has and continue to be a mentor.

A single great mind provides only one perspective, but the great minds of many fosters knowledge and a broader understanding. Sincerest thanks goes to Dr. Jean Eric Tarride and Prof. Gord Blackhouse who provided these different views as my committee members. I truly value the discussions they brought forth, the guidance they were willing to share and, despite their busy schedules, their constant availabilities in times of need.

Life is also filled with people who come into our lives: some for only brief moments, others for a longer stay. Each, nonetheless, leaves a mark that shapes who we are and the direction we go. I am grateful to the colloquially-known PATHoids: the faculty, researchers, administrative staffs and students at the PATH research institute who provided the entertaining debates, the exchanges in knowledge (in which some were loosely related to our research) and the immense support.

In addition, I express gratitude to Prof. Ron Goeree who was instrumental in shaping my interests on my thesis topic and in ensuring that I had the necessary tools to be successful in my PhD; and to Dr. Nathaniel Osgood from the University of Saskatoon for first bringing into perspective how to construct agent-based and system dynamics models and for providing me the support for the last paper of my thesis.

To my friends and fellow HRM colleagues, thank you for being a part of this journey and for staying for the next one! Last, but not least, to my family and loved ones - especially my parents - I am truly indebted. You have always been my omnipresent and yet, least acknowledged, cheerleaders and support team. Not only have you taught me the value of hard work and the importance of pursuing knowledge but, more importantly, the biggest lesson to date has been to never compete and be better than everybody else, but rather, to always be better than my past self. Without holding this lesson dear to me, I would not be where I am presently.

|  |
| --- |
| **TABLE OF CONTENTS** |

[DESCRIPTIVE NOTE ii](#_Toc428204562)

[ABSTRACT iii](#_Toc428204563)

[ACKNOWLEDGEMENTS v](#_Toc428204564)

[LIST OF FIGURES ix](#_Toc428204565)

[LIST OF TABLES x](#_Toc428204566)

[LIST OF APPENDICES xii](#_Toc428204567)

[PREFACE xiv](#_Toc428204568)

CHAPTER 1: [Decision-analytic economic modelling in healthcare decision-making: Why do we need more approaches? 0](#_Toc428204570)

[1.1 Decision Analysis in the Context of Health care 0](#_Toc428204571)

[1.2 Mathematical Modelling in Health Care Decision Analysis 1](#_Toc428204572)

[1.3 Fundamentals to Constructing Decision-Analytic Health Economic Models 3](#_Toc428204573)

[1.4 Existing Approaches to Decision-Analytic Modelling 5](#_Toc428204574)

[1.4.1 Decision Trees 5](#_Toc428204575)

[1.4.2 State-Transition Models 5](#_Toc428204576)

[1.4.3 Discrete Event Simulation 8](#_Toc428204577)

[1.4.4 Agent-based Models 9](#_Toc428204578)

[1.4.5 System Dynamics 10](#_Toc428204579)

[1.4.6 Infectious-disease Compartmental Models (‘SIR’ and its variants) 11](#_Toc428204580)

[1.5 Validity of Economic Modelling 11](#_Toc428204581)

[1.6 Outline of the Thesis 14](#_Toc428204582)

[1.7 References 19](#_Toc428204583)

[CHAPTER 2:](#_Toc428204584) [Systematic narrative review on DECISION FRAMEWORKS to select the appropriate modelling approachES for HEALTH ECONOMIC evaluations 22](#_Toc428204585)

[CHAPTER 3:](#_Toc428204586) [Do different decision-analytic modelling approaches produce different results? A systematic review of cross-validation studies 73](#_Toc428204587)

[CHAPTER 4:](#_Toc428204588) [Comparison of cost-effectiveness results by decision trees, agent-based and system dynamics models: A case of influenza vaccination 113](#_Toc428204589)

[CHAPTER 5:](#_Toc428204590) [Discussion and Conclusion 165](#_Toc428204591)

[5.1 Summary and Major Contributions 165](#_Toc428204592)

[5.2 Future Areas of Research 172](#_Toc428204593)

[5.3 Main concluding points of the thesis 174](#_Toc428204594)

[5.4 References 176](#_Toc428204595)

|  |
| --- |
| LIST OF FIGURES |

CHAPTER 2

|  |  |
| --- | --- |
| **Figure 1:** PRISMA diagram of the literature search process for articles on decision frameworks for model approach selection | 45 |

CHAPTER 3

|  |  |
| --- | --- |
| **Figure 1:** PRISMA diagram of the literature search process for documents that have conducted cross-validation between different modelling approaches | 95 |

CHAPTER 4

|  |  |
| --- | --- |
| **Figure 1:** Example of a causal-loop diagram describing the transmission of a hypothetical infectious disease based on the susceptible-infectious-recovered (SIR) structure  **Figure 2:** Model structure for immunization  **Figure 3:** Diagrammatic representation of three common network topologies in ABM  **Figure 4:** Cost-effectiveness acceptability curve from the two aggregate-level models | 143  144  146  147 |

|  |
| --- |
| LIST OF TABLES |

CHAPTER 2

|  |  |
| --- | --- |
| **Table 1:** Description of modelling approaches employed in health economic evaluation  **Table 2:** Overview of the decision framework and the modelling approach covered within the respective frameworks  **Table 3:** Summary of the decision criteria considered within each decision framework  **Table 4**: Nomenclature and definition of commonly-mentioned decision criterion for selecting a modelling approach  **Table 5:** Classification of structural elements, specific to each modelling approach, according to the decision frameworks | 46  49  52  55  61 |

CHAPTER 3

|  |  |
| --- | --- |
| **Table 1:** Approaches to decision-analytic modelling  **Table 2:** List of selection criteria commonly used in existing frameworks to guide choice on modelling approach  **Table 3:** Summary of empirical studies that have compared two or more modelling approaches  **Table 4**: Current understanding of the impact of each structural decision criterion | 96  100  102  104 |

CHAPTER 4

|  |  |
| --- | --- |
| **Table 1:** Comparison of the main characteristics of agent-based and system dynamics model  **Table 2:** Description of common transitions in agent-based modelling  **Table 3:** Model parameters that differ by modelling approach  **Table 4:** Summary of scenarios studied  **Table 5:** Expected costs, effects and ICER from all three modelling approaches following calibration  **Table 6:** Probabilistic estimates from the cohort-level models for children vaccinated between the ages of 2 to 5  **Table 7:** Expected results under a varied vaccination scenario for children age 2 to 5  **Table 8:** Expected results of ABM across varying network topology for children vaccinated between the ages of 2 to 5  **Table 9**: Deterministic results of decision tree and ABM when incorporating heterogeneity by expanding the ages of patients modelled. | 148  149  150  151  152  153  154  155  156 |

|  |
| --- |
| LIST OF APPENDICES |

CHAPTER 2

|  |  |
| --- | --- |
| **Appendix 2.I**: Search strategy  **Appendix 2.II:** Level I selection criteria  **Appendix 2.III:** Copy of the existing decision frameworks | 63  64  65 |

CHAPTER 3

|  |  |
| --- | --- |
| **Appendix 3.I**: Search strategy  **Appendix 3.II:** Level I selection criteria  **Appendix 3.III:** Summary of the results of empirical studies that have compared two or more modelling approaches | 106  107  108 |

CHAPTER 4

|  |  |
| --- | --- |
| **Appendix 4.I:** Model parameters (for children, aged 2-5)  **Appendix 4.II:** Model parameters that differ across age groups (specific to the decision tree and agent-based model) | 157  163 |

|  |
| --- |
| **LIST OF ABBREVIATIONS** |

**ABM** agent-based model

**DES** discrete event simulation

**HTA** health technology assessment

**HIV** human immunodeficiency virus

**ICER** incremental cost-effectiveness ratio

**ISPOR** International Society for Pharmacoeconomics and Outcomes Research

**LAIV** intranasal live attenuated vaccine

**QALYs** quality-adjusted life years

**SIR** susceptible-infected-recovered

**SMDM** Society for Medical Decision Making

**SD** system dynamics

**TIV** injectable trivalent inactivated influenza vaccine

|  |
| --- |
| PREFACE |

This thesis represents a “sandwich thesis” that combines three individual projects that have been prepared for publication in peer-reviewed journals. Two of the papers have been published while one has been submitted for publication. The contributions of Bernice Tsoi to all three papers within this thesis include: developing the research ideas and research questions, performing the analyses, interpreting the results, writing the manuscripts, submitting the manuscripts for publication and responding to reviewer comments. The work in this thesis was conducted between Winter 2012 to Autumn 2014.

|  |
| --- |
| CHAPTER 1:Decision-analytic economic modelling in healthcare decision-making: Why do we need more approaches? |

**1.1 Decision Analysis in the Context of Health care**

Decision analysis is a discipline that systematically identifies, represents and assesses aspects of decision-making in order to prescribe a recommended course of action according to the axioms of expected utility theory. Despite originating from game theory, its popularity has grown beyond economics with application extending to other disciplines, including the health sector where it was first applied in the 1960s1.

Given that decisions on health care policies and management can be complex, decision analyses have been used under a variety of settings such as in addressing resource allocation. For instance, with the development of new health technologies, their introduction and diffusion may offer opportunities to realize improved health outcomes. However, this has proven problematic at a broader health systems level. Notwithstanding that these new technologies are frequently more costly, they may also increase demand due to expanding indications or changes in demography2. Indeed, several studies have shown that technological progress is a primary factor for the observed rise in health care expenditures3,4. Furthermore, medical innovations can introduce disruptions to the existing health care system by changing existing processes and deliveries of care. These disruptions can either be beneficial such as by introducing efficiencies5, or detrimental6 such as by increasing resource demands.

Given this environment, policy-makers face a considerable challenge. Their mandate to meet their population’s demands and expectations often conflicts with their needs to manage a pre-specified and constrained health care budget. Scarce resources must therefore be allocated strategically by investing in technological innovations that deliver the best possible health outcomes. Furthermore, intrinsic uncertainty exists in the demands and outcomes associated with a health technology. Decision analysis provides a structured and systematic approach to estimate both the costs and the health consequences associated with a health technology relative to its alternatives, and to characterize the uncertainty associated with a particular decision. It provides answers to key questions such as whether other savings can be expected elsewhere to offset the higher prices of a given treatment; or, if not, whether the improved effectiveness justifies the added incremental costs when compared to other treatment options.

**1.2 Mathematical Modelling in Health Care Decision Analysis**

Models are a set of decision-analytic techniques that rely on mathematics to describe the operation of real-world processes. The process under analysis is defined as the system; with the model representing a set of mathematical and/or logical equations that are combined in a structured framework to describe system mechanics7. The objective of a model is to abstract reality into simpler representations in order to simulate complex processes, to describe system behaviours and/or to quantify outputs according to predefined inputs8,9.

There are a number of circumstances where a mathematical model may be employed in health care decision-analysis, including:

* Temporal extrapolation beyond the data observed in a clinical trial
* Linkage of intermediate clinical endpoints to final health outcomes
* Contextual extrapolation (or “generalization”) of the results obtained in one clinical setting to another that is considerably different
* Bridging of the treatment effects observed in a restricted patient setting to a wider population
* Head-to-head comparisons of alternative interventions when such direct comparisons do not exist

Its application can further include situations where clinical studies are unethical, cost prohibitive or impractical10.

Before continuing further, it is necessary to clarify how the term *mathematical model* is defined in this thesis given its broad definition. Mathematical models that quantify the effects of a health intervention can be distinguished into empirical, mechanistic or decision-analytic11-13.

Empirical models are those that describe the relationships between predictive factors and outcomes based on the experimental data that are available, such as regression models derived from observational or clinical trials. No attempts are made to understand the underlying mechanism and such models are built according to data availability. For instance, many empirical models exist that predict disease progression and time until a particular clinical event by correlating the outcome to clinical and demographic indicators that have been previously collected.

Mechanistic models are less data-dependent than empirical models as they explicitly describe the mechanisms underlying a system. Constructed based on both experimental data and judgement, such models often combine data from multiple sources according to a hypothesized understanding of the relationship between factors. For example, the first data source could be quantifying the relationship between vaccination and likelihood of contracting an infectious disease for a particular individual while the second data source may combine findings from an observation study that characterizes disease transmission on a broader population. By combining different sources of data, one can better understand the potential impact of individual vaccination strategies on population-specific disease spread. Mechanistic models can further address counter-factual “what if” scenarios when the described mechanism is well-validated11.

In contrast, decision-analytic models aim to inform decision-making by determining the value associated with alternative courses of action13. Such models compare the respective value of different health technologies by characterizing how each influences disease course and the subsequent link between disease progression and the outcomes of interest. For the purpose of this thesis, the term “model” refers solely to decision-analytic models. This restricted definition thereby concentrates on those models that are used as normative decision-making aids. These include health economic models that place a value on both the clinical and cost outcomes in order to help decision-makers prioritize expenditures and/or allocate scarce resources. It excludes from consideration several useful and scientifically sound modelling forms (e.g., regression models). For example, an infectious disease transmission model would be considered outside the scope of this definition if it only predicts the course of an epidemic. However, it would be considered within the scope if it is used to evaluate what can be done to affect or change the course of an epidemic.

**1.3 Fundamentals to Constructing Decision-Analytic Health Economic Models**

The process of model building can be seen as two distinct components:

(1) the conceptualization of a problem, which draws on knowledge of the health care process and the desired problem at hand, and

(2) the conceptualization of the model itself by constructing it in a way such that it captures the attributes and the characteristics of the problem being studied14.

To construct a model, three fundamental elements must be defined: I) the model basics; II) the mathematical model itself; and III) its parameters (i.e., data)15. The first, *model basics*, entails scoping and defining the decision problem and making basic methodological choices such as the perspective to be adopted and the time horizon to be captured. The second element, the *mathematical model*, comprises the model structure and is related to the first step of scoping and defining the decision problem. In particular, in this step, three major considerations are of particular importance: the approach to modelling which is the focus of this thesis; the model substructure (e.g., elements that comprise the architecture of a model which can be general, such as the time interval, or can be intervention/disease-specific, such as the specific health states); and the model’s structural assumptions. Practicalities to modelling may have an influence on this step. The final aspect, *model* *parameters*, describes the data sources, the data modelling processes, and the data incorporation methods13,16. At each step, careful consideration must be made as uncertainty and bias may exist that would need to be addressed through the conduct of appropriate sensitivity analyses.

This thesis addresses the evidence on selecting a modelling approach given that this issue is rarely recognized or justified in existing decision-analytic health economic evaluations17. In particular, we address whether the choice of an approach truly matters. Given the availability of ‘alternative’ modelling approaches, it is important to identify aspects and types of problems that would be most appropriate for each approach. Similarly, with the acceptance and use of the newer modelling approaches, it is important to be cognizant of the situations where two different approaches are likely to produce similar conclusions such that the simpler of the two approaches may be preferred to model future decision problems that share similar properties.

**1.4 Existing Approaches to Decision-Analytic Modelling**

Existing approaches to modelling in medical decision-analysis have been adopted from disciplines of mathematics and operational research. These include traditional techniques (i.e., decision trees and Markov models) and, with growing computational capacity, a large number of alternative approaches that are now feasible (e.g., discrete event simulation (DES), agent-based models (ABM), susceptible-infected-recovered (SIR) models, and system dynamics (SD)).

**1.4.1 Decision Trees**

Decision trees primarily model sequenced decision problems and their possible consequences. It describes the decisions to be made, the resulting chance events that may occur and the outcomes associated with the combinations of decisions and events. Events are arranged in temporal order, thereby, illustrating the logic and potential outcomes arising from competing courses of action. The decision tree consists of three components: (i) decision nodes; (ii) chance nodes; and (iii) terminal nodes. The decision node presents the set of decision alternatives that are being considered; the chance node presents the set of mutually exclusive and collectively exhaustive consequences that may be possible for a particular decision; and lastly, the terminal node presents the value of the individual pathways (i.e., the combination of decision and events, colloquially known as branches)18. Conventionally, the population modelled is closed (i.e., no entry or exit of individuals in the model; or, in other words, the population size remains constant).

**1.4.2 State-Transition Models**

State-transition models conceptualize the decision problem as sets of health states and transitions among these states. State-transition models represent an over-arching term encompassing a wide collection of modelling approaches (e.g., Markov cohort and Markov microsimulation) that share certain key features in its model structure. State-transition models describe the system under study according to states, transitions and transition probabilities. The decision problem is described by a set of distinct mutually exclusive and collectively-exhaustive states with movement, or transition, from one state to the next dictated by transition probabilities19,20.

1.4.2.1 Markov Cohort

The term cohort refers to the fact that a single population cohort enters the model together and progresses through the states simultaneously19. Movement within the model occurs at discrete fixed periods, known as Markov cycles18-20.

As an entire cohort is modelled, an assumption on homogeneity is required whereby patients within the same states are considered homogeneous, or indistinguishable, in terms of both observed and unobserved characteristics19-21. This creates difficulty in capturing disease complexity, especially interrelated risks22. A related, fundamental assumption to such models is the Markovian assumption, whereby transition probabilities are independent of history, such as past states or time spent in the current state. Movement in the next cycle depends solely on the present health state19,20. In other words, the model is “memory-less” and transition probabilities between health states are constant over time and applied equally to all members within the same state21,23,24. This may be limiting in situations where history is predictive of future clinical events19. Although there are strategies to embed heterogeneity and knowledge of disease history into the model structure, there is often a sacrifice in model efficiency (i.e., the extent to which a model is manageable). For instance, heterogeneity or memory can be incorporated by introducing new health states. However, every additional factor considered in the model will expand the number of health states required and may lead to an unmanageable number of health states, a situation commonly referred to as state explosion19. Memory can also be mimicked by the use oftemporary tunnel states. Tunnel states are structural temporary side paths where patients are shunted to a different part of model until such time as the conditional change is over. However, such an approach only introduces temporary memory for as long as a patient remains within the tunnel. Another method to bypass the assumption of constant transition probabilities is to model non-homogeneous Markovian stochastic processes, also referred to as semi-Markov state-transition model. In such situations, state-transitions vary18,20 according to the specific time-dependent transition matrices23. Markov cohort models are conventionally modelled assuming a closed population7.

1.4.2.2 Markov Microsimulation

Microsimulations differ from Markov cohort models as they simulate, at the individual-level, one person at a time. Individual’s transitions are generated using a random process and are drawn from probability distributions, a process referred to as Monte Carlo simulation. This mechanism therefore captures first-order, stochastic uncertainty (i.e., random variability in outcomes between identical patients)25. As a result, some authors have termed this modelling approach as “first-order Monte Carlo simulation”7,14,21 or “individual sampling models without interactions” 23 as individuals strictly progress through the model independently of each other and of environmental constraints. However, similar to Markov cohort models, movement occurs at discrete time cycles19.

Microsimulation provides greater flexibility as individual characteristics and history can be tracked, and can subsequently impact the progression or state value of an individual as they move though the model. Such models are not limited by the Markovian assumption19. In other words, memory and heterogeneity can be embedded easily in this modelling approach through the incorporation of logic statements without requiring additional health states. Similar to Markov cohort models, Markov microsimulations typically simulate a closed population.

**1.4.3 Discrete Event Simulation**

Discrete event simulation (DES) is a flexible modelling technique8,23 adopted from industrial engineering and operations research. This approach is capable of modelling complex behaviour and interactions between individuals and their environment8,9,14. Time is considered continuous, thereby permitting more accurate implementation of continuous risk functions and incorporation of time-to-event data14. As its name suggests, movement in these models is driven by discrete time intervals with time-to-next-event data sampled from parametric or empirical distributions20,26. Although the model is conducted at the individual-level, rather than running individuals one at a time, multiple individuals may be present in the system simultaneously and interacting with each other.

New fundamental concepts introduced in DES include entities, attributes, events, resources and queues. Entities represent objects that evolve through the simulation (e.g., individuals, items). They are defined by a set of assigned attributes and may experience events, consume resources or enter queues over time. Entities may be dynamic or static within a system and have the ability to interact with other entities18,23. Attributes are features specific to each entity. It carries information, such as history or prognostic factors18 that can be modified at any time during the simulation due to the occurrence of an event or due to the length of time between events20. Attributes are a mechanism to introduce memory into the model 18,23. Events are defined as anything that happens over the course of the simulation and extends beyond simply health states in a Markov model as events in DES may correspond to either changes to the patients or the environment. Events may occur, sequentially or simultaneously, and recur26. DES further introduces the concept of resources that represent objects that provide service to an entity. Resources may be described by spatial factors26,27 and may be unconstrained or limited in capacity (i.e., a maximum number of entities can be served simultaneously)18. Consequently, queues can emerge if an entity must wait for an occupied resource and queue behaviour can be defined. As DES explicitly embeds queuing theory, interaction and competition for resources may impact the state of not only the entity involved but also the system as a whole18. DES can model either open or closed populations.

**1.4.4 Agent-based Models**

Agent-based models (ABM) are individual-level simulations adapted from system sciences that are focused on autonomous agents and their interactions with each other and the environment. They model and simulate dynamic systems. The notion is that internal rules governing individual communication and interaction will lead to an emergence of complex system-level behaviour28. Time, in this modelling approach, can either be treated continuously or as discrete units. As an individual-level approach, the model may either run each agent separately (whereby, no interaction exist between agents) or run multiple individuals simultaneously (whereby, interaction could be captured).

ABM introduces the concept of agents that are similar to entities in DES modelling. However, agents have greater flexibility compared to entities as, not only do they have autonomy in communicating and interacting with each other and their environment, agents may also present with the following properties: reactive (i.e., capable of perceiving changes in the environment and respond accordingly); pro-active (i.e., goal-seeking behaviour in order to achieve a goal); social ability (i.e., adept in interacting and communicating with others); and adaptivity (i.e., learning from memory and adapting behaviours based on previous experience)29. Each agent contains information about their state and will respond accordingly to pre-specified rules9. This granularity therefore permits easy incorporation of heterogeneity into agent’s attributes and their rules.

A second important concept is the system’s environment. Agents exist within a network that may adopt a variety of static and dynamic network structures with hierarchical, nested or spatial properties. Therefore, the emerging behaviour may be in response to hierarchical or geographical constraints30. Through the definition of the system’s environment and rules dictating agent’s behaviour, the simulation will lead to an emergence of broader system behaviour. ABM may simulate either open or closed populations.

**1.4.5 System Dynamics**

System dynamics (SD) is another method adopted from system sciences that typically focuses at a cohort level. SD qualitatively visualizes a system through a causal loop diagram consisting of a series of feedback loops where information or change from one point travels through a system and triggers a cascading series of changes that ripple through the system and eventually returns in some form to its point of origin31. The presence of multiple feedback loops can lead to an emergence of complex behaviour. Time is typically handled continuously but may also be discretized18.

SD is quantified by translating the causal loop diagram into stock and flow models31. A system is defined by a set of mutually exclusive and collectively exhaustive stocks that represent sinks where accumulations occur and that define the state and memory of the system. Flows represent the rates at which stocks are drained or replenished, thereby, changing the system over time31. Movement between stocks are based on the rate of flow, and is commonly defined by differential equations*.* Stocks are distinct from the concept of queues in DES as capacity is infinite. SD, if modelled at an aggregate level, encounters the same issues noted by other aggregate-level models in that a necessary assumption is that individuals belonging in the same stock are homogeneous. SD is therefore inherently memory-less. To overcome such limitations, similar strategies proposed in the Markov cohort model section can be applied for SD models.

Therefore, through a series of differential equations, movement between stocks can be described. A feedback mechanism will dictate the rate of change within a system, which is a function of the state of the system itself23. SD may be modelled as either open or closed populations.

**1.4.6 Infectious-disease Compartmental Models (‘SIR’ and its variants**)

Compartmental models, or dynamic transmission models, are used in health economics specifically to model infectious diseases. They rely on differential equations and may be considered a subset of system dynamics models. Compartmental models describe the stages of an infectious disease as compartments with the fundamental states typically being: susceptible, infectious and recovered (SIR). It is classified as a dynamic model since movement of individuals through each compartment varies over time according to the state of the current system.

**1.5 Validity of Economic Modelling**

As summarized above, each modelling approach is based on a different theory of how a system is to be described. Thus, in theory, each approach may differ in its overall performance and its capabilities of capturing a system under study. When constructing or critically appraising the quality and value of a model, validity of a model is a property that must be considered.

Validity reflects the degree to which a model accurately portrays a system, or, in other terms, how well a model reproduces reality32. Given the role of decision-analytic models, a model must be valid in order to produce valid conclusions and predictions that can be trusted when informing decision-making processes. Traditionally, the concept of validity relates to bias and, in most areas of economics, where modelling is typically undertaken using regression-based approaches, there exists a prescribed set of tools for validity testing. Broadly, these can be classified into either: (1) examination of residuals within sample (e.g., test for non-normality, heteroskedasticity and autocorrelation) or, (2) prediction of data points out of sample (e.g., adjusted r2, Akaike’s Information Criteria (AIC), Bayesian Information Criteria (BIC)). These validity tests are possible since regression-based models are estimated from both explanatory and dependent data in which a gold standard comparison is available. However, in decision-analytic economic modelling, such tests are not appropriate as typically these models are constructed to estimate the dependent data (e.g., costs) in the absence of observable data33.

Given that standard approaches to model validity are not appropriate for decision-analytic economic models, several constructs to validity have been proposed, adapted from the field of psychometrics. These include: face validity, internal validity, cross validity, external validity and predictive validity32.

Face validity involves experts’ assessments of the model structure, data sources, assumptions and results. It involves a subjective judgement of how well a model aligns to the existing knowledge of science, to the available evidence and to the clinical or administrative research question32. Face validity helps ensure a model is constructed and used in accordance with the most current medical science and that the best available evidence has been incorporated in order to increase the model’s credibility with experts and enhance users’ confidence in the results32.

Internal validity, known also as verification, refers to whether a model behaves as hypothesized32. To address this, one must assess the extent to which the mathematical calculations are performed correctly and are consistent with the model’s specifications. A two-step process must be followed as individual equations must be verified against their source and, subsequently, its implementation into the code must be assessed for correctness and accuracy32. Coding accuracy can be explored using quality assurance and control methods adapted from software engineering (e.g., structured walk-throughs, double programming, extreme analysis, trace analysis). This type of validity thus seeks to prevent any unintentional computational errors32.

The remaining three types of validity are less frequently reported. Cross validity involves a comparison of a model’s results with other models that analyze an identical or similar problem. Not only does this provide confidence if similar results are calculated by models that have employed different methods, it can also serve a methodological purpose in better understanding the causes to differences in the results of models32. External validity assesses a model’s ability to calculate actual outcomes by simulating real-world scenarios that have already occurred in order to compare the model’s outcomes against known real-world results. Multiple validations may be required to cover the model’s intended application across its expected range of populations, interventions, outcomes and time horizons32. Lastly, predictive validity entails the use of a model to forecast events and, in which, the forecasted outcomes are compared against the actual ones once they unfold. Predictive validity is considered one of the most desirable types of validity as it directly addresses a model’s purpose for prediction and is independent, avoiding data-fitting problems32. However, with both external and predictive validity, a challenge remains in that there is presently few appropriate observational data sets to assess this form of validity.

Validity, itself, is a property that is dependent on a model’s objectives as a model may be valid under one application but not in another. A common thread thus shared by all modelling approaches is their objective to validly represent a decision problem in order to better characterize the consequences of competing treatment strategies34. Advocates for alternative modelling approaches contend that these newer methods have better properties (e.g., flexibility, interactivity) that can better capture important aspects of a system and, thus, produce more valid estimates of the economic value of competing health technologies. For instance, alternative approaches to modelling can account for certain complexities such as: patient history, resource/capacity constraints, health system efficiency and individual’s interactions with others and their surrounding environment23; thereby, being more representative of real-world conditions. In contrast, others have argued that models should be kept as simple as possible to aid in understanding the underlying system35. The question therefore arises when selecting a modelling approach: does it truly matter? And if yes, under what circumstances?

**1.6 Outline of the Thesis**

This is a sandwich thesis of three papers that explores the evidence surrounding the role of different approaches to decision-analytic modelling in the context of healthcare decision-making. The overall aim is to better understand the settings or circumstances whereby each modelling approach would be most suitable for the conduct of health care decision-analytic economic modelling. This is an important endeavour for two reasons: 1) the approach to modelling is rarely justified in current health economic papers; and 2) there is little guidance on how newer modelling approaches should be used and applied in health-care decision-analysis. A more consistent and better understanding of when a particular modelling approach should be used may improve model validity in order to ensure decision-analytic models are addressing their purpose to predict the best course of action amongst a set of alternatives. The first paper compared the similarities and differences between theoretical frameworks that have offered prescriptive guidance on how a modelling approach should be selected (Chapter 2). The second paper systematically identified empirical cross-validation papers published to date, that have addressed the question of whether different modelling approaches produce different quantitative results and qualitative conclusions (Chapter 3). The last paper conducted a cross-validation exercise that compared a decision tree with two modelling approaches less frequently implemented (i.e., agent-based and system dynamics models) using an example comparing two vaccines for childhood influenza (Chapter 4). The final chapter discusses the main conclusions based on the work presented in this thesis as well as directions for future research (Chapter 5).

In Chapter 2, a systematic literature review on the theoretical frameworks and taxonomies that offer guidance to modellers in addressing how to select the most suitable modelling approach for the conduct of decision-analytic health economic evaluations is presented. Eight frameworks were identified in which a multitude of selection criteria were noted. These factors could be classified into either structural features (i.e., objective facts underpinning the fundamental theories of a specific modelling approach) or practical considerations (i.e., subjective, context-dependent elements that arise from the given research question and a modeller’s expertise). This paper is unique in that it systematically reviewed and compared these selection frameworks for the first time. Specifically, given that structural features are objective in nature, the expectation would be that considerable agreement amongst frameworks exist in how the structural features of a modelling approach are defined. However, it was found that there was poor agreement in how frameworks classified the structural features corresponding to each modelling approach.

Given the concerns brought forth in Chapter 2, Chapter 3 addressed the same question on how to differentiate amongst the many approaches available in decision-analytic modelling, but, by employing a different lens. Instead of guidelines and frameworks, we explored this issue from an empirical perspective through the work of published cross-validation studies. Cross-validation is an aspect of model validity in which two or more modelling approaches are compared for the same decision problem. Traditionally, the value of cross-validation has been in comparing models with a low degree of dependency (e.g., those that share little commonality by using alternative mathematical models and selecting model parameters from different sources) as a means to address the impact of structural uncertainty32. However, the contrary-highly dependent models, can also be meaningful as it can highlight situations or scenarios whereby the results may be divergent or convergent depending on the approaches to modelling. In order words, mathematical models that share identical parameters and assumptions but that produce different results can highlight any inherent differences that exist between modelling approaches.

A systematic literature review was conducted to identify all published cross-validation studies. Included papers were classified according to the selection criteria, defined in Chapter 2. Most work on cross-validation has focused on two structural features to modelling: population resolution, and interactivity. However, the empirical work demonstrated consistent findings on the impact of the structural feature on the practicalities of model development and its subsequent model output (i.e., quantitative results and qualitative policy conclusion).

Despite consistent results observed across published cross-validation exercises, gaps in the literature remained as certain modelling approaches were rarely or never compared (e.g., SD, ABM). As a result, Chapter 4 compared the results of both a SD and an ABM with the results of a previously constructed decision tree that assessed the economic value of different childhood influenza vaccines. The ABM was constructed at an individual level whereas, the SD model was at a cohort level. Thus, the factor of population resolution could be addressed. Similarly, both SD and ABM represent a form of dynamic modelling whereas, the decision tree is a static model; thereby, allowing the factor of interactivity to also be addressed.

Model calibration was successful as, under the same set of assumptions, all three approaches produced similar clinical and cost-effectiveness estimates. In terms of the impact of the resolution of a model, the individual-level ABM was found to be more flexible as it relied on fewer simplifying assumptions but this was at the practical expense of requiring more time and data. ABM, with its individual-level resolution, was found to produce different cost-effectiveness conclusions when different network topologies were modelled. With respect to interactivity, the cost-effectiveness results remained similar when strict assumptions were imposed. The clinical estimates were found to be highly sensitive to the proportion and the schedule of vaccination under both dynamics models. This work corroborates, to an extent, the systematic review on empirical cross-validation work but further highlights the flexibility of ABM compared to any existing approaches in modelling as it can capture the impact of network topologies. Furthermore, this work illustrates one of the first attempts to use ABM in the field of health economics in an empirical example.

Finally, in Chapter 5, a summary of the key findings, limitations and implications of this thesis, and directions for future research are presented. This thesis therefore summarizes the existing theoretical and empirical work done so far to distinguish between the many approaches available in decision-analytic health economic evaluations. The decision of which modelling approach to select involves considerations on both the objective structural features and the subjective practicalities to modelling. Despite a lack of agreement between frameworks on how to classify the features to describe a particular modelling approach, the empirical evidence has been shown to be consistent. The impact of a particular structural feature on the choice of a modelling approach and, consequently, on the results generated by such a model have been consistent - even across different disease areas. This remained true even in our cross-validation exercise, the first ever that compared agent-based and system dynamics modelling across the criteria of population resolution and interactivity. Research on this topic remains in its infancy, as other technical features remain unexplored. Despite this, the consistency observed from empirical work of cross-validation studies suggests that the direction forward should be focused on evidence-based guidelines and to move away from current existing heuristic-based guidelines. This is an important and necessary discussion as the validity of any model, and ultimately the decision in which it seeks to inform, is dependent on proper specification of the modelling approach such that it can properly capture the features that are necessary to that particular decision problem.

**1.7 References**

1. McKibbon A, Wilczynski N. PDQ Evidence-based Principles and Practic*e.* People's Medical Publishing House-USA, Shelton, USA; 2009.

2. Sorenson C, Drummond M, Khan BB. Medical technology as a key driver of rising health expenditure: disentangling the relationship. *Clinicoecon Outcomes Res* 2013; 5, 223-34.

3. Barros PP. The black box of health care expenditure growth determinants. *Health Econ* 1998; 7(6), 533-544.

4. Okunade AA, Murthy VN. Technology as a ‘major driver’of health care costs: a cointegration analysis of the Newhouse conjecture. *J Health Econ* 2002; 21(1), 147-159.

5. Cutler DM, McClellan M. Is technological change in medicine worth it? *Health Aff* 2001; 20(5), 11-29.

6. Skinner JS, Staiger DO, Fisher ES. Is technological change in medicine always worth it? The case of acute myocardial infarction. *Health Aff* 2006; 25(2), w34-w47.

7. Kreke JE, Schaefer AJ, Roberts MS. Simulation and critical care modeling. *Curr Opin Crit Care* 2004; 10(5), 395-398.

8. Fone D, Hollinghurst S, Temple M, Round A, Lester N, Weightman A, Roberts K, Coyle E, Bevan G, Palmer S. Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *J Public Health Med* 2003; 25(4), 325-335.

9. Kim S-Y, Goldie SJ. Cost-effectiveness analyses of vaccination programmes. *Pharmacoeconomics* 2008; 26(3),191-215.

10. Coyle D, Lee KM, O’Brien BJ. The role of models within economic analysis. *Pharmacoeconomics* 2002; 20(1),11-19.

11. Cohen JT, Neumann PJ. Decision analytic models for Alzheimer's disease: state of the art and future directions. *Alzheimers Dement* 2008; 4(3), 212-222.

12. Box GEP, Hunter, W.G., Hunter, J.S. Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*.* John Wiley & Sons, New York, USA; 1978.

13. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modeling practices: A suggested framework. *Value Health* 1998; 1(2), 131-147.

14. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a model A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. *Med Decis Making* 2012; 32(5), 678-689.

15. Kim S-Y, Goldie SJ, Salomon JA. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. *Med Decis Making* 2010; 30(5), E1-E28.

16. Afzali HHA, Karnon J, Merlin T. Improving the accuracy and comparability of model-based economic evaluations of health technologies for reimbursement decisions: A methodological framework for the development of reference models. *Med Decis Making* 2013; 33(3), 325-332.

17. Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. *J Health Serv Res Policy* 2005; 10(4), 245-250.

18. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment. *Pharmacoeconomics* 2008; 26(2), 131-148.

19. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3. *Value Health* 2012; 15(6), 812-820.

20. Cooper K, Brailsford S, Davies R. Choice of modelling technique for evaluating health care interventions. *J Oper Res Soc* 2007; 58(2), 168-176.

21. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004; 9(2), 110-118.

22. Yi Y, Philips Z, Bergman G, Burslem K. Economic models in type 2 diabetes. *Curr Med Res Opin* 2010; 26(9), 2105-2118.

23. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006; 15(12), 1295-1310.

24. Le Lay A, Despiegel N, François C, Duru G. Can discrete event simulation be of use in modelling major depression? *Cost Eff Res Alloc* 2006; 4(1): 19.

25. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health* 2012; 15(6), 835-842.

26. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–4. *Med Decis Making* 2012; 32(5), 701-711.

27. Caro JJ. Pharmacoeconomic analyses using discrete event simulation. *Pharmacoeconomics* 2005; 23(4), 323-332.

28. Conte R, Hegselmann R, Terna P. Simulating Social Phenomena. Springer, Berlin, Germany; 1997.

29. Nilsson F, Darley V. On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings: Experiences from a packaging company. *Int J Oper Prod Manage* 2006; 26(12), 1351-1373.

30. Commendatore P, Kayam S. Complexity and Geographical Economics: Topics and Tools. Springer International Publishing, Switzerland; 2014.

31. Sterman JD. Business Dynamics: System Thinking and Modeling for a Complex World. Irwin/McGraw-Hill, Boston,USA; 2000.

32. 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. *Med Decis Making* 2012; 32(5), 733-743.

33. McCabe MC, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000; 17(5), 501-513.

34. Drummond MF, Sculpher MJ, Torrance GW, O’Brien B, Stoddart G. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press, Oxford, UK; 2008.

35. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray M. Modelling in ecomomic evaluation: an unavoidable fact of life. *Health Econ* 1997; 6(3), 217-227.

# CHAPTER 2:

# Systematic narrative review on DECISION FRAMEWORKS to select the appropriate modelling approachES for HEALTH ECONOMIC evaluations

Tsoi B BSc, MSc, PhD(cand)1,2ŧ, O’Reilly D BSc, MSc, PhD1-3, Jegathisawaran J BMSc, MHEcon,1,2 Tarride J-E BA MA, PhD1 Blackhouse G BA, MBA, MSc1,2, Goeree R BA, MA1-3

1. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

2. PATH Research Institute, St. Joseph’s Healthcare Hamilton, Ontario, Canada

3. Centre for Evaluation of Medicines (CEM), St. Joseph’s Healthcare Hamilton, Ontario, Canada

ŧ **Corresponding author**

Cited as: Tsoi et al. Systematic narrative review of decision frameworks to select the appropriate modelling approaches for health economic evaluations. 2015. *BMC Res Notes:* accepted.

**Abstract**

Background: In constructing or appraising a health economic model, an early consideration to modelling is whether the selected approach is appropriate for the given decision problem. Frameworks and taxonomies that distinguish between modelling approaches can help make this decision more systematic and this study aims to identify and compare the decision frameworks proposed to date on this topic area.

Methods: A systematic review was conducted to identify frameworks from peer-reviewed and grey literature sources. The following databases were searched: OVID Medline and EMBASE; Wiley’s Cochrane Library and Health Economic Evaluation Database; PubMed; and ProQuest.

Results: Eight decision frameworks were identified, each focused on a different set of modelling approaches and employing a different collection of selection criterion. The selection criteria can be categorized as either: i) structural features (i.e., technical elements that are factual in nature) or ii) practical considerations (i.e., context-dependent attributes). The most commonly mentioned structural features were population resolution (i.e., aggregate vs. individual) and interactivity (i.e., static vs. dynamic). Furthermore, understanding the needs of the end-users and stakeholders was frequently incorporated as a criterion within these frameworks.

Conclusions: There is presently no universally-accepted framework for selecting an economic modelling approach. Rather, each highlights different criteria that may be of importance when determining whether a modelling approach is appropriate. Further discussion is thus necessary as the modelling approach selected will impact the validity of the underlying economic model and have downstream implications on its efficiency, transparency and relevance to decision-makers.

K**ey words:** Decision analysis; health economic evaluation; systematic review; decision trees; state-transition models; Markov model; microsimulation; agent-based models; system dynamics; compartmental models

**2.1 Background**

The use of decision-analytic modelling to estimate the cost-effectiveness of health care interventions is becoming widespread to inform health policy decision-making. A model, referred to in this article, is defined as the use of analytical methodology to quantitatively compare health technologies. Models may have a range of uses including extrapolating from primary data sources and transferring results from one jurisdiction to another1. By incorporating event probabilities, resource utilization, costs and patient outcomes, a model synthesizes the data to identify the best option for decision-makers.

However, with the growing reliance on economic evaluations to support decision-making, concerns have risen on the validity, reliability and comparability of the results generated from such models1. To respond to these criticisms, the research community has focused considerable efforts in setting best practice guidelines for the development and conduct of health economic models. This is evident from the guidelines published by respective health technology assessment (HTA) agencies (e.g., Canadian Agency for Drugs and Technologies in Health (CADTH)2) and from non-profit research organizations (e.g., International Society For Pharmacoeconomics and Outcomes Researc*h* (ISPOR)3-5).

With the growing diversity of modelling approaches available (Table 1), a particular question is the relative merits of each approach in health economic modelling. Historically and still to date, decision trees and Markov cohort models are the most commonly used approaches in economic evaluation. However, due to their limitations, awareness has grown on alternative modelling approaches. Guidance documents recommend transparent reporting of a modeller’s rationale for selecting a model type, although it may not always be clear which approach would be most suitable for a given decision problem. This is an important issue since each approach can introduce constraints to a model’s development and its conceptualization in terms of what elements can be captured and the ease to which they can be incorporated into the model6. This may lead to a different focus on the decision problem and, thereby, generate conflicting results and diverging policy recommendations7.

To provide guidance on how to select a particular modelling approach, frameworks have emerged that categorize and distinguish between them. However, no attempt has been made to compare and contrast these frameworks. The purpose of this paper is therefore to conduct a systematic literature review to identify and critically appraise these published frameworks.

**2.2 Methods**

**2.2.1 Search Methods**

A literature search was performed for articles published up to January 21, 2014 with the following bibliographic databases searched: OVID Medline (1946-present; In-Process & Other Non-Indexed Citations) and EMBASE (1996-present); Wiley’s Cochrane Library (Issue 1 of 12, Apr 2014) and Health Economic Evaluation Database; PubMed (for non-Medline records); and ProQuest Dissertations. Controlled vocabulary terms, such as the National Library of Medicine’s Medical Subject Headings, and keywords were used to construct the search strategy (Appendix 2.I). The electronic search was supplemented by cross-checking the bibliographies of relevant publications and grey literature searches (e.g., working papers, commissioned reports, policy documents, websites).

**2.2.2 Selection of Relevant Articles**

Records were screened for inclusion based on the pre-defined criteria presented in Appendix 2.II. To be included, a paper had to describe, in whole or in part, a decision framework (e.g., algorithm, taxonomy) on how to select between economic modelling approaches in the context of health care policy decision-making. Studies were limited to those published in English.

The titles and abstracts of the records identified from the bibliographic search were initially screened for relevance by one reviewer (B.T.) with a 50% random check conducted by a second independent reviewer (J.J.). If either reviewers identified a citation as being potentially relevant, its full-text was obtained. In the second phase of screening, one reviewer (B.T.) assessed the full-text version of all included articles, with a second independent reviewer (J.J.) completing a 50% random sample. Any discrepancies at this stage were resolved through discussion and consensus.

**2.2.3 Data Extraction**

A standardized data abstraction form was developed to extract data from the relevant studies. The form captured: bibliographic information (e.g., author, year); framework type (e.g., flow-chart, table); framework description, including its selection criteria; and the main conclusions. Their evolution and history, if discussed, was further noted.

The selection criteria specific to each decision framework were identified. These had to be present within the framework; criteria that were simply mentioned in the paper but not explicitly incorporated into the framework were excluded. These criteria were separated into either structural features or practical considerations. Structural features were defined as those relating to principles or theories behind a model. These are the technical elements that lay bare the intricacies of modelling concepts and the nature of the decision problem will dictate the structural features desired within a model. Practical considerations are defined as elements that impact the effectiveness or feasibility of developing and constructing a model and are, to a degree, context-dependent.

**2.2.4 Data Analysis**

Data were analyzed and synthesized with the intent to:

1. Understand the evolution of the frameworks;
2. Tabulate and identify the frequency to which selection criteria were discussed across these frameworks;
3. Evaluate the extent to which the frameworks agree or disagree on the structural features specific to each modelling approach.

**2.3 Results**

Of the 3,342 unique publications identified from the literature search, eight met the full inclusion criteria (Figure 1). Most studies were excluded either because it made no mention to decision-analytic modelling or it did not present a selection framework to guide the choice between modelling approaches. Overall, the agreement between independent reviewer for study inclusion was considered moderate (Cohen’s kappa: 0.60).

Table 2 provides an overview of the decision frameworks in terms of the country of publication, the framework’s focus and the modelling approaches that were covered. All decision frameworks were published in the past ten years with two specific to infectious diseases6,8 and the remainder being generic/ non-disease specific7,9-13. Each framework covered different model types, although all of them involved a decision between a traditional modelling approach (i.e., decision tree and Markov cohort model) and one or more alternative approaches (e.g., discrete event simulation; agent-based model; system dynamics).

Decision frameworks were visually represented by flow charts6,9,12,13, radar graphs7, or tables8,10,11 and Table 3 further details the selection criteria that were considered within each framework. The definition of common structural features and practical considerations that were identified from this review of frameworks are presented in Table 4. The structural elements include: the resolution of the population; the capture of first-order uncertainty; the nature of the interactions; the handling of resource constraint; and the dimension of time. From Table 3, it was found that the most common structural features considered amongst these frameworks were interactivity (i.e., static vs. dynamic) and population resolution (i.e., aggregate vs. individual) (n=6/8; 75%), followed by how time is handled (n=4/8; 50%) (Table 3). Practical considerations (Table 4b) were explicitly included within most flowcharts and the most common practical consideration were the end-user requirements and simplicity (n=3/8; 37.5%) (Table 3).

Below, a narrative summary of each framework is presented. A copy of each decision framework can be further found in Appendix 2.III**:**

**2.3.1 Generic Frameworks**

The first paper within the health care field on this topic was by Barton et al.9. Based on the following four criteria, their flowchart assists in the selection between decision tree, Markov cohort model, Markov microsimulation, discrete event simulation and system dynamics: (i) interactivity – importance of capturing interaction between patients; (ii) population resolution – the necessity of individual-level modelling; (iii) validity – the adequacy of pathways represented by a decision tree; and (iv) simplicity – the number of states required in a Markov cohort model. The authors highlight the trade-off between simplicity and clinical validity. They recommend a more complex and computational-demanding model only if it provides a more accurate representation of the decision problem and leads to more valid results9. Simplification, according to the authors, may involve fixing one or more parameters in the model and two conditions may justify such a practice: when the results are robust to variation with that particular set of parameters or if the parameter is derived from good and accurate data.

Brennan et al.10 have proposed a taxonomy table describing the relationships between modelling approaches according to their structural features. The columns in their taxonomy highlight the assumptions of population resolution; expected value/memory and first-level uncertainty while the rows describes the interaction between individuals and the handling of time. Each cell in the table lists the modelling approach with those corresponding structural features. Some of the model structures described in this taxonomy can be considered subclassification of specific modelling approaches. For instance, depending on the dimension of time and first-order uncertainty, system dynamics was separated into finite difference equation system dynamics, ordinary differential equation system dynamics, discrete time Markov chain model and continuous time Markov chain model. The authors state that the identification of health states and risk factors, and their underlying relationships should precede the selection of a modelling approach. If multiple approaches are suitable, the simplest model that accurately addresses the decision problem should be chosen with further consideration on practical factors such as software availability, implementation skills, time constraints and end-user requirements10.

Chick11 simplifies Brennan’s proposed taxonomy by removing the rows specific to the dimension of interactivity; thereby, reducing the subclassification of certain modelling approaches seen in Brennan et al.’s original taxonomy (i.e., microsimulation, system dynamics). However, it remains unclear why, for one set of features (i.e., stochastic Markovian individual discrete time), the cell is empty in Chick’s framework and is not associated with any particular modelling approach11. For this particular set of features, Brennan et al.’s described the model structure as: ‘discrete-time individual event history model’.

Similarly, Heeg et al.7 adapts Brennan’s10 framework. However, rather than using a taxonomy table, they displayed their framework as a radar diagram that ranks the relative ability of decision tree, Markov cohort model, discrete event simulation and Markov microsimulation in addressing a collection of selection criterion—including practical considerations. Each spoke on the radar diagram represents a particular selection criterion and modelling approaches that are better at addressing that criterion appear further away from the origin of the radar diagram. Their framework incorporates all of the technical features proposed by Brennan although different terminologies are employed: ‘randomness’ is now referred to as ‘variability’ (i.e., first order uncertainty) while ‘expected value’ is referred to as ‘memory’7. An additional technical feature included is the interaction due to covariates and nonlinear associations between individual risk factors and outcomes. The following practical considerations were also included in their framework: time (i.e., to collect data, build and simulate the model); experience and validity (i.e., clinical representativeness) (Table 3)7.

An independent framework developed by Cooper et al.12 similarly intertwined practical and structural considerations (Table 3) to help guide the decision between Markov cohort model, decision tree and discrete event simulation. The authors state that the nature and the complexity of the disease, and the health care intervention, may influence which structural features to consider (e.g., interaction between individuals; queuing and resource constraints)12. Rather than considering population resolution explicitly as a structural feature, this framework mentioned the impact of dimensionality in terms of the differences in time required to build and conduct simulations between aggregate-level and individual-level models. Outside of their framework, the modeller’s experience and data availability were additional factors that, together, may impact the speed and the ease of model development. The authors recommend that the analysis should be built based on the simplest model that can adequately address the research question12. A unique trait in Cooper’s framework is that it recognizes that modelling may not always be possible, and further incorporates an ‘abandon’ scenario when it is futile to pursue modelling given the disconnect between practical constraints and the desired technical attributes (e.g., significant heterogeneity and/or when queuing or interaction between individuals is important). In such cases, when the practical elements and the structural features conflict, construction of a model should be stopped until such issues are resolved12.

By moving through a series of decisions pertaining mainly to the desired structural features, Stahl’s13 hierarchical flowchart filters the choice of modelling approaches down to one to two suitable ones. Similar to Cooper12, Stahl also advocates that simplicity should be a guiding principle - referring to it as, ‘keep it simple stupid (KISS)’ with a model only as complex as necessary for the question(s) of interest13.

**2.3.2 Infectious Disease Specific Frameworks**

Brennan et al.’s framework10 was modified by another group of researchers for the evaluation of vaccines. Models were categorized according to three structural features: population resolution; first-order uncertainty; and interactivity8. As the selection criteria are dichotomous, eight possible categories exist (n=23) although only six categories were linked to modelling approach(es) as some combinations were deemed unrealistic. Kim et al. further recommend that model choice should be based on not only the nature of the decision problem (e.g., research question, natural history and features of the disease) but on practical concerns such as data availability, an analyst’s experience and time8.

The last framework, by Jit et al.6, utilized a series of questions organized into a flowchart to highlight the key distinctions between static (referred to as cohort models) and dynamic models in the context of infectious disease modelling. According to the authors, infectious diseases have several complexities that make it unique compared to other illnesses: transmissibility (i.e., interaction between infected and susceptible individuals); natural immunity; and the epidemiology of the illness (i.e., an infection proceeds through several stages, such as: susceptibility, latency/incubation, infectious/ symptomatic and recovery)6. These distinctions result in the need for dynamic modelling when the force of infection is not constant over time. Instances include if an intervention changes the profile of the infected individuals (e.g., increase pathogenicity or transmissibility by shifting the age profile of the disease) or induces selective evolution on a subset of the organisms (e.g., antibiotic resistance)6.

**2.3.3 Consistency between Decision Frameworks**

Given that several decision frameworks were identified, it was of interest to assess the concordance in the frameworks’ recommendations. To conduct this, the structural features were evaluated across frameworks to assess their consistency in how they categorize each modelling approach in terms of their structural traits. As previously mentioned, structural features are expected to remain the same across decision frameworks for each modelling approach since they are based on theories and facts.

Table 5 presents the degree to which the decision frameworks are consistent in how they classify the structural assumptions specific to each modelling approach. As only two frameworks included agent-based models, both agreed that it is an individual-level approach that can incorporate interactions. System dynamics was seen as an aggregate-level approach that could handle interactions. Amongst the frameworks that do discuss the mechanism of time, system dynamic was considered able to model at a discrete unit or continuously although their capacity for handling resource constraints has yet to be addressed. Markov microsimulations have been characterized by the majority of the frameworks as an individual-level approach with time handled discretely or continuously. Few frameworks have addressed first-order uncertainty and the capability of Markov microsimulations in handling resource constraints except for one that suggested that microsimulations can assume unlimited resources7. Disagreement between frameworks remained on whether it is capable of handling interaction. For discrete event simulation, of those that addressed resource constraints and first order uncertainty, they all agreed on its capacity to incorporate resource constraints and that it is stochastic. The majority considered discrete event simulation as being capable of handling interactions between patients. However, discrepancies lay on how to classify the resolution of such models. For compartmental models, only the features of population resolution, first order uncertainty and interactivity have been discussed so far with the sole agreement being that this approach can incorporate interactions (Table 5).

For traditional modelling approaches, an even greater degree of disagreements was observed in how structural features were specified. Most frameworks did not discuss the notion of resource constraints for decision trees. Of the frameworks that describe the dimension of time and interactivity, they were consistent in characterizing decision trees as static, fixed time horizon (i.e., “untimed”) models. However, for the remaining two structural features (i.e., population resolution, first-order uncertainty) evaluated, less clarity emerged. For Markov cohort models, as per its name, the frameworks all agreed that this modelling approach is not an individual-level modelling approach but rather focused at the aggregate-level. Markov cohort models were considered not capable of handling interaction or resource constraints in most except in two of the frameworks (Table 5)7,12.

**2.4 Discussion**

Despite the prevalence in the use of traditional modelling approaches to conduct health economic evaluations, these frameworks all highlight the need for alternative modelling approaches under certain circumstances. For instance, discrete event simulation permits explicit incorporation of queuing theory and may be suitable if the question partly involves resource constraints. Agent-based models, on the other hand, can integrate agent-to-agent interactions and are thus suitable when behavior is considered an important characteristic with the problem at hand (e.g., infectious disease modelling). Indeed, it may be safe to extend that there is no single modelling approach that is capable of answering all types of research questions. HTA agencies and other policy organizations that rely on economic modelling to guide reimbursement and resources allocation decision-making must therefore develop the capacity to construct and critically appraise models outside of what is considered the traditional modelling approaches.

Although several frameworks have been published to distinguish between modelling approaches, there is no clear over-arching or universally-accepted one. Each framework has, in fact, highlighted different selection criteria that may be of importance when choosing the most-suitable approach. A recurring theme that emerged across these frameworks is the necessity for the approach to reflect the underlying theory of the health condition and the characteristics of the health technologies being compared. The modelling approach selected should align with the purpose of the model and the level of detail desired with minimal complexity 4.

However, Table 5 highlighted a concerning observation: there is a general lack of agreement between the decision frameworks on how the structural features specific to each modelling approach are described. This suggests that, by using different frameworks, one may come to a different decision on what constitutes the most appropriate modelling approach. For instance, consider a model that is interested in exploring the cost-effectiveness of therapies for lowering blood pressure in patients with essential hypertension in terms of the prevention of cardiovascular and cerebrovascular events. The model aims to simulate a cohort of patients with heterogeneous characteristics and, given the existing understanding of hypertension, the model must capture the impact of different risk factors on the development of clinical events as these risk factors evolve over time. Additional factors to consider for this decision problem is that resources will be assumed unlimited and that time will be handled discretely. Employing the six generic frameworks without consideration of the practical constraints, we find that four frameworks7,9-11 advise for a Markov microsimulation, one framework13 recommends a Markov or discrete event simulation without specification on whether the Markov model is an aggregate-level or an individual-level model (i.e. microsimulation) while the last specifies for a discrete event simulation or a Markov cohort model12. Other cases exist of applying these frameworks to a decision problem and encountering different recommendations in terms of which modelling approach would be recommended.

It may be that, in certain cases, the frameworks do not entirely represent what the authors would consider as best practice, but rather what is recommended and accepted practice in the jurisdiction in which they work (i.e., many countries now have national reimbursement bodies that provide guidelines on economic modelling and may influence the choices of how the researchers in those countries developed their framework). Indeed, different frameworks were found to address different sets of modelling approaches (Table 2). It would be expected that frameworks would characterize the structural features specific to a particular modelling approach similarly although this was not observed (Table 5).

Despite this, another consistent recommendation emerged from these studies in that the decision of which modelling approach to select is dependent not only on the structural assumptions but often also on the practical considerations. It is rarely possible to consider one without the other. Even amongst the frameworks that solely incorporated structural features6,8,10,11, half included a separate discussion on the practical considerations to modelling8,10. The selection of the appropriate modelling approach is therefore iterative. The clinical research question (i.e. characteristics of the disease and its intervention) dictates which structural features are important. This filters down the range of suitable modelling approaches and subsequently, practical elements such as simplicity, computational efficiency, end-user requirements and transparency may impact the decision on the best-suited modelling approach.

One unresolved question remains: the trade-off between simplicity and validity. In most of the frameworks and in other broad economic evaluation guidelines, the majority support the notion that the model structure should be kept as simple as possible13,14. Barton and colleagues mention that more complex models are only justifiable when the increased complexity leads to more valid results9. Another interpretation to the above recommendation is that, when selecting a simpler technique, a modeller should ensure that any error incurred from omitting certain aspects of the disease and its intervention will not materially bias a study’s results15. But, how much simplification is possible without compromising a model’s validity? Unfortunately, this is not a straightforward issue as it is based on several factors including the nature of the decision problem (i.e. clinical condition and the treatment alternatives being modelled) and several practical considerations (e.g. available data, time and budget)16. Greater research and education is thus necessary for both modellers and decision-makers to better characterize and understand the implications of such a trade-off.

The observed discrepancy observed between frameworks in the recommendations they provide on which modelling approach is appropriate leads to the question of whether selecting different modelling approaches do in fact impact the model’s results and conclusions? When does it truly matter which modelling approach is used? For instance, to what extent does patient heterogeneity have an impact such that a Markov microsimulation or a Markov cohort model would produce diverging results? Similarly, to what extent does queuing and constrained resources impact the cost-effectiveness of an intervention such that it warrants the need for a discrete event simulation? These frameworks were all found to lack a sufficient evidence-base as most were based on general heuristics. A means to answer the above questions empirically would be to assess a model’s validity. One approach, based on the concepts of cross-validation, would be to compare the results between highly-dependent models that employ different modelling approaches to otherwise address the same research problem by using the same data parameters and share common assumptions. Such exercises may inform when it empirically matters whether a particular modelling approach is selected and some of the early pioneers in such activities include the Mount Hood Challenge for diabetes modellers17.

A recent systematic review was published focused on cross-validation work in health economic models, evaluating the impact of structural features on the choice of the modelling approach18. Population resolution was found to have minimal impact empirically as both aggregate- and individual-level models generated nearly identical results. Rather, consideration on this structural feature was relevant in terms of a practical trade-off between validity and feasibility (e.g., individual-level models required fewer simplifying assumptions, thus increasing its face validity but at the expense of being more time- and data-intensive; and vice versa). In terms of the criterion of interactivity, infectious-disease models have consistently showed that, depending on the assumptions regarding the probability of disease exposure, dynamic and static models will produce dissimilar results and lead to opposing policy recommendations18*.* Further research in this area is still required as it may provide the evidence that is necessary to better guide the development of evidence-based decision frameworks.

One challenge that arose over the course of this study was the heterogeneity in the terminologies employed to describe the modelling approaches. For instance, for Markov cohort model, Chick’s11 framework used the term “finite difference model” while the original framework by Brennan referred to it as “simulated Markov model”13. This was even more evident for Markov microsimulation as it was referred to by a wide range of terms including: “individual sampling model”9,10, “patients evolve on discrete time grid”11, “patient-level simulation”11, “Monte Carlo Markov models”13, “Monte Carlo simulation/ microsimulation”8 and “First-order Markov model”7. This is concerning as continued use of unclear and inconsistent terminologies can hamper communication between modellers and mislead understanding on these frameworks. It is possible (and we acknowledge) that the differences observed between frameworks may not only lie with their recommendations but may also be partly due to differences in their semantics. Given the cross-disciplinary nature of this field, greater effort is necessary to standardize the terminology that is being used. Some excellent work has indeed emerged from ISPOR-SMDM good research practice guidelines3-5 although much remains to be done.

A limitation with this study is that it focused mainly on literature from the health care context. As previously mentioned, the modelling approaches used in health economic evaluations originated from the fields of mathematics, operations research and industrial engineering. Consequently, a vast and rich source of literature is likely to exist within those fields that have not been included in this study. By not including studies from other disciplines, this paper may not capture the decision frameworks outside of health care. We acknowledge that this is a limitation to this study although it was necessary to restrict the literature search within the field of health to capture the decision criteria that are specifically relevant to the health context.

**2.4 Conclusions**

To reiterate, the aim of this systematic review was not to propose a new framework that unifies the existing frameworks or to provide support towards a single one. Rather, this review was intended to identify and critically appraise the collection of decision frameworks that are currently available to health economic modellers and their users. Although most were developed independently, at a minimum, all frameworks were found to involve a comparison of the structural features as a means to distinguish between the approaches. Nearly all frameworks considered the criteria of population resolution and interactivity; which may perhaps be indicative as the absolute minimum needed to be considered when selecting a modelling approach. Furthermore, most authors explicitly considered or discussed the practicalities to modelling as part of their framework. Emerging from this review, we find that the process of selecting an appropriate approach for health economic models involves the consideration of multiple criteria. One must not only align the nature of a given decision problem with the structural features of a modelling approach; practical constraints that are context-dependent must further be examined.

Although decision frameworks are intended to provide a systematic and transparent approach in which to pursue the question of which modelling approach should be chosen, this review found a concerning lack of agreement between frameworks in terms of how structural elements are classified. Thus, by employing different frameworks, different recommendations may emerge. In this case, the use of decision frameworks may provide a false sense of confidence that the appropriate methods were employed for the conduct of an economic evaluation and blindly shut down any further debate on the process and the considerations for which a modelling approach was chosen. It is intended that this work will hopefully open dialogue between researchers and policy-makers in terms of providing or requiring greater transparency on how a particular modelling approach is selected. Until better agreement exists amongst frameworks or more empirical research is conducted, we strongly recommend that modellers properly and transparently justify why a particular modelling approach was selected over the others. The choice of a modelling approach is an important and necessary step to any health economic modelling exercise with broad implications on the subsequent model development and evaluation. Given its potential impact on a model’s validity, the choice should be carefully considered, debated and reported.

**Competing Interest**

Bernice Tsoi is supported through an Award from the Father Sean O’Sullivan Research Centre, St. Joseph’s Healthcare Hamilton and Pfizer Graduate Student Award in Health Technology Assessment. The authors have no other potential relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

**Author Contributions:**

BT and RG conceived of this study with all authors participating in its design. BT further conducted the search, performed both level I and II screenings, analyzed the data and drafted the manuscript. JJ performed the role of the second reviewer in conducting the 50% random screen at both level I and II. RG,DO, JET and GB assisted in the interpretation of the data. All authors read and approved the final manuscript.

**Acknowledgements:**

We would like to thank Kaitryn Campbell for her assistance in developing the search strategy for this systematic review.

Reference List

(1) Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, Vray MP: Modelling in economic evaluation: an unavoidable fact of life. *Health Econ* 1997, 6:217-227.

(2) Canadian Agency for Drugs and Technologies in Health: G*uidelines for the Economic Evaluation of Health Technologies* 3rd Ed*.* Ottawa; 2006.

(3) Caro JJ, Briggs AH, Siebert U, Kuntz KM: Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health* 2012, 15:796-803.

(4) Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M: Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2. *Value Health* 2012, 15:804-811.

(5) Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, Kuntz KM: State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Value Health* 2012, 15:812-820.

(6) Jit M, Brisson M: Modelling the epidemiology of infectious diseases for decision analysis: a primer. *Pharmacoeconomics* 2011, 29:371-386.

(7) Heeg BM, Damen J, Buskens E, Caleo S, de CF, Van Hout BA: Modelling approaches: the case of schizophrenia. P*harmacoeconomics* 2008, 26:633-648.

(8) Kim SY, Goldie SJ: Cost-effectiveness analyses of vaccination programmes: a focused review of modelling approaches. *Pharmacoeconomics* 2008, 26:191-215.

(9) Barton P, Bryan S, Robinson S: Modelling in the economic evaluation of health care: selecting the appropriate approach. J *Health Serv Res Policy* 2004, 9:110-118.

(10) Brennan A, Chick SE, Davies R: A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006, 15:1295-1310.

(11) Chick SE: Taxonomy of model structure for health economics [http://www2.wmin.ac.uk/hscmg/qmmhealth2007/talks/Chick\_S.-.IMAHealth2007.-.Keynote.pdf]

(12) Cooper K, Brailsford SC, Davies R: Choice of modelling technique for evaluating health care interventions. *J Oper Res Soc* 2007, 58:168-176.

(13) Stahl JE: Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008, 26:131-148.

(14) Sculpher M, Fenwick E, Claxton K: Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics* 2000, 17:461-477.

(15) Karnon J, Brown J: Selecting a decision model for economic evaluation: a case study and review. *Health Care Manag Sci* 1998, 1:133-140.

(16) Halpern MT, Luce BR, Brown RE, Geneste B: Health and economic outcomes modeling practices: a suggested framework. *Value Health* 1998, 1:131-147.

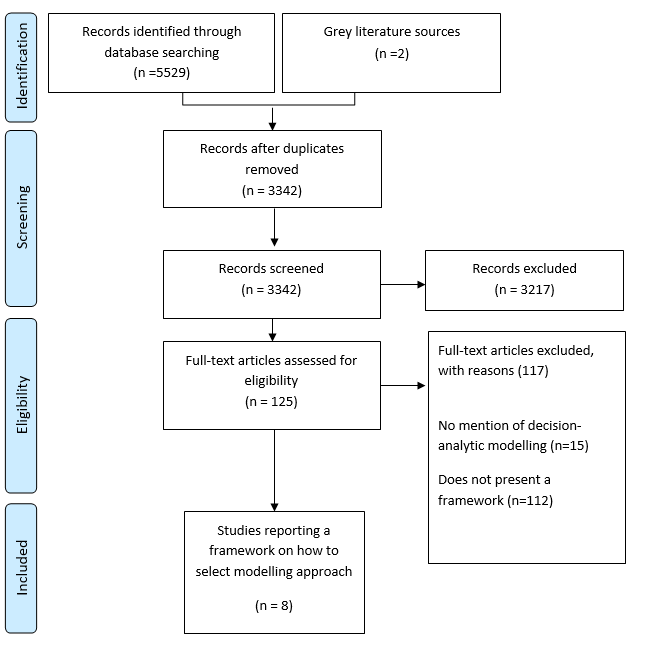
(17) Palmer AJ, Clarke P, Gray A, Leal J, Lloyd A, Grant D, Palmer J, Foos V, Lamotte M, Hermann W, Barhak J, Willis M, Coleman R, Zhang P, McEwan P, Betz BJ, Gerdtham U, Huang E, Briggs A, Carlsson KS, Valentine W: Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. *Value Health* 2013, 16:670-685

(18) Tsoi B, Jegathisawaran J, Tarride J-E, Blackhouse G, O’Reilly D. Do different decision-analytic modelling approaches produce different cost-effectiveness results? Insights from a systematic review of existing cross-validation studies. *Expert Rev Pharmacoecon Outcomes Res* 2015, 15: 451-461.

(19) Groot KB, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG: Uncertainty and patient heterogeneity in medical decision models. *Med Decis Making* 2010, 30:194-205.

(20) Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB: Model transparency and validation: a report of the ISPOR-SMDM modelling good practices task force. *Med Decis Making* 2012, 32:733-743.

**Figure 1:** PRISMA Diagram of Literature Search for Articles on Decision Frameworks to Select the Appropriate Modelling Approach



**Table 1:** Description of modelling approaches employed in health economic evaluation

|  |  |
| --- | --- |
| **Model Approach** | **Description (key terminology highlighted)** |
| Decision tree | Decision trees embody the central paradigm of decision analysis. Events in the tree are typically arranged in temporal order from left to right. Decisions are broken down into three components:   1. **Decision node**: decision point between competing strategies 2. **Chance node**: consequence to a given decision. Typically indicates point where two or more alternative chance events for a patient are possible. May contain sequential chance events. 3. **Terminal node**: Terminal branch, representing the value of a particular strategy.   **Branches** connect the nodes and represent the pathways through the tree. At each chance node, the probabilities of each consequence will determine the proportion of patients progressing down each unique path.  Consequences such as costs and effects of events and decisions may be attributed at each chance node of the tree or accumulated at the terminal nodes. The expected effect and/or costs associated with each treatment option or branch is estimated by **‘rolling’ back** the tree whereby a weighted average of the value of all branches emanating from a decision node is calculated. |
| Markov cohort model | Markov cohort models describe the transition of patients as they move through health states over time. **Health states** are mutually exclusive events, representing the entirety of the disease process and patients are assumed to be in one of a finite number of health states (known as the **unitary state requirement**). Patients within the same health state are assumed homogeneous.  Movement between health states are governed by **transition probabilities** that occur only once per **Markov cycle** (i.e. a defined time period). The transition probabilities depend only on the starting state and not on any of the previous health states (i.e., memoryless assumption).  The model is run over many cycles to build a profile of how many patients are in each state of the model over time. Estimates of costs and health outcomes are attached to the states within the model. **Cycle sum** are calculated as the weighted average of the proportion of a cohort in a health state multiplied by the value for that particular health state, summing across all health states. Expected costs and QALYs are then calculated by summing all cycle sums over the model’s time horizon. |
| Markov microsimulation | Markov microsimulation simulates individual patients over time. As individuals are modelled separately, microsimulation can store information to what has happened to the individual (i.e., memory). Similarly, as individuals are modelled, there is no need to assume homogeneity between patients. The unitary state requirement remains as patients can only be in one of a finite number of health states during each cycle. Transitions govern patient prognosis and are calculated by model parameters that reflect actual event/transition rates and may be conditional on previous and current risk factors and historical outcomes. Transitions occur only once per cycle.  Consequences such as costs and effects of events are attributed to health states and are summed over each cycle. Each patient has their own respective costs and outcome following a run through the model and the expected costs and QALYs can be calculated as the average from a large number of patients that have gone through the model. |
| Discrete event simulation | Discrete event simulation describes the flow of **entities** through the treatment system. Entities are objects, such as individuals, that may interact indirectly with other entities within the system when waiting for resources to become available.. Entities may be given **attributes**, such as characteristics or memory, which may influence their route through the simulation and/or the length of time between events. Another important concept is r**esources**, representing an object that provides service to a dynamic entity.  Life (and disease) histories of individuals are simulated one-by-one or simultaneously. If simulated simultaneously, one can model entity interactions or resourcecompetition, thereby, explicitly embedding the effects of **queues**.  Consequences such as costs and effects can be attached to events, resource use or time with a particular condition. |
| Agent-based model | This approach focuses on the **agent**. Agents are aware of their state and follow decision rules on how to communicate and interact with other agents or their environment. Agents are flexible as they may adapt over time, learn from experience and/or exist within a hierarchical structure. From simple rules governing individual actions and communication, complex behaviour may emerge.  As agents exist within a **network**, social network analysis may be used to examine interventions that impact inter-agent relationships and communication. It further provides a means for spatial considerations and can examine interventions that have a geographic impact.  Consequences such as costs and effects can be attributed to the events or patient attributes. |
| System dynamics model | The **causal loop diagram** provides a qualitative visualization of a system’s structure. Its basic building block is the **feedback loop**, describing change at one point within a system that triggers a cascading series of changes that ripple through and eventually returns in some form to either reinforce or push back against that original change. Complex behaviour may emerge from the interaction of multiple feedback loops.  The system dynamics model is quantified by stock and flow diagrams. As per its name, these diagrams consist of two main variable types: **stocks** (also referred to as levels or state or accumulations) and **flows** (i.e. rates at which stocks are either drained or replenished). Movement between stocks is defined by the rate of flow and, together, a system’s behaviour may be described through a set of differential equations.  Costs and outcomes may be attributed to the time-in-stocks or movements between stocks that are continuously updated. |
| Compartmental model | Compartmental models are historically used to model the epidemiology of infectious disease. The population is divided into various **compartments**, representing their average state. Individuals within a single compartment are considered homogeneous. Most commonly, it contains compartments of the population whom are at different stages of the illness (e.g. susceptible – exposed - infectious- recovered). |

**Table 2:** Overview of the decision framework and the modelling approaches covered within the respective frameworks

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author** | **Country** | **Generic or Disease-specific framework** | **Modelling Approaches mentioned within the Frameworks** | | | | | | | **Notes (i.e., relating to semantics and terminology)** |
| **Decision tree** | **Markov cohort model** | **Markov microsimulation** | **Discrete event simulation** | **Agent-based Models** | **System Dynamics** | **Compartmental Models** |  |
| Jit, 20116 | United Kingdom; Canada | Infectious-disease | X | X |  |  |  |  | X | Static model aggregates decision tree and Markov cohort models; dynamic model refers specifically to compartmental models. |
| Kim, 20088 | USA | Infectious disease | X | X | X | X | X | X | X | System dynamics models further separated into *discrete difference equation, ordinary differential equations* and *partial differential equations models*. |
| Heeg, 20087 | Netherlands | Generic | X | X | X | X |  |  |  | Markov microsimulation referred to as *first-order Markov model*. |
| Stahl, 200813 | USA | Generic | X | X |  | X | X | X |  | Markov model, in this framework, does not distinguish between Markov cohort models and Markov microsimulations. |
| Chick, 200711 | France | Generic | X | X | X | X |  | X |  | System dynamics model separated into *finite difference model* and *ordinary differential equation* (ODE). Markov microsimulation referred to as *patient-level simulation* DES includes both discrete event simulation and patients evolve on discrete-time grid. Decision tree described in many ways including: decision tree, stochastic decision tree, stochastic decision tree with covariates. |
| Cooper, 200712 | United Kingdom | Generic | X | X |  | X |  |  |  |  |
| Brennan, 200610 | United Kingdom | Generic | X | X | X | X |  | X |  | Markov model also referred to as simulated markov model. Markov microsimulation separated into *simulated patient-level Markov model* and *individual event history model*. System dynamics model also referred to as *Markov chain models*. |
| Barton, 20049 | United Kingdom | Generic | X | X | X | X |  | X |  | Markov microsimulation referred to as *individual sampling model*. |

**Table 3:** Summary of the decision criteria (i.e. structural features and practical considerations) considered within each decision framework

1. **Generic Decision Frameworks**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author** | **Type of Framework** | **Framework Elements** | | | | | | | | | | |
| **Structural features** | | | | | | **Practical considerations** | | | | |
| **Population resolution** | **First order uncertainty** | **Interactivity** | **Resource constraints** | **Dimension of Time** | **Other:** | **Time** | **End-user requirement** | **Simplicity** | **Validity** | **Other** |
| Heeg, 20087 | Radar graph | X | X |  | X | X | • Memory | X |  |  | X | • Experience |
| Stahl, 200813 | Flow diagram | X |  | X | X | X | • Agent autonomy  • Spatial consideration |  | X |  |  |  |
| Chick, 200711 | Table | X | X |  |  | X | • Expected value |  |  |  |  |  |
| Cooper, 200712 | Flow diagram |  |  | X† | X† |  | •Modelled duration  • Recurrence  •Aggregation of cohort†† | X | X | X | X | • Model error |
| Brennan, 200610 | Table | X | X | X† | X† | X | • Expected value |  |  |  |  |  |
| Barton, 20049 | Flow diagram | X |  | X† | X† |  |  |  |  | X | X |  |

† Interaction, as defined in this framework, includes both interaction between individuals or constraints in resources that affect individuals.

††Aggregation of cohort refers to whether a single or multiple cohort of patients are modelled. It is commonly referred to whether the population is open (i.e., new individuals can enter model) or closed (i.e., no new additions are made in the model).

1. **Decision Frameworks Specific to Infectious Disease Modelling**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Author** | **Type of Framework** | **Framework Elements** | | | | | | | | | | |
| **Structural features** | | | | | | **Practical considerations** | | | | |
| **Population resolution** | **First order uncertainty** | **Interactivity** | **Resource constraints** | **Dimension of Time** | **Other:** | **Time** | **End-user requirement** | **Simplicity** | **Validity** | **Other** |
| Jit, 20116 | Flow diagram |  |  | X |  |  |  |  |  |  |  |  |
| Kim, 20088 | Table | X | X | X |  |  |  |  |  |  |  |  |

**Table 4:** Nomenclature and definition of commonly-mentioned decision criterion for selecting a modelling approach

1. Structural features

|  |  |  |  |
| --- | --- | --- | --- |
| **Structural Element** | **Question to differentiate between attributes:** | **Typical Classification** | **Definition** |
| Population resolution | What level is the model arising? | Aggregate (may also be referred to as cohort) | The model is at a macro-level with a population aggregated and run through the model together. Variables represent population averages8. Relies on a homogeneity assumption that individuals within a particular health state are homogeneous10,12. To incorporate individual factors or memories into the model, separate health states are required10,12. Interactions are also modelled at an aggregate level. |
| Individual | The model is at a micro-level with individuals going through the model separately8,10,13.This easily incorporates individual factors and memory. Patient characteristics may be retained as continuous variables4.  Permits exploration of first-order uncertainty. |
| First order uncertainty19 | To what extent is the model capable of incorporating and analysing patient-level variability within its structure? | Deterministic | No variability in the outcomes between identical patients. Within a given sample of patients, individuals facing the same probabilities and outcomes will experience the effects of a disease or intervention identically. |
| Stochastic | Permits random variability in outcomes between identical patients as there exists uncertainty in patient-level outcomes that is entirely due to chance. Within a given sample of patients, individuals facing the same probabilities and outcomes will experience the effects of a disease or intervention differently.  This can be perceived as a form of random error and, with increased sample size, the extent of this uncertainty can be reduced. |
| Interactivity | Are actors in a model or the overall system independent? | Static/ Independent | No interaction present between or within actors as each actor is independent. No interactions at the system level9. |
| Dynamic/ Dependent | Interaction exists between or within actors or at the level of the system. Feedback and interdependencies may exist within the modelled system9. |
| Resource constraint | Are constrained resources or queuing important to the decision problem? | Unlimited | There exist no constraints in the system. |
| Constrained | Resource constraints has impacts on features within the model13. |
| Dimension of time | How is time handled by the model? | Untimed | Time is not explicitly modelled. Another term used to describe this concept of time is “aggregate” as changes in time are not considered important to the model13. |
| Discrete | Time separated into discrete units with an event occurring during one of the discrete time steps8,13. To handle simultaneous events, requires smaller fixed time intervals10. |
| Continuous | Time is continuous with an event occurring at any point in the continuum of time; thereby, permits modelling of multiple simultaneous events8. |

1. Practical considerations

|  |  |
| --- | --- |
| **Practical Consideration** | **Definition** |
| Data availability | The availability of the necessary data to populate the economic model7. |
| End-user Requirement | This considers whether the model meets the need of its end-users and decision-makers. It is dependent on how well the model structure reflects and is able to capture all relevant aspects of the underlying reality and the corresponding uncertainties that exist6,10. End user requirement may capture whether the modelling approach is considered acceptable and whether funding is present to support a particular project. |
| Experience | The extent to which the modeller has accumulated knowledge and implementation skills to construct the model7. |
| Model error | The degree of imprecision in the model that is deemed acceptable by either the modeller and/or its end-users12. Model error can either be systematic or unsystematic. Unsystematic error, synonymous to uncertainty, can be explored through the application of sensitivity analysis. The feasibility of conducting sensitivity analysis is dependent on the model structure and its underlying parameters. |
| Modelling software availability | The accessibility of the necessary software(s) to construct and evaluate the model. Different software may support different modelling approaches and are associated with licensing fees. Softwares for health economic modelling include Microsoft Excel (for decision trees and Markov cohort models); Treeage (for decision tree, Markov cohort model and Markov microsimulation); Arena (for discrete-event simulation); AnyLogic (for discrete-event simulation, agent-based model, system-dynamics and compartmental models); and Berkeley Madonna (for system-dynamics and compartmental models). |
| Simplicity | The degree of complexity in a model. This is essentially dependent on the size of the model (e.g. the number of states/transitions in state-transition models) and the number of parameters present9. Simpler models are more likely to be understood and accepted by stakeholders12. |
| Time | This considers the speed of model development and captures several aspects including the time required to programme the model (building time), the time required to collect the necessary data to fill the model (data collection) and the time required to generate simulation results (simulation time)7. |
| Transparency | The degree to which the end-user of the model can review the model structure, equations, parameter values and the underlying assumptions. This is considered important by modellers for two reasons: 1) to provide non-quantitative description of the model to those interested in understanding how a model works; and ii) to provide technical information to those interested in evaluating a model at the higher level mathematical and programming detail, possibly with the interest to replicate the results. Transparency promotes an understanding on the model’s accuracy, limitation and potential application. This is deemed important to build trust and confidence in a model by the appropriate decision-makers20. |
| Validity | The clinical representativeness of a model to the actual decision problem7,12. This addresses how adequately a chosen modelling approach reflects and captures all relevant aspects of the underlying reality and the corresponding uncertainties that exist. |

**Table 5:** Classification of structural elements, specific to each modelling approach, according to the decision frameworks

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Modelling Approach** | **Number of Frameworks that include this modelling approach**  n [%‡] | **Structural Assumptions**  **n [%†]reference** | | | | | | | | | | | | | | | |
| **Population resolution** | | | **First-order uncertainty** | | | **Interactivity** | | | **Resource constraints** | | | **Dimension of Time** | | | |
| Not specified within framework | Classification | | Not specified within framework | Classification | | Not specified within framework | Classification | | Not specified within framework | Classification | | Not specified within framework | Classification | | |
| Cohort | Individual | Deterministic | Stochastic | Static/ Independent | Dynamic/ Dependent | Unlimited resources | Constrained resources | Untimed | Discrete | Continuous |
| **Decision tree** | 8 [100] | 3  [37.5]  6,9,12 | 5 [62.5]  7,8,10,11,13 | 3 [37.5]  10,11,13 | 4  [50]  6,9,12,13 | 4  [50]  7,8,10,11 | 2  [25]  10,11 | 2  [25]  7,11 | 6  [75]  6,8-10,12,13 | 0 | 6  [75]  6,8-11,13 | 2  [25]  7,12 | 0 | 4  [50]  6,8,9,12 | 4  [50]  7,10,11,13 | 0 | 0 |
| **Markov cohort model** | 8 [100] | 3  [37.5]  6,9,12 | 5  [62.5]  7,8,10,11,13 | 0 | 4  [50]  6,9,12,13 | 3  [37.5]  7,8,10 | 2  [25]  10,11 | 2  [25]  7,11 | 6  [75]  6,8-10,12,13 | 0 | 6  [75]  6,8-11,13 | 2  [25]  7,12 | 0 | 4  [50]  6,8,9,12 | 0 | 4 [50]7,10,11,13 | 1  [12.5]  13 |
| **Microsimulation** | 6 [75] | 1  [16.6]  9 | 0 | 5 [83.3]  7,8,10,11,13 | 3  [50]  9,10,13 | 0 | 3  [50]  7,8,11 | 2  [33.3]  7,11 | 4  [66.7]  8-10,13 | 2  [33.3]  8,10 | 5  [83.3]  8-11,13 | 1 [16.7]  7 | 0 | 2  [33.3]  8,9 | 0 | 3 [50]  7,10,13 | 3  [50]  10,11,13 |
| **Discrete event simulation** | 7 [87.5] | 1  [14.3]  12 | 1 [14.3] 13 | 6 [85.7]  7-11,13 | 4  [57.1]  9,10,12,13 | 0 | 3  [42.9] 7,8,11 | 2  [28.6]  7,11 | 1  [14.3] 13 | 5  [71.4]  8-10,12,13 | 4  [57.1]  8,9,11,12 | 0 | 4  [57.1]  7,10,12,13 | 3  [42.9]  8,9,12 | 0 | 2 [28.6]  11,13 | 4  [57.1]  7,10,11,13 |
| **Agent-based models** | 2 [25] | 0 | 0 | 2 [100] 8,13 | 1  [50]  13 | 0 | 1  [50]  8 | 0 | 0 | 2  [100]  8,13 | 1  [50]  8 | 0 | 1  [50]  13 | 1  [50]  8 | 0 | 1  [50]  13 | 1  [50]  13 |
| **System Dynamics** | 5 [62.5] | 0 | 5 [100] 8-11,13 | 0 | 2  [40]  9,13 | 3  [60]  8,10,11 | 2  [40]  10,11 | 1  [20]  11 | 0 | 4  [80]  8-10,13 | 5  [100]  8-11,13 | 0 | 0 | 2  [40]  8,9 | 0 | 3  [60]  10,11,13 | 3  [60]  10,11,13 |
| **Compartmental models** | 2 [25] | 1  [50]  6 | 1  [50]  8 | 0 | 1  [50]  6 | 1  [50]  8 | 1  [50]  8 | 0 | 0 | 2  [100]  6,8 | 2  [100]  6,8 | 0 | 0 | 2  [100]  6,8 | 0 | 0 | 0 |

†: Denominator out of the number of frameworks that have discussed that specific modelling approach (i.e. second column).

‡ : Denominator out of8 (total number of decision frameworks identified)

**APPENDIX 2.I:** Search Strategy

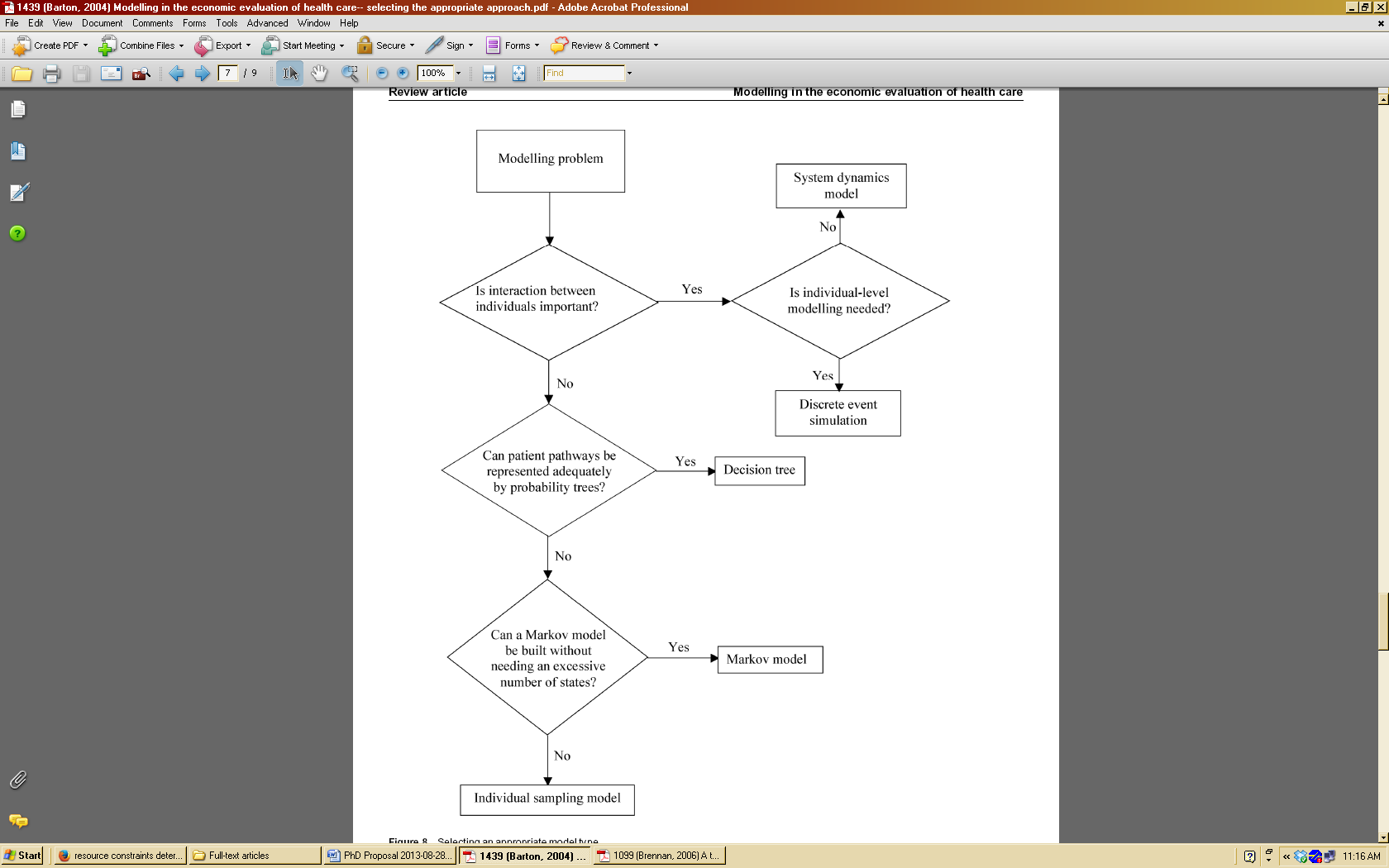
|  |  |
| --- | --- |
| Database: Ovid MEDLINE(R) <In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present> | |
|  | [(Economic evaluation\*[Title/Abstract] OR economic outcome\*[Title/Abstract] OR economic analys?s[Title/Abstract] OR health economic\*[Title/Abstract] OR pharmacoeconomic\*[Title/Abstract] or pharmaco-economic\*[Title/Abstract])  (37,258) |
|  | (decision\* analy\* [Title/Abstract] OR model?ng adj (approach\* OR method\* OR practice\* OR technique\*)[Title/Abstract]  (33,906) |
|  | (Guid\*[Title/Abstract] OR select\*[Title/Abstract] OR choos\*[Title/Abstract] OR choic\*[Title/Abstract] OR categoris\*[Title/Abstract] OR framework\*[Title/Abstract] OR taxonom\*[Title/Abstract] OR recommend\*[Title/Abstract]  (4,608,561) |
|  | 1 AND 2 AND 3  (802) |

**APPENDIX 2.II:** Level I Selection Criteria

|  |
| --- |
| **Modeling Framework ~ Selection Criteria (LEVEL I SCREENING)**   1. Is this study focused on economic evaluation methods?   No (exclude) Yes (include) Maybe (include)   1. Does this study discuss either: 2. frameworks that assist in selecting the approach to modeling?   No (exclude) Yes (include) Maybe (include)   1. Is this study published in English?   No (exclude) Yes (include) Maybe (include)  NOTE: Markov model, state-transition model, decision tree, discrete event simulation, agent-based modeling, microsimulation, system dynamics modeling |

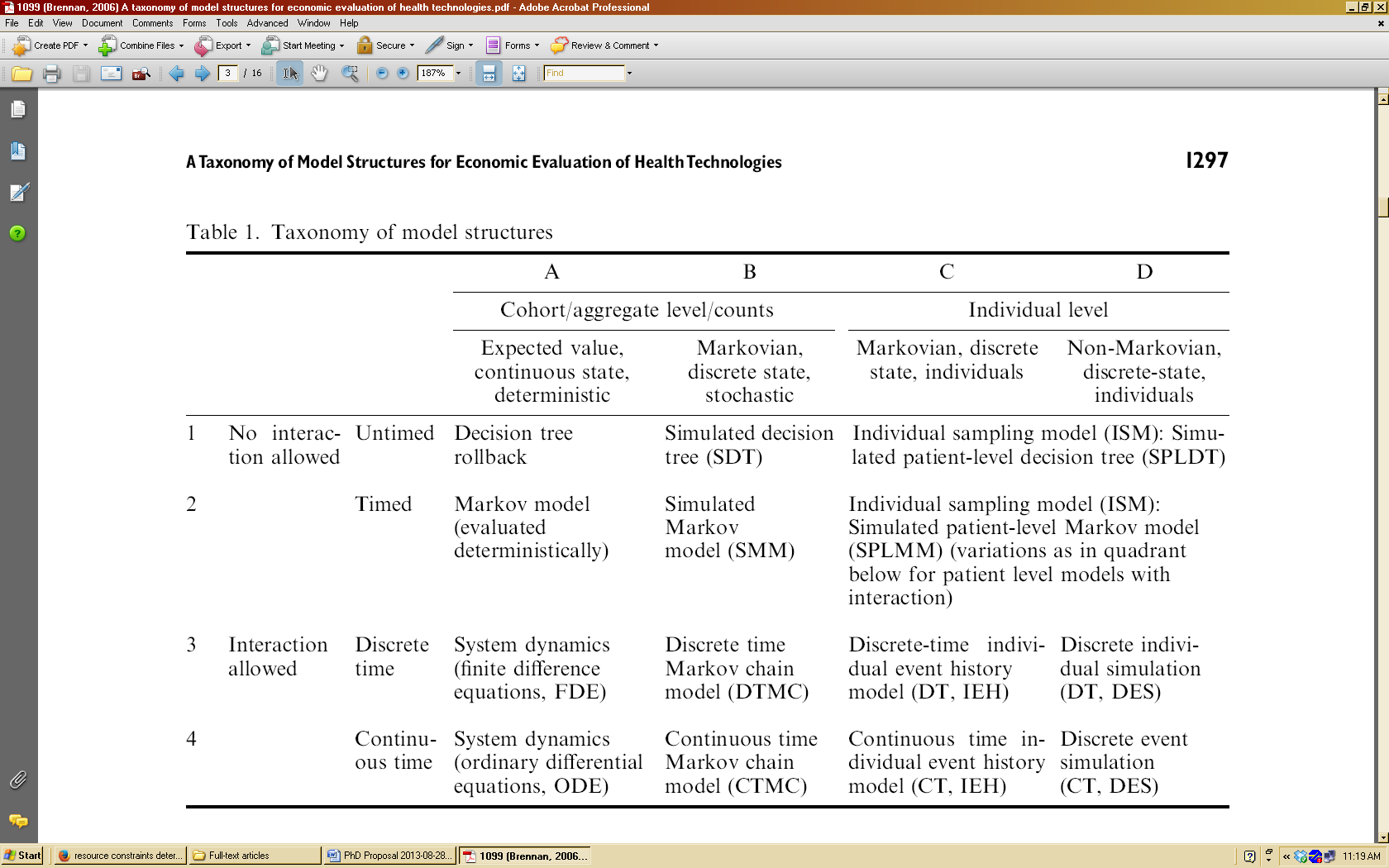
**Appendix 2.III:** Copy of the Existing Decision Frameworks

**Figure 2:** Barton et al’s decision flowchart(9). Based on a series of questions, this guides the decision on which modelling approach would be most suitable for the modelling problem at hand.



*(copyright ©2004 Barton et al, “Selecting an appropriate model type.” With kind permission from SAGE Publications Ltd)*

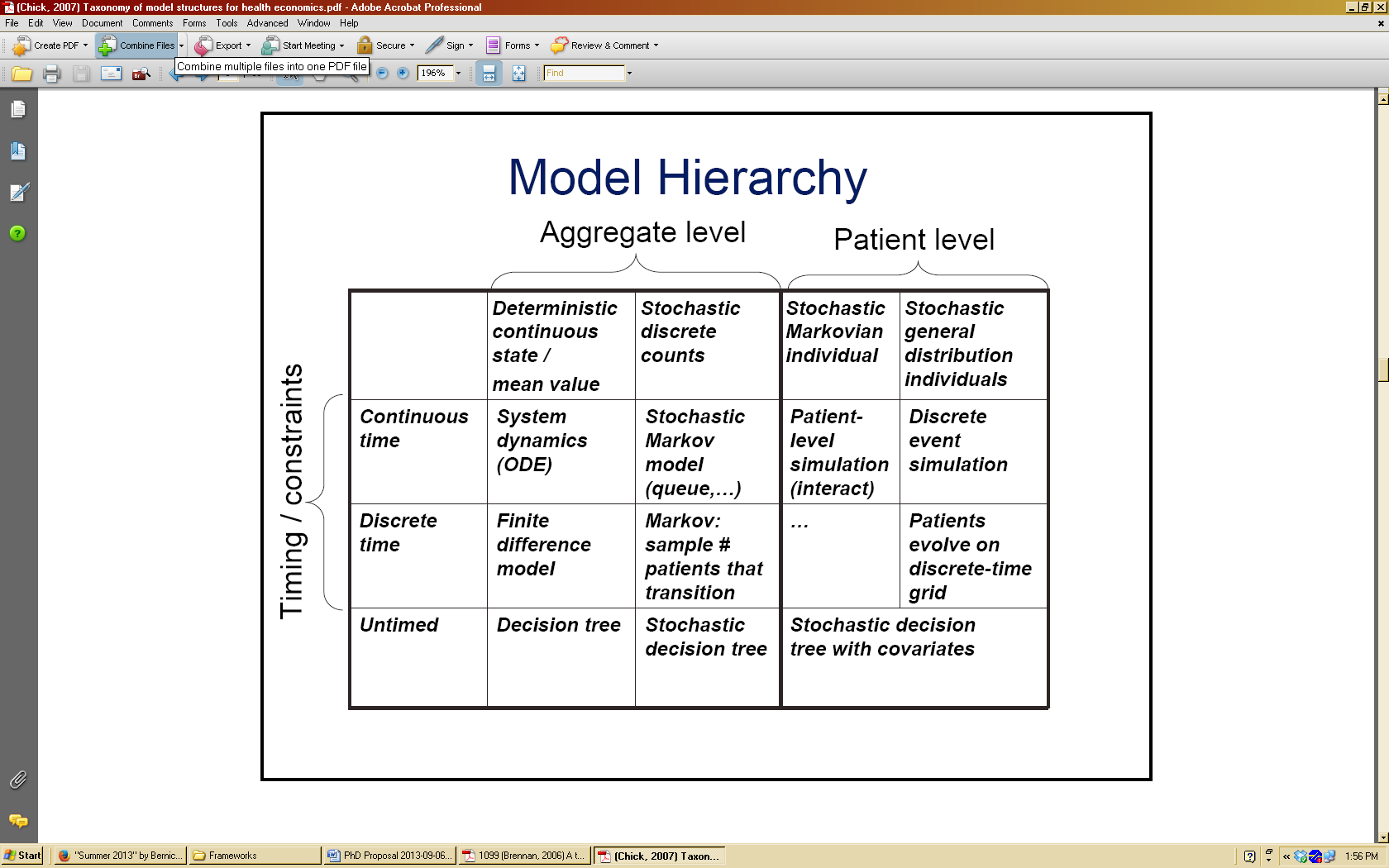
**Figure 3:** Brennan et al’s classification framework(10). This taxonomy emphasizes the distinguishing structural assumption within each modelling approach and highlights the commonalities and differences between modelling approaches used in health economic evaluation.



*(copyright ©2006 Brennan et al, ‘Taxonomy of model structures’ with kind permission from Springer Science and Business Media)*

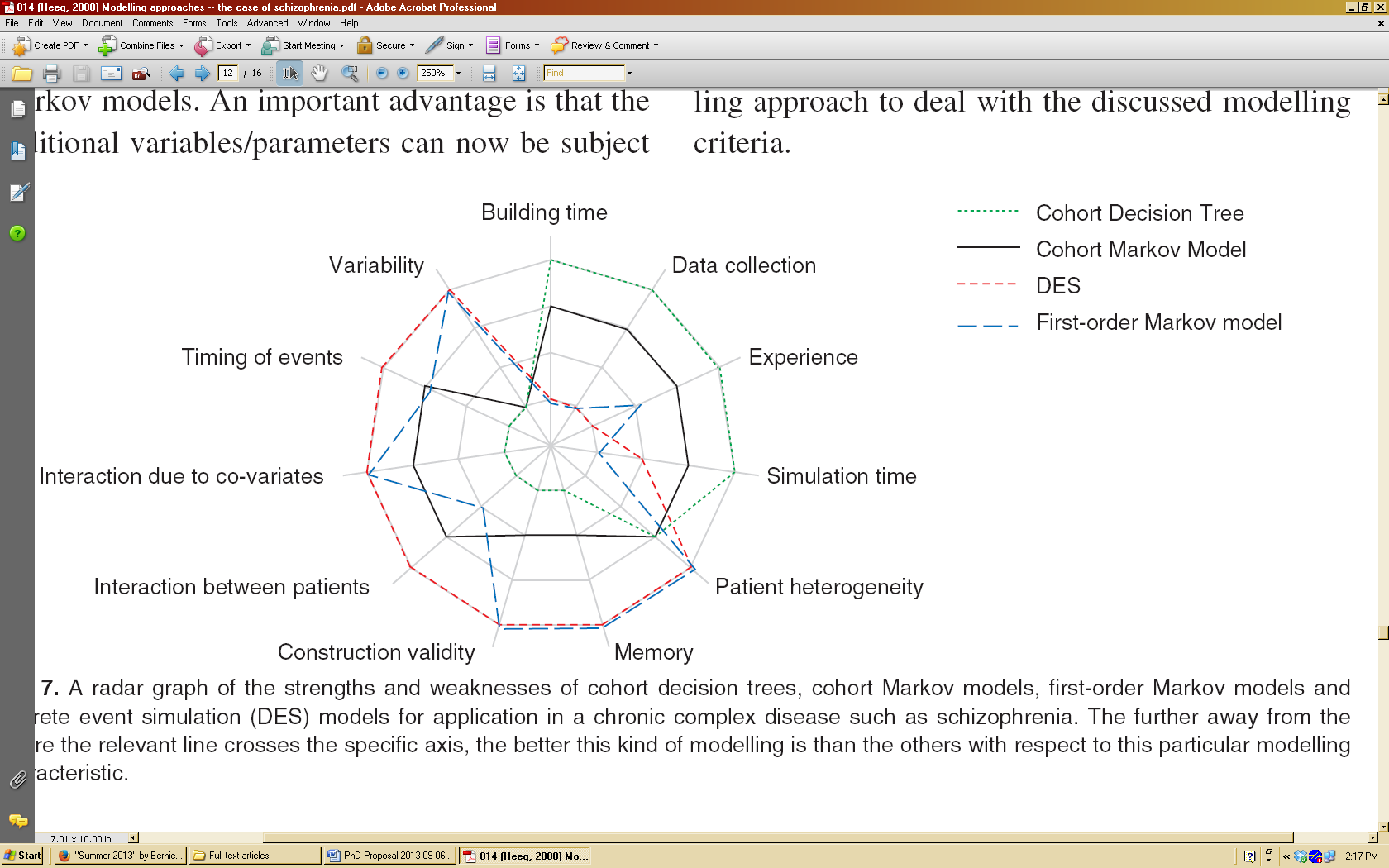
*Licence number:* 3612220480339

**Figure 4:** Chick’s framework(11), adapted from Brennan’s taxonomy



(copyright ©2007 Chick)

**Figure 5:** Heeg’s radar diagram(7), adapted from Brennan’s taxonomy. The further away from the origin the model’s line crosses the axis of the selection criteria, the better the model is able to incorporate or handle that corresponding structural feature or practical characteristic.DES: discrete event simulation

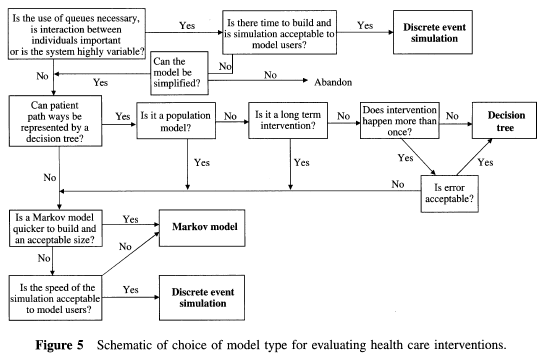


DES: discrete event simulation

*(copyright ©2008 Heeg et al, “A radar graph of the strengths and weaknesses of cohort decision trees, cohort Markov models, first-order Markov models and discrete event simulation (DES) models for application in a chronic complex disease such as schizophrenia”, With kind permission from Springer Science and Business Media)*

*Licence number:* 3612220245969

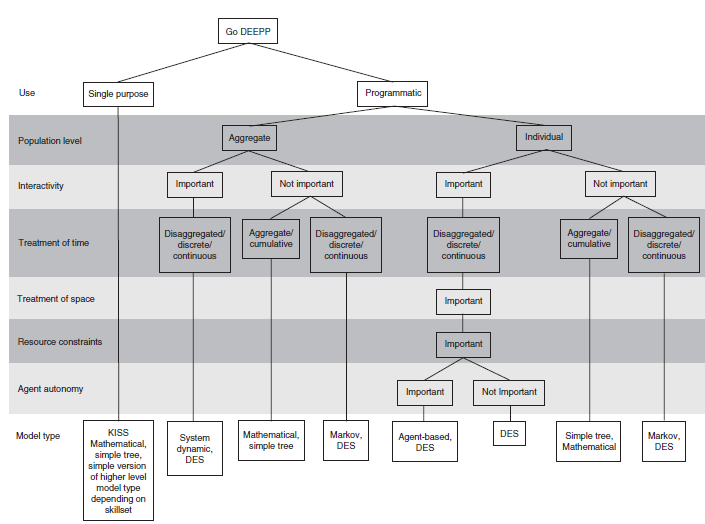
**Figure 6:** Cooper’s flowchart weaves together practical aspects and structural elements, suggestive that the decision on the choice of modelling approach needs to consider both(12).

****

Reprinted by permission from Macmillan Publishers Ltd: Journal of the Operational Research Society, advance online publication, 1 Feb 2007 (doi: 10.1057/sj.*J Oper Res Soc*.2602230) published by Palgrave Macmillan

Licence number: 3612211273444

**Figure 7:** Stahl’s flowchart to assist in selecting the modelling approach(13).

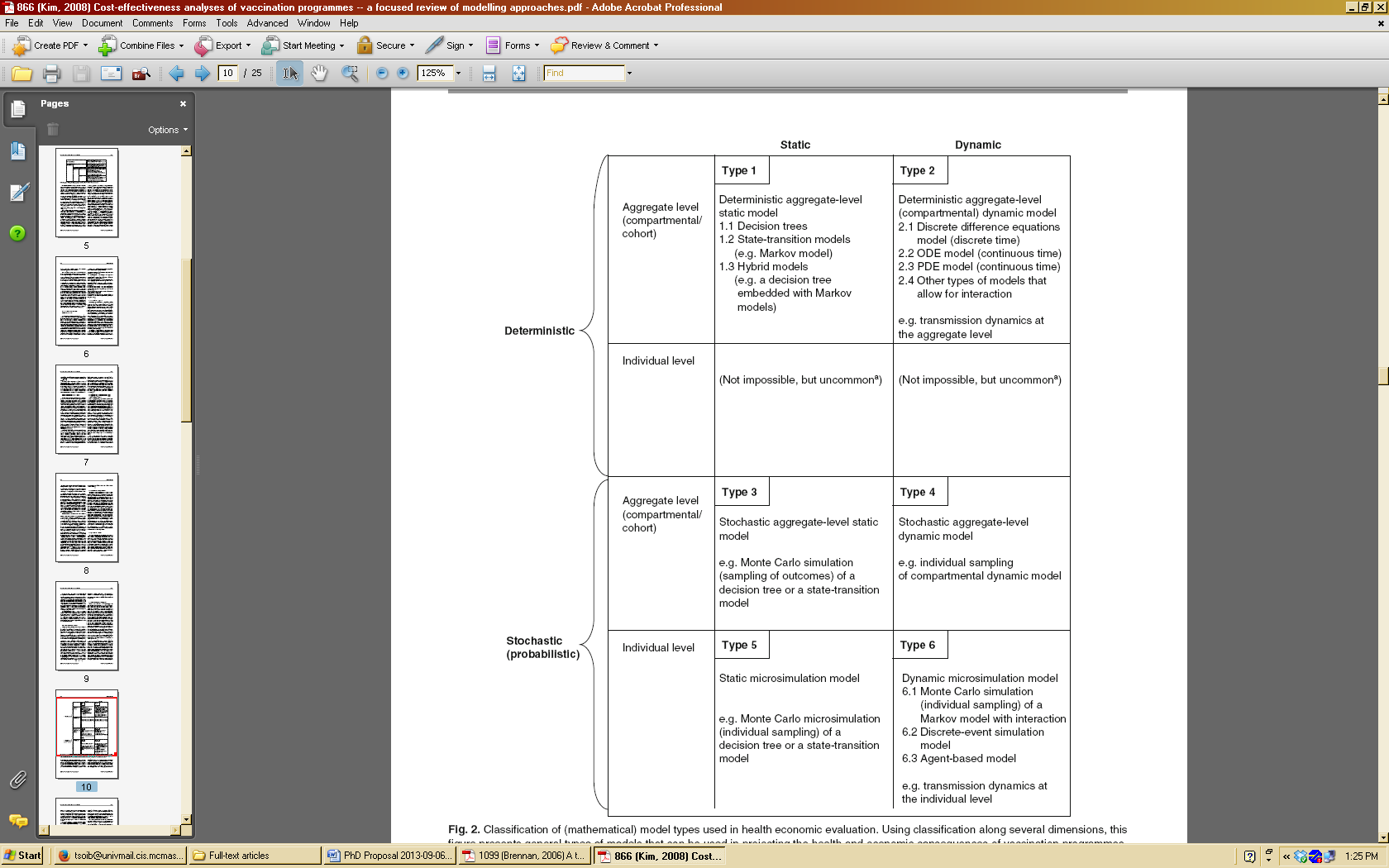
****

DEEPP: describe, evaluate, explore, predict and persuade; KISS: keep it simple stupid; DES: discrete event simulation

*(copyright ©2008 Stahl, “A decision algorithm for choosing a simulation method”. With kind permission from Springer Science and Business Media)*

Licence number: 3612210945819

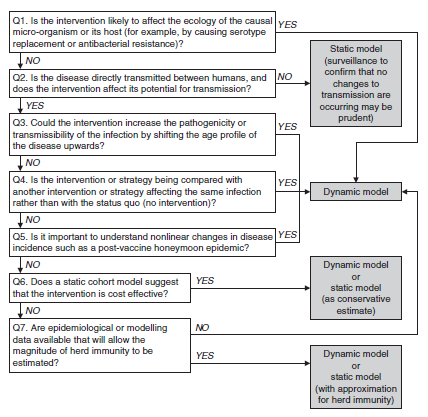
**Figure 8:** Kim et al’s(8) framework specific to infectious-disease modelling, based on an adaptation of Brennan’s framework(10).



*(copyright ©2008 Kim et al, “Classification of (mathematical) model types used in health economic evaluation.” With kind permission from Springer Science and Business Media)*

*License number:* 3612210728201

**Figure 9:** Jit et al’s(6) framework primarily distinguished between the instances in which a dynamic or static model would be appropriate in infectious disease modelling. The key decision criterion relates to interaction and whether the force of infection is constant or variable.



*(copyright ©2011 Jit et al, “Flow diagram showing how the choice of a static model, static model with approximation for herd immunity or dynamic model could be made based on answers to seven key questions” With kind permission from Springer Science and Business Media)*

*License number:* 3612210275431

# CHAPTER 3:

**Do different decision-analytic modelling approaches produce different results? A systematic review of cross-validation studies**

Tsoi B BSc, MSc, PhD(cand)1,2ŧ Goeree R BA, MA1-3,, Jegathisawaran J BMSc, MHEcon,1,2 Tarride J-E BA MA, PhD,1 Blackhouse G BA, MBA, MSc1,2 O’Reilly D BSc, MSc, PhD1-3,

1. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

2. PATH Research Institute, St. Joseph’s Healthcare Hamilton, Ontario, Canada

3. Centre for Evaluation of Medicines (CEM), St. Joseph’s Healthcare Hamilton, Ontario,

Canada

Cited as: Tsoi et al. Do different modeling approaches produce different results? A systematic review of cross-validation studies. 2015. *Expert Rev Pharmacoecon Outcomes Res.* 15: 451-461

Doi: 10.1586/14737167.2015.1021336

© 2015 Tsoi et al.; Licensee Informa Healthcare

**Summary**

When choosing a modelling approach for health economic evaluation, certain criteria are often considered (e.g. population resolution, interactivity, time advancement mechanism, resource constraints). However, whether these criteria and their associated modelling approach impacts results remain poorly understood. A systematic review was conducted to identify cross-validation studies (i.e. modelling a problem using different approaches with the same body of evidence) to offer insight on this topic. With respect to population resolution, reviewed studies suggested that both aggregate- and individual-level models will generate comparable results, although a practical trade-off exists between validity and feasibility. In terms of interactivity, infectious-disease models consistently showed that, depending on the assumptions regarding probability of disease exposure, dynamic and static models may produce dissimilar results with opposing policy recommendations. Empirical evidence on the remaining criteria is limited. Greater discussion will therefore be necessary to promote a deeper understanding of the benefits and limits to each modelling approach.

**3.1 Introduction**

Economic evaluations of health technologies are often conducted by employing decision-analytic modelling techniques. In particular, the choice of a modelling approach is crucial since this can impact the validity of the underlying model and its results1,2. Furthermore, since economic evaluations are conducted to inform decision-makers on the optimal allocation of scarce resources, the modelling approach chosen may have a significant impact on the subsequent policy decisions.

Several approaches to modelling are available: from ‘traditional’ methods (i.e., decision tree and Markov model) to more recent developments of ‘alternative’ modelling techniques (e.g., Markov microsimulation, discrete-event simulations (DES), agent-based models, system dynamics and compartmental models). Table 1 provides a description of each of these modelling approaches. Researchers often encounter common questions during the developmental stages when selecting a modelling approach. For instance, when does patient heterogeneity truly matter and to what extent might the results be impacted if patient heterogeneity is incorporated? Similarly, to what degree do queuing and constrained resources impact the cost-effectiveness of an intervention such that it warrants the need to be explicitly modelled? Should time be managed as realistically as possible by treating it as a continuous measure or would simplification of time as a discrete input be suitable and not introduce significant bias to a model?

Given the numerous options available in a modeller’s toolkit and the above frequently raised questions, frameworks have emerged to help guide the choice on the most suitable modelling approach3,4. A recent systematic review on these frameworks found that each featured a different set of selection criteria5 and the most common criteria within these frameworks are outlined in Table 2. Overall, the choice of a modelling approach is multi-factorial, as the decision does not simply consider the structural features desired within a model but is also dependent on practical considerations.

However, the decision criteria proposed within these frameworks are mostly based on heuristics5. Few have directly linked their recommended decision criteria back to empirical evidence that otherwise would provide deeper insight on the extent to which the criterion impacts the construction and analysis of the decision-analytic model. One way to gather such insight is through cross-validation exercises whereby, for a single research question, different modelling approaches are built based on the same data sources6. Although the conventional rationale for cross-validity is to garner greater confidence in a particular model when different modelling approaches produce similar results7, the opposite may also be useful towards understanding the underlying differences between these approaches. Indeed, it is well understood that a particular modelling approach may introduce certain constraints to a model’s development and conceptualization in terms of what elements can be captured and the ease with which these can be incorporated. Indeed, as Meadows and Robinson8 explained, “every modeling discipline [modelling approach] depends on unique underlying assumptions; that is, each modeling method is itself based on a model of how modeling should be done”. As a result, the divergence and convergence in the performance and the results of different modelling approaches can provide insight into the technical assumptions inherent to each approach along with a better understanding of their relative strengths and weaknesses.

To our knowledge, although cross-validation studies have been conducted in a handful of health economics models, limited attempts have been made to assess the consistency in the results and the conclusions drawn from these exercises9. This literature review is thus interested in appraising existing cross-validation studies to examine the extent to which different decision criteria, which distinguishes the modelling approaches being compared, impacted the quantitative results and qualitative conclusions. The focus will be on the structural features, although practical considerations will also be noted. We identify the conditions under which discrepant results emerged between modelling approaches or, equally as important, when similar results were produced, in order to understand the relative importance of each decision criterion. From this work, analysts will be able to better understand the relative strengths and limitations between decision-analytic modelling approaches and how each decision criterion can influence the choice on which modelling approach is most suitable for a given research question.

**3.2 Methods**

To identify relevant case studies, a literature search was performed on April 1, 2013 with updates up to 21 January 2014. The following bibliographic databases were searched: OVID Medline (1946-present; In-Process & Other Non-Indexed Citations) and EMBASE (1996-present); Wiley’s Cochrane Library (Issue 4 of 12, Apr 2012) and Health Economic Evaluation Database (HEED); PubMed (for non-Medline records); and Thomson Reuters’ BIOSIS Previews. The search strategy was constructed using controlled vocabulary terms, such as the National Library of Medicine’s Medical Subject Headings and additional keywords (Appendix 3.I). No restrictions were applied on the date or type of publication. Grey literature (e.g., working papers, commissioned reports, policy documents, websites) were also searched and, for comprehensiveness, reference lists of publications included after full text screening were reviewed for additional citations.

Records were screened for inclusion based on pre-defined criteria presented in Appendix 3.II. To be included, the study had to have conducted a cross-validation of modelling approaches, defined as an empirical direct comparison of two or more modelling approaches to address the same health economic research question that explicitly used the same body of evidence to populate the models. Studies were limited to those published in English. In the first stage of screening, the titles and abstracts of the records identified from the bibliographic search were screened by one reviewer, with a 50% random check conducted by a second independent reviewer. If, at least one reviewer identified a citation as being potentially relevant, the full text of the published literature was obtained. In the second stage of the screening, one reviewer screened the full text of all included papers, with a second reviewer completing a random independent verification of 50% of the records. Any discrepancies between the two reviewers at this stage were discussed until consensus was reached. For irresolvable differences, a third researcher was asked to make the final decision regarding a study’s eligibility.

Data extracted from relevant studies include: bibliographic information (e.g., author, year of publication), disease topic, methodology (e.g., modelling approaches, parameters, scenario analysis and sensitivity analysis), the selection criteria studied and the results/conclusions generated by each separate modelling approach. The selection criteria, separated into structural elements and practical considerations, were identified according to the definitions presented in a previous publication5 (Table 2). In the case of missing information, authors of the relevant publications were contacted.

Relevant studies that met the full inclusion criteria were analyzed narratively. The studies were first organized according to the structural selection criteria that was being explored and subsequently by the disease area (i.e., communicable vs. non-infectious illnesses). The study was first examined individually on the degree of agreement between modelling approaches in their base case results and, furthermore, the conditions that led to a divergence or convergence between the models. The key findings that emerged from each study were then assessed to determine whether these were study-specific (i.e., findings apply only to the study at hand) or were generalizable (i.e., similar findings were found across the other studies that also explored the same structural decision criterion).

**3.3 Results**

The search identified 3,346 publications of which nine met the inclusion criteria (Figure 1). All studies compared two modelling approaches. Five were in the realm of infectious diseases10-14; two in the area of breast cancer15,16; and the remaining two on a hypothetical disease17,18: one exploring a communicable disease and the other on a non-communicable disease (Table 3). In the studies on infectious diseases, the structural decision criteria of interactivity (dynamic vs. static models)10,12,14,18 and population resolution (aggregate/cohort- vs. individual-level models)11,13 were explored using a variety of modelling approaches. All three cases of non-communicable diseases compared DES and Markov models, with one each focused on the features of population resolution16, time advancement mechanism17 and resource constraints15.

In addition, all studies that addressed population resolution further addressed practical trade-offs, such as time11,16, model simplicity13,16 and data availability11 (Table 3). Detailed summaries of each study can be found in Appendix 3.III and further down, we present the results grouped by their respective structural criterion.

**3.3.1 Population Resolution**

Three studies explored this structural decision criterion: two in the realm of HIV11,13 and one on breast cancer16. In particular, the two HIV studies looked at different disease stages, with one focused on treatment-experienced13 and the other on treatment-naïve patients11. All three studies selected a Markov cohort model as their aggregate-level model, which was compared against, at an individual-level, a Markov microsimulation13 or a DES11,16 in one and two of the studies, respectively.

The authors noted structural differences in how the decision problem was modelled between the approaches. Individual-level models could more concisely capture heterogeneity and time-dependencies. Rather than introducing additional health states that would have been necessary in aggregate-level models, heterogeneity could be incorporated in individual-level models by tracking and storing an individual’s history as each patient progressed through the model13,16. Furthermore, individual-level models permitted a more flexible representation of the data in its existing format. For instance, Karnon et al.16 noted that disease progression based on survival data had to be transformed in their Markov cohort model to constant transition probabilities, whereas, in DES, the data in its original form could be incorporated into their model.

Despite these structural differences, in all three studies, the authors found that the expected outcomes were similar between the aggregate- and individual-level models (Appendix 3.II). Although individual-level models produced slightly higher estimates than aggregate-level models, the overall incremental cost-effectiveness ratios (ICER) were alike and the qualitative conclusions remained the same irrespective of the model’s population resolution11. The authors in all studies concluded that the implications when considering the structural criterion of population resolution rather laid in the trade-off between validity - feasibility13 (also referred to as a flexibility - analytic input trade-off16). The authors noted that aggregate-level models tended to have shorter computing time and could more easily conduct probabilistic sensitivity analysis13,16. However, this came at a sacrifice of validity. The classical Markovian assumption of “memory-less” restricted how patient pathways were modelled16 and difficulty existed in incorporating memory without state explosion (i.e., the creation of a large number of states that would lead to a large and difficult-to-manage model)16. Contrary, the face validity of individual-level models (e.g., Markov microsimulation, DES) was often found to be superior given that it predicted the course of disease more naturally and flexibly. These models were able to incorporate a fuller set of inputs without requiring aggregation and averaging11,13, avoid over-simplifying assumptions as data were incorporated in its original form11,13,16, and represent the problem more compactly without requiring additional health states13,16. However, all authors concluded that these benefits were at the expense of greater difficultly in conducting probabilistic sensitivity analysis and their extensive data requirements and time needed for verification and validation 11,13,16.

Given that only a subset of trial patients was used to parameterize the Simpson et al. model11, the authors were able to assess predictive validity using the remaining patients. It was found that, over a shorter time horizon, the clinical estimates of both models aligned with real-world observations, although this diverged over time with DES providing more accurate predictions11. Although no explanation was provided to explain this trend, the authors noted that the results still fell within a margin of error less than 3% even at a longer five-year time horizon. A further advantage noted specifically in DES is their modelling flexibility as they were able to predict subtle clinical details (in this case: predictions on CD4+ T-cell count, viral loads at various time points) that Markov cohort models could not have estimated otherwise11.

In a study on breast cancer screening, Karnon et al.16 noted substantial differences between the two modelling approaches for one of the clinical outputs: metastases-specific outcomes. Despite this, the aggregated outcomes were comparable and differences lay in the same direction (i.e., expected cost, effects and ICERs were higher in the DES). Both models led to a similar resource allocation decision16. The authors’ interpretation is that DES only provides more accurate estimates when areas of flexibility (e.g., memory-dependent transition probabilities, patient heterogeneity) are required for a large portion of the model. Outside of this, the greater flexibility conferred by a DES may be outweighed by their extensive analytic input requirements and the time needed to construct these models16.

**3.3.2 Interactivity**

The concept of interactivity is specific to infectious disease modelling and cross-validation exercises have been on the following diseases: two on influenza10,12, one on Chlamydia14 and one on a hypothetical close-contact infectious disease18. Interactivity, in this context, refers to whether interactions among individuals is simulated as this would have a bearing on how the force of infection (other synonyms include the instantaneous rate of infection, or the probability of disease exposure) is handled. A static model is one where interaction is omitted from the model and consequently, the force of infection is assumed constant over time10,18 (Table 2). If a model incorporates interactions between individuals (e.g., contact patterns or transmissibility is dependent on the distribution of infected) or with the infectious agent (e.g., evolution of the pathogen)6,19, the force of infection is instead a dynamic parameter18. Dynamic models can factor in the impact of an intervention on disease transmission. In the existing cross-validation exercises, the decision tree was always chosen to represent the static model while a compartmental or system dynamics model was selected as the dynamic model.

All studies led to the same conclusion: dynamic modelling is required in the presence of large indirect effects. Results between a static and a dynamic model closely approximated each other only when indirect effects (e.g., herd immunity) were marginal12,18. These represented instances when the force of infection was constant, such as high reproductive rates (i.e., highly contagious agent)10,18, where either no one (i.e., high clinical attack rate) or everyone was immunized (i.e., low clinical attack rate)18) or when the number of susceptible patients was lower than the disease’s epidemic threshold (i.e., herd immunity threshold is reached)12. If indirect effects were significant, the cost-effectiveness estimates in most cases led to considerably different policy recommendations14.

The magnitude of bias introduced by a static model is context-dependent: three studies found that the static model underestimated the cost-effectiveness of the new intervention, leading to a higher ICER value14, whereas one study found that the static model overestimated the cost-effectiveness10. Of greater concern, however, are the cases where the ICER estimates were found to be on different quadrants of the incremental cost-effectiveness plane12. These studies therefore highlight that, when the force of infection is nonlinear, a dynamic model may produce markedly different results compared to a static model with the magnitude and the direction of bias difficult to predict.

All authors concluded that dynamic modelling is required when both direct and indirect effects of an intervention are of interest. If interest lies only in the direct effects, a static model may suffice. Welte et al.14 further noted that static models may be appropriate when the focus is on a subset of the population with little interest in the implications on the broader population. In such cases, as only a small group within the population is studied, the likelihood of disease transmissibility amongst this subset should be low.

One paper further considered the practical implications, such as time and data availability, which are imposed when choosing between a static and dynamic model14. The authors suggest that an underlying reason for the conflicting results between dynamic and static models was due to a trade-off between validity and feasibility. Static models generally require less data and are quicker to construct. However, dynamic models are better able to capture characteristics of the infectious agent. In this context, given the sexual transmission of Chlamydia, the dynamic model was able to factor disease transmission based on sexual behaviours of partner formation and the indirect protection of screening on lowering the force of infection14. This increased validity was, however, more complex and data/time demanding. Welte et al.14 thus concluded that, by not incorporating indirect effects, the static model may have resulted in the screening of the wrong target group: the dynamic model found that screening was a dominant strategy irrespective of the age group screened, whereas the static model recommended screening for a more restrictive age.

**3.3.3 Time Advancement Mechanism**

A single study explored time advancement mechanisms within models. In discrete time models, events occur within distinct points in time; whereas, in continuous time models, events can occur at any instant. Assumming a hypothetical, three-state non-communicable illness, Chrosny et al.17 constructed both a Markov cohort and a DES model to explore the potential bias introduced into the cost-effectiveness estimates due to the manner in which time is handled. Whereas a DES model runs on continuous time, the Markov cohort model progresses on fixed cycle lengths that were varied at: one month, six months and one year cycles. The authors found that the biases in both absolute costs and absolute quality-adjusted life years estimated in the Markov cohort model, when compared to the DES, were a function of cycle duration. The biases reduced as the time in each Markov cycle shortened; in other words, the model’s results converged as discrete time more closely approached a continuous measure. A reason offered by the authors for this observation is that aggregate-level models require a simplification of competing risks.However, given the practical disadvantage of DES in that it requires a number of model iterations to stabilize the mean value, the authors recommend a Markov cohort model. Only if the ICER values are close to the willingness-to-pay threshold do they recommend either a Markov cohort model with a shorter cycle length or a DES17.

No studies were identified that explored this decision criterion for communicable diseases.

**3.3.4 Resource Constraints**

One study compared the impact of resource constraints in modelling the economic value of a 21-gene assay for breast cancer screening and its subsequent impact on personalized decision-making regarding adjuvant chemotherapy15 DES captured a constrained resource setting while a Markov cohort model represented unconstrained resources. Although the relative differences between the cost and quality-adjusted life year outcomes did not exceed 2.5% in the DES compared to the Markov cohort model, the efficiency frontiers generated by both modelling approaches were not the same. Different sets of non-dominated test-treatment strategies were suggested, depending on which modelling approaches was selected. Limited implications of the results were drawn by the author since this study was presented as a poster.

**3.4 Discussion and Conclusion**

This study highlights cross-validation exercises conducted so far in health economic modelling. For the structural decision criterion examined by more than a single study, the manner in which that particular criterion impacts a model was found to be consistent despite differences in the disease areas studied and modelling approaches used. As such, the impact of each decision criterion in terms of the practical trade-offs and the circumstances whereby resource allocation decisions changes are highlighted in Table 4.

In terms of population resolution, the differentiating elements between individual- and aggregate-level models were their capability in incorporating individual heterogeneity and memory into the model. Despite this, the cost-effectiveness results were all found to be similar between aggregate- and individual-level models. The emerging concern impacting the choice in this decision criterion was the trade-off in validity and practicality/feasibility. Studies suggested that unless flexibility is required for a large section of the model, individual-level models may not be beneficial in providing more accurate estimates compared to a simpler, aggregate-level model given their higher practical costs16.

With respect to interactivity, the conclusions were also similar: when indirect effects are large, dynamic modelling is required as a static model may severely bias the effectiveness and the economic value of the intervention10,12,18. These findings are in agreement with studies that have treated indirect effects as an element of structural uncertainty within a model. Several studies have used a single modelling approach with a switch to turn on or off the components constituting indirect effects20-23. For instance, a model on the vaccination of varicella-zoster virus to prevent two distinct diseases, varicella and herpes zoster, was found to be sensitive to whether indirect effects were incorporated. The economic value of vaccination was independent of vaccine coverage in the static model but varied according to the rate of vaccination coverage in the dynamic model21. Other studies across a range of infectious diseases have reported similar findings and specifically an underestimation in the economic value of vaccination when not incorporating herd immunity into their model, including in the areas of: meningococcal serogroup C conjugate vaccination23, childhood hepatitis A immunization20 and rotavirus vaccination22.

Traditional economic models have often assumed unlimited resources and we found one study that explored the impact of incorporating resource constraints on the economic value of breast cancer treatment. As this work came from a poster presentation, certain details remain unpublished. Although we were unable to assess this work completely, the authors have published similar work where capacity constraints were treated as a structural uncertainty within a DES model24. In that study, neglecting capacity restrictions and queuing led to disparate results that had broad implications on decision-making. The ICER estimates laid in different quadrants of the incremental-cost-effectiveness plane depending on whether resource constraints were included or not within the model24. The authors’ findings in this publication aligned with their unpublished study as different efficiency frontiers were generated from a resource-unlimited versus a resource-constrained setting15. However, more work is required in this area to determine whether these conclusions are generalizable and to what extent resource constraint may impact and lead to differences in the modelled results.

As mentioned previously, existing frameworks on how to select a modelling approach have been guided mainly by heuristics and lack empirical support5. This systematic review has identified empirical evidence that has explored some of these decision criterion used to guide the choice on a suitable modelling approach. Although there are few published cross-validation exercises and a predominant focus on only two selection criteria (namely, interactivity and population resolution), the existing body of literature has drawn similar conclusions. Although the relative importance of each of the differential factors identified are likely to vary between treatment area, on the basis of the converging results presented in this systematic review, researchers and analysts should be able to make informed choices with respect to the relative advantages of alternative decision modelling techniques to their respective treatment area.

Through the areas of convergence and divergence in cross-validated exercises, one can better understand the strengths, limits and restrictions of different modelling approaches. However, given that the ultimate purpose of modelling is to help decision-makers anticipate and take action by providing quantitative information on the consequences of the options being considered, to truly understand which modelling approach is producing the ‘best’ answer, one must consider other forms of validity. Simply knowing that different modelling approaches will produce different results would not be sufficient if one does not know which result is indeed the correct one! Two additional forms of validity that can address this issue are external and predictive validity. External validity involves comparing the model’s predicted results with real-world data while predictive validity involves comparing a model’s forecast with prospectively observed events25 Such validation is necessary to be confident that a model is indeed able to calculate the actual outcome if given the appropriate data (i.e., external validity) and is fulfilling its purpose in prediction (i.e., predictive validity). The majority of the cross-validation studies have, to an extent, conducted external validity exercises but only one was found to address predictive validity11. We thus encourage modellers to consider, when conducting cross-validation exercises, to not simply address cross-validity but to extend their exercises to these other two forms of validity when possible.

Indeed, with the growth in health economic modelling, it is important for modellers to now come together to discuss differences in their models for the same decision problem in terms of structure and performance. Greater cross-collaboration should be fostered with an emphasis on comparing models and their underlying validities. Some promising work has emerged in the area of diabetes economic modelling. The Annual Mount Hood Diabetes Challenge has brought together modellers to discuss and explain differences between their models when modelling the same set of decision problems26. Each challenge involves systematic comparisons and validation exercises that not only addresses cross-validity between modelling approaches but also considers external and predictive validity of each model. Results from the Fifth Mount Hood challenge highlight that, in comparing the population resolution, individual-level models do have greater computing demands but also permit additional flexibility such as capturing the covariance between different risk factors. Despite using the same data to estimate risks, models with greater flexibility were found to produce different results and it remains unclear which has higher predictive validity26.

Current empirical work on cross-validation has focused on comparing a traditional modelling approach (i.e., decision tree or Markov cohort) against an alternative form (i.e., microsimulation, DES or dynamic compartmental modelling). However, studies are somewhat lacking on addressing certain modelling approaches such as agent-based modelling. Future research should continue to expand this knowledge through comparative exercises to address whether different modelling approaches do indeed produce different results and, if so, under which circumstances. Conversely, this may address the situations whereby different modelling approaches may produce similar results, such that, in these cases, a more simplistic model would be preferable. With greater empirical knowledge, this may foster the development of evidence-based frameworks on how to choose a modelling approach among the many options available rather than the existing approach of heuristic-based guidelines. Presently, there has been little work focused on this area as highlighted by the limited number of studies identified from the systematic review. Despite the plethora of decision-analytic health economic models published and the growing reliance by policy makers in using evidence generated by such models, few have actually debated the value and the importance that different decision criteria may play on the selection of a modelling approach, its subsequent process of construction and its predicted results. Additional research on this topic does, however, represent an important endeavour, as selecting a suitable modelling approach will impact the validity of the underlying health economic model and ultimately the resource allocation decision it seeks to inform.

|  |
| --- |
| **Key messages:**   * Each modelling approach will impose certain constraints and assumptions that may impact the results and consequently the resource allocation decision it seeks to inform. As such, specific circumstances may exist on when each approach would be the most appropriate. * Decision criteria have been published to distinguish the features characterizing each specific modelling approach. However, few studies have addressed whether and how these decision criteria are important empirically. Cross-validation studies can provide insight into this topic. * Only nine studies were identified that have conducted cross-validation between different modelling approaches. Among the decision criteria studied, relatively consistent themes emerged from the empirical work regarding the impact of a decision criteria on the construction and overall results to a model. * Aggregate- and individual-level models often produce similar results and qualitative conclusions. Rather, the choice on a model’s resolution is important in terms of the potential trade-off between model validity and feasibility. * Interaction must be captured when indirect effects are important and are expected to be large. Studies consistently show different results estimated by dynamic (i.e., captures both direct and indirect effects) and static models (i.e., captures only direct effects) that often lead to conflicting policy recommendations. * Overall, there has been limited work done on this topic as only nine studies were identified from a systematic literature review of bibliographic databases and grey literature. Additional research will thus be necessary to not only verify the existing understanding of the aforementioned decision criteria that have been studied more extensively but also to address the importance of the remaining less studied decision criteria (e.g., time advancement mechanism, resource constraints). * Understanding the implications of the various structural features specific to a modelling approach is necessary to foster knowledge on the benefits and limits of each modelling approach. This would provide insight into how to select a suitable modelling approach for a particular decision problem. |

**Acknowledgement/ Sources of financial or other support:**

We would like to thank Kaitryn Campbell for her assistance in developing the search strategy for this systematic review.

Bernice Tsoi is supported through awards from the Father Sean O’Sullivan Research Centre, St. Joseph’s Healthcare Hamilton and Pfizer Graduate Student Award in Health Technology Assessment.

The authors have no other potential relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.

**References**

1. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modeling practices: a suggested framework. *Value Health* 1998; 1(2), 131-147.

2. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008; 26(2), 131-148.

3. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004; 9(2), 110-118.

4. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006; 15(12), 1295-1310.

5. Tsoi B, O'Reilly D, Jegathisawaran J, Tarride JE, Blackhouse G, Goeree R. Systematic narrative review of decision frameworks to select appropriate modelling approaches for economic evaluations. *BMC Res Notes*; 2015; *manuscript submitted.*

6. Kim SY , Goldie SJ. Cost-effectiveness analyses of vaccination programmes : a focused review of modelling approaches. *Pharmacoeconomics* 2008; 26(3), 191-215.

7. 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(6), 843-850.

8. Meadows D, Robinson J. The Electronic Oracle. Computer Models and Social Decisions. John Wiley & Sons, Chichester, UK; 1985.

9. Standfield L, Comans T, Scuffham P. Markov modeling and discrete event simulation in health care: a systematic comparison. *Int J Tech Assess Health Care* 2014; 30(02), 165-172.

10. Lugner AK, Mylius SD, Wallinga J. Dynamic versus static models in cost-effectiveness analyses of anti-viral drug therapy to mitigate an influenza pandemic. *Health Econ* 2010; 19(5), 518-531.

11. Simpson KN, Strassburger A, Jones WJ, Dietz B, Rajagopalan R. Comparison of Markov model and discrete-event simulation techniques for HIV. *Pharmacoeconomics* 2009; 27(2), 159-165.

12. Pradas-Velasco R, Antonanzas-Villar F, Martinez-Zarate MP. Dynamic modelling of infectious diseases: an application to the economic evaluation of influenza vaccination. *Pharmacoeconomics* 2008; 26(1), 45-56.

13. Kuehne FC, Chancellor J, Mollon P. Microsimulation or cohort modelling? A comparative case study in HIV infection. iHEA 2007 6th World Congress: Explorations in Health Economics Paper. 2007. Available from: <http://ssrn.com/abstract=994349>.

14. Welte R, Postma M, Leidl R, Kretzschmar M. Costs and effects of Chlamydial screening: dynamics versus static modelling. *Sex Transm Dis* 2005; 32(8), 474-483.

15. Jahn B, Rochau U, Arvandi M *et al.* Lessons learned from a cross-validation between a discrete-event simulation model and a markov model for personalized breast cancer treatment. *Value in Health* Conference: ISPOR 15th Annual European Congress Berlin Germany. Conference Start: 20121103 Conference End: 20121107. Conference Publication:(var.pagings), A283. 2012.

16. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ* 2003; 12(10), 837-848.

17. Chrosny W, Stevenson M, Munzer A. Comparison of Markov and discrete event simulation modeling techniques with application to cost effectiveness analysis. *Value Health* 2013; 16(7), A587.

18. Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. *Stat Med* 199; 18(23), 3263-3282.

19. Jit M , Brisson M. Modelling the epidemiology of infectious diseases for decision analysis: a primer. *Pharmacoeconomics* 2011; 29(5), 371-386.

20. Armstrong GL, Billah K, Rein DB, Hicks KA, Wirth KE, Bell BP. The economics of routine childhood hepatitis A immunization in the United States: the impact of herd immunity. *Pediatrics* 2007; 119(1), e22-e29.

21. Brisson M , Edmunds WJ. Impact of model, methodological, and parameter uncertainty in the economic analysis of vaccination programs. *Med Decis Making* 2006; 26(5), 434-446.

22. Kim SY, Goldie SJ, Salomon JA. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. *Med Decis Making* 2010; 30(5), E1-E28.

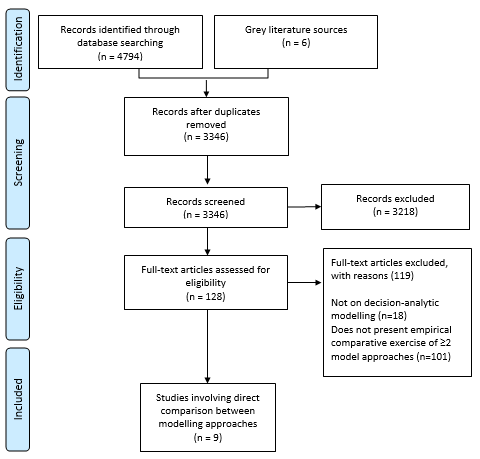
23. Trotter CL , Edmunds WJ. Reassessing the cost-effectiveness of meningococcal serogroup C conjugate (MCC) vaccines using a transmission dynamic model. *Med Decis Making* 2006; 26(1), 38-47.

24. Jahn B, Pfeiffer KP, Theurl E, Tarride JE, Goeree R. Capacity constraints and cost-effectiveness: a discrete event simulation for drug-eluting stents. *Med Decis Making* 2010; 30(1), 16-28.

25. 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(6), 843-850.

26. Palmer AJ, Clarke P, Gray A *et al.* Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. *Value Health* 2013; 16(4), 670-685.

**Figure 1:** PRISMA diagram of the literature search process for documents that have conducted cross-validation between different modelling approaches



*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**Table 1:** Approaches to decision-analytic modelling.

|  |  |  |
| --- | --- | --- |
| **Modelling approaches** | **Technical description** | **Typical areas of**  **application** |
| Decision tree | Decision trees typically arrange events in temporal order, from left to right (i.e., steps on the left occur earlier in time than those on the right). A decision can be broken down into three broadly defined components:   1. **Decision node**: indicates competing strategies. 2. **Chance node**: consequences of a particular decision, which are mutually exclusive. 3. **Terminal node**: value of a branch.   **Branches** connect the nodes and represent the pathways through the tree. At each chance node, the probabilities of each consequence will determine the proportion of patients progressing down each unique path.  Consequences such as costs and outcomes of events and decisions may be attributed to each chance node or be accumulated at the terminal nodes. The expected costs and/or effects associated with each strategy are estimated by “**rolling back”** the tree (weighted averaging of the values of all branches emanating from a decision node). | Decision trees illustrate the logic and sequencing of decisions in an easy way to interpret, especially if the number of branches is kept low.  A common usage is when the disease is acute and events are not recurrent. If the model time horizon increases or too many branches exist, the tree can become ‘bushy’ and difficult to manage. |
| Markov cohort model | Markov cohort models describe the movement of patients over time as they progress through **health states**. Movement occurs only once per **Markov cycle** and is governed by specific **transition probabilities**. The model is run through many **fixed-time** Markov cycles up to a specified model time horizon. Patients are assumed to be in one of a finite number of health states (known as the **unitary state requirement**). Patients within the same health state are considered homogeneous.  The basic Markov cohort model relies on the **Markovian assumption** (also referred to as the ‘memoryless’ assumption) whereby the transition probabilities depend only on the current state and not on any previous health states.  Typically, the entire cohort will enter the model together at the same time although they can be distributed among various states. Costs and health outcomes are attached to health states as transition probabilities are value-free. The expected outcome for a particular time point, the **cycle sum**, is the weighted average of the value of the states. Overall costs and QALYs can be calculated by adding the cycle sums over the model’s time horizon. | A Markov cohort can be mathematically equivalent to a decision tree. However, the Markov model provides a more manageable representation if the modelled time horizon is long or if events recur. The basic Markov cohort model is often said to be memory-less (i.e., can only remember the previous health state and progression of disease is not impacted by prior history).  If the decision problem can be represented with a manageable number of health states and incorporates all characteristics relevant to the decision problem, a cohort simulation should be chosen. |
| Markov microsimulation | Markov microsimulation focuses on simulating an individual patient over time. The assumption of homogeneity is not necessary, as each patient is modelled individually. As this approach is capable of storing the history of each individual patient over the course of the model, **memory** is explicitly captured. Patient prognosis is governed by event/transition rates although these may be conditional on previous and existing risk factors and events. Transitions occur only once per cycle and, similar to the Markov cohort model, the unitary state requirement remains as a patient can only be in one of the health states during each cycle.  Consequences, such as costs or health outcomes, are attached to health states and the values may further be dependent on a patient’s characteristics or history.  Following the simulation, each patient has their own respective costs and outcomes. The expected costs and QALYs can then be calculated as the average from a large number of simulated patients. | Compared to cohort-level models, Markov microsimulation differs in that it models each patient individually. It solves one of the key issues with Markov cohort modelling as individual memory can exist.  If a valid representation of any aspect of the decision problem would lead to an unmanageable number of health states in a Markov cohort model, then a microsimulation is preferred. |
| Discrete event simulation (DES) | Discrete event simulation describes the progression of **entities**, which in the case of healthcare modelling, refers to individuals. Rather than a fixed-time, time is **continuous** with patient progression sampled according to parametric or empirical time-to-event distributions. However, other approaches to advancing the clock include step or periodic approach, which is similar to the Markovian cycles. Entities may be characterized with **attributes** and common attributes include individual characteristics or memory. This may influence their progression through the model and/or the length of time between events. Another important concept is r**esources**, an object that provides service to an entity.  Individuals may either be simulated one-by-one or simultaneously. If simulated simultaneously, one can model entity interactions indirectly through resourcecompetition, thereby, explicitly capturing effects of **queuing**.  Consequences, such as costs and effects, can be attributed to anything that is sensible such as events, time with a particular condition or simply having a particular patient attribute within the model. Following the simulation, since each patient has their own respective costs and outcomes, the expected costs and QALYs is simply the average among this group. | Discrete event simulation has greater flexibility as the model runs on continuous time (e.g., simultaneous events can occur). Furthermore, it is useful in modelling situations of constrained resources (e.g., number of hospital beds, organ transplants) or process-driven situations (e.g., queues, waitlists). As individuals have collections of memory, heterogeneity can be captured. However, this is a data-intensive method. |
| Agent-based models | This approach focuses on the **agent**. Simple rules govern their communication and interaction with other agents or with their environment. Rules may relate to adaptation over time, learning or a variety of other options. From these simple rules that govern individual actions and communication, complex behaviour may emerge.  Individuals are simulated simultaneously through the model. Consequently, agents exist within a **network** that may adopt a variety of structures including hierarchical (e.g., social network). Agents (and their behavioural rules) may further present with spatially considerations.  Consequences such as costs and effects can be attributed to specific events or patient attributes that are updated in continuous time. | Few applications in health economics although it can be useful if the decision problem have important features of space, geography or network structures (e.g., transmission of infectious diseases, social health promotions). Given the individual level of the model, heterogeneity and memory can be easily incorporated. |
| System dynamics | System dynamics describes a system through feedback loops, stocks and flows. The **causal loop diagram** provides a qualitative visualization of a system. Its basic building block is the **feedback loop**, describing movement (i.e., flow) from one pool (i.e., stock) and eventually returning in some form to its origin. Individuals within the same stock are assumed homogeneous and thus the model does not easily handle memory or heterogeneity.  To quantify a system dynamics model, stock and flow diagrams are used. As per its name, these diagrams consist of two main types of variables: **stocks** (i.e., levels or state) and **flows** (i.e., rates at which stocks are either drained or replenished). Movement between stocks is defined by the rate of flow, dictated by differential equation, and time changes continuously.  Costs and outcomes may be linked to the time spent in a particular stock or by the flow between stocks. | Little application observed except in the area of infectious diseases where differential equations are taken from mathematical models of infectious disease epidemiology. Useful when feedback is an important feature in the decision problem (e.g., patient flow through emergency department, human homeostasis processes). |
| Compartmental models | The population is divided into various **compartments**, representing their average state. Individuals within a single compartment are considered homogeneous. Most commonly, it contains compartments of the population whom are at different stages of the illness. | Compartmental models have exclusively been used to model the transmission and epidemiology of infectious disease (e.g. **susceptible – exposed – infectious –recovered**). |

Key terms specific to a modelling approach are bolded.

QALYs: quality-adjusted life years

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**Table 2:** List of selection criteria commonly used in existing frameworks to guide choice on modelling approach.

|  |  |  |  |
| --- | --- | --- | --- |
| **Structural element** | | | |
| **Name of criteria** | **Description (key term bolded)** | **Significance to modelling** | **Examples of modelling approaches falling under each criteria** |
| Population resolution | **Individual** or **aggregate/cohort** level data (i.e., means, medians) are used to populate the model parameters. Cohort characterizes the average patient experience while individual level models capture the variability between patients with their pathway dictated by their history and attributes. | The choice depends on whether patient variability and history is important. | **Individual:** Markov microsimulation  **Aggregate/Cohort**: Markov cohort model |
| Interactivity | Interactivity refers to whether interaction among individuals or their environment needs to be simulated. **Independent** models ignore any interaction effects between individuals and their environment while **dependent** model captures these effects. The terminology, **static** and **dynamic**, are respectively used in infectious-disease modelling. Interactivity in infectious-disease modelling refers to the force of infection as this is a function of the number of infectious individuals in the population at a given time in a dynamic model | Interactions may impact the disease progression and the development of the outcomes of interest. | **Independent:** Decision tree  **Dynamic:** Compartmental model |
| Resource constraints | The capability of a model to integrate utilization of multiple resources. Resources may be considered as infinite (**unlimited**) or finite (**limited**). | Simultaneous resource use and/or resource constraints may introduce unexpected behaviours, such as queues or delays, that impact patient outcomes. | **Unlimited:** Decision tree  **Limited:** Discrete event simulation |
| Time advancement mechanism | **Untimed** models do not explicitly consider the progression of time. If time is advanced explicitly, it can be treated as **discrete** time step (i.e., distinct value, such as an integer) or on a **continuous** scale (i.e., can take on any non-negative value). | Method by which time advances impacts the format of the input and output data. | **Untimed:** Decision tree  **Discrete:** Markov models (cohort and microsimulation)  **Continuous:** System dynamics |
| **Practical considerations** | | | |
| **Name of criteria** | **Description** | **Significance to modelling** | |
| Time | The speed of model development, which is impacted by the time required to: construct the model, collect the necessary data and generate simulation results (simulation time). | The time required to have a complete model should align with the time that is available. | |
| End-user requirement | The end-user requirement is dependent on may factors including the validity of the model and their understanding of what is modelled. | This impacts the acceptability and usage of the model, and may impact the funding support available for the modelling project. | |
| Simplicity | The degree of complexity in a model, which is essentially dependent on the size of the model and the number of parameters required. | Simpler models are more likely to be easier explained and understood. | |
| Validity | The clinical representativeness of a model to the actual decision problem. | The extent to which the model reflects and captures all relevant aspects of the decision problem. | |

These criteria have been separated into: structural elements (i.e., technical features specific to the concepts and principles behind a model) and practical considerations (i.e., factors that impact the effectiveness or feasibility of developing and constructing a model). The desired structural elements are often dictated by the nature of the decision problem while practical considerations are, to an extent, context-dependent.

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**Table 3:** Summary of empirical studies that have compared two or more modelling approaches

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Study (year)** | **Modelling approaches** | | | | | | | **Decision criterion explored** | |
| **Decision tree** | **Markov cohort** | **Markov microsimulation** | **DES** | **Agent-based model** | **System Dynamics** | **Compartmental model** | **Structural element** | **Practical considerations** |
| Infectious diseases | Hypothetical Disease | | | | | | | | | |
| Edmunds et al. (1999)18 | X |  |  |  |  |  | X | Interactivity | N/A |
| HIV | | | | | | | | | |
| Kuehne (2007)13 |  | X | X |  |  |  |  | Population resolution | Simplicity  Ease of conducting probabilistic sensitivity analysis |
| Simpson et al. ( 2009)11 |  | X |  | X |  |  |  | Population resolution | Time  Data availability |
| Influenza | | | | | | | | | |
| Pradas-Velasco et al. (2008)12 | X |  |  |  |  |  | X | Interactivity | N/A |
| Lugner et al. (2010)10 | X |  |  |  |  |  | X | Interactivity | N/A |
| Chlamydia | | | | | | | | | |
| Welte et al. (2005)14 | X |  |  |  |  | X |  | Interactivity | Time  Data availability |
| Non-communicable diseases | Hypothetical | | | | | | | | | |
| Chrosny et al. (2013)17 |  | X |  | X |  |  |  | Handling of time† | N/A† |
| Breast Cancer | | | | | | | | | |
| Karnon  (2003)16 |  | X |  | X |  |  |  | Population resolution | Time  Simplicity |
| Jahn et al. (2012)15 |  | X |  | X |  |  |  | Resource constraints† | N/A† |

DES: discrete event simulation; HIV: human immunodeficiency virus

† Only a poster abstract was available for this study and therefore, it is uncertain whether the authors looked at other decision criteria

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**Table 4**: Current understanding of the impact of each structural decision criterion

|  |  |  |  |
| --- | --- | --- | --- |
| **Structural decision criteria**  **(# of studies)** | **Practical trade-offs in model construction** | **Implication for resource allocation decision** | **Recommendation** |
| Population resolution (3) | Validity-feasibility trade-off: Individual-level models more easily and compactly reflected the disease (e.g., memory, heterogeneity, time-dependencies) while incorporating data in its existing form. Aggregate-level models, although disadvantaged in those respects, had shorter time in terms of computing, validation and conduct of probabilistic sensitivity analysis. | Little difference in the estimated outcomes (i.e., cost, effects, ICER) between individual-level and aggregate-level models. | Choice on a model’s resolution depends on practical consideration in terms of the extent to which a given modelling approach can flexibly reflect the necessary details of a particular research question. |
| Interactivity (4) | Validity-feasibility trade-off: Static models generally require less data and are quicker to construct. However, dynamic models may have greater validity as it can more easily capture characteristics of the decision problem that may be relevant and important. | Results closely approximates when indirect effects are marginal. Such instances may include:   * Constant force of infection (e.g., highly contagious agent and/or limited vaccination rate as the entire population is expected to be infected) * Low likelihood of disease transmission (e.g., when the critical threshold in the population are vaccination since herd immunity is attained).   If indirect effects are significant, models may generate different and opposing resource allocation recommendations. | Dynamic models are required in the presence of large indirect effects unless the decision problem is interested in only the impacts and implications of direct effects. |
| Handling of time (1) | Not addressed in any cross-validated work. | As discrete cycle lengths shortened in Markov cohort model (i.e., approached continuous time such as in the DES), modelled results converged. | There is limited evidence that one should begin with a model that handles time discretely due to practical advantages. If the ICER values is however found to be close to the willingness-to-pay threshold, one should consider shortening the cycle length or switching to a model that runs on continuous time. |
| Resource constraint (1) | Not addressed in any cross-validated work. | Relative difference in the expected costs and QALY outcomes for each treatment strategy were similar (<2.5%) between an unconstrained Markov model and a constrained DES. However, efficiency frontiers for the given decision problem differed between modelling approaches. | There is limited evidence that resource constraints has an important role in modelling the resource allocation decision. |

DES: discrete event simulation; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life years

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**APPENDIX 3.I:** Search Strategy

|  |  |
| --- | --- |
| Database: Ovid MEDLINE(R) <In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present> | |
|  | [(Economic evaluation\*[Title/Abstract] OR economic outcome\*[Title/Abstract] OR economic analy\*[Title/Abstract] OR health economic\*[Title/Abstract] OR pharmacoeconomic\*[Title/Abstract] or pharmaco-economic\*[Title/Abstract])  Filters: English (13550) |
|  | (decision\* analy\* [Title/Abstract] OR modeling approach\* [Title/Abstract] OR modeling method\* [Title/Abstract] OR modeling practice\*[Title/Abstract] OR modeling technique\*[Title/Abstract] OR modelling approach\*[Title/Abstract] OR modelling method\* [Title/Abstract] or modelling practice\*[Title/Abstract] OR modelling technique\*[Title/Abstract])]  Filters: English (63812) |
| 3. | Exp Models, Economic (2962) |
| 4. | 2 and 3 (2423) |
| 5. | 4 or 5 (5207) |

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**APPENDIX 3.II:** Level I Selection Criteria

|  |
| --- |
| **Modelling Framework ~ Selection Criteria (LEVEL I SCREENING)**   1. Is this study focused on economic evaluation methods?   No (exclude) Yes (include) Maybe (include)   1. Does this study directly compare one modelling approach\* over another?   No (exclude) Yes (include) Maybe (include)   1. Is this study published in English?   No (exclude) Yes (include) Maybe (include)  \*NOTE: Markov model, state-transition model, decision tree, discrete event simulation, agent-based model, microsimulation, system dynamics |

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**Appendix 3.III:** Summary of the Results of Empirical Studies that have Compared Two or More Modelling Approaches

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Study (year)** | **Objective** | **Key methodological differences between modelling approaches** | **Results:** | |
| **Quantitative** | **Qualitative**  *(e.g., Different qualitative conclusions reached between modelling approaches?)* |
| Edmunds et al. (1999)18 | To highlight the impact on cost-effectiveness estimates when indirect effects are incorporated by comparing static (solely direct effects) and dynamic model (inclusion of both direct and indirect effects). | To ensure models are comparable, the dynamic compartmental model (i.e., Susceptible-Infectious-Recovered) was modified. Rather than the conventional monitoring of different birth cohorts over their lifetime, the dynamic model was set to monitor a single birth cohort with vaccination applied to only this cohort. Earlier and later birth cohorts were either assumed to have zero vaccination coverage or continuous vaccination that would lead to steady state equilibrium. | • Identical results between modelling approaches at the extreme ends of coverage (i.e., when entire cohort or no one is immunized). At other levels of immunization, cohort model underestimates number of cases prevented and underestimates the ICER (if the effectiveness outcome is cases of infection).  • ICER when single cohort vaccinated at 90% level with no vaccination in the other cohorts:  Cohort: $13.95/case avoided  SIR: $13.91/case avoided | • Cohort model closely approximates dynamic model if outcome of interest is specific to a single cohort, the basic reproductive rate is high, and either very few or many within the cohort are immunized.  • Cohort model sufficient when force of infection not substantially altered by mass immunization &/ when the decision maker is interested in a subset of the population. |
| Kuehne (2007)13 | To compare and validate a Markov cohort and Markov microsimulation with respect to its cost-effectiveness. | Progression in both models was based on CD4 cell count, viral load set point, treatment, response, age, and opportunistic infections. The Markov cohort model required more disease states to incorporate history and time dependency. The Markov microsimulation mitigated the need for additional disease states by handling transitions through tracker variables that captured an individual’s case history (e.g., prognostic factors) and the time elapsed. | • Differences in the total costs and total QALYs were in the same direction: Markov cohort model estimates were lower compared to DES.  Costs [Cohort/ microsimulation]:  New: $180100/ $201600  Standard: $148500/ $156700  QALEs [Cohort/ microsimulation]:  New: 4.29/ 4.59  Standard: 3.64/ 3.60 | Comparable estimates although each model has individual strengths/ limitations in terms of a trade-off between validity and feasibility. |
| Simpson et al. (2009)11 | To compare how Markov cohort model and DES perform when using data from a clinical trial to estimate and extrapolate ICER for two competing treatments over life times. | None described. | • Results similar when timeframe short (1 year) but DES has slightly better predictive ability for longer durations:  % with viral load<400 copies/mL at year one [Markov/DES/ actual]: 65/ 61-66/ 65  % with viral load<400 copies/mL at year one [Markov/DES/actual]: 50/ 57-60/ 56  Results are within 3% margin of error.  • Variations in outputs were in the same direction with DES providing higher estimates:  Costs [Markov/ DES]:  New: $318882/ $352843  Standard: $310194/ $340022  QALYs [Markov/ DES]:  New: 10.11/ 12.11  Standard: 10.55/ 12.40 | Both modelling approaches found that one strategy dominated the other. |
| Pradas-Velasco et al. (2008)12 | To evaluate the economic efficiency between static (i.e., decision tree) and dynamic model (i.e., compartmental model). | None described. Dynamic model took into account the variable force of infection (i.e., incorporated both direct and indirect effects) while static model maintained a constant force of infection (i.e., only direct effects captured). | • Base-case rate of return per € invested:  Dynamic model: 1.22  Static: 0.28  • Results converge when indirect effect becomes less important (i.e., when number of susceptible < epidemic threshold of susceptible). Rate of return per € invested:  Dynamic model: 1.88  Static: 1.21 | As rate of return >1 indicates a health care intervention with an efficient outcome, results from static model suggest intervention is not efficient; dynamic model suggests the contrary. |
| Lugner et al. (2010)10 | To examine the differences and sensitivities to epidemiological factors between static (i.e., decision tree) and dynamic models (i.e., compartmental model) in terms of estimating the incremental cost-effectiveness ratio (cost/ life year gained (LYG)) for the treatment of an influenza pandemic. | Effects of treatment in static model based on prevalence of influenza. Dynamic models extend this by modelling the number of individuals whom develop influenza based on the characteristics of the virus, the contact structure in the population and the treatment effects on transmission. To ensure model comparability, prevalence of influenza in the static model was calibrated based on the overall clinical attack rate that was predicted in the dynamic model. | • Static models predicted 1.4 times higher number of deaths than dynamic model under both treatment strategies.  • In the base case analysis, ICER between modelling approaches were similar:  Static model: €1695/ LYG  Dynamic model: €1637/ LYG  • Estimates between the two modelling approaches differed for:  i) Drug usage: lower use of treatment led to divergence with dynamic model predicting a higher ICER than the static model.  ii) Size of epidemic/ clinical attack rate: ICER in dynamic model is sensitive to this parameter but remains constant in static model. | • Both models led to the same conclusion as ICER< €20,000/LYG, even across different scenarios (e.g., drug usage, pandemic size). However, if a different threshold was use, recommendation may differ. |
| Welte et al. (2005)14 | To compare static (i.e., decision tree) vs. dynamic models (i.e., system dynamics) in terms of the type of data required, the computed results and model parameter’s sensitivity. | The dynamic model simulates prevalence of the disease according to disease characteristic, patient behaviour and the treatment. This parameter varies with time. The static model requires prevalence of the disease as a fixed input variable that remains constant. To ensure model comparability, the number of tested women per year was calculated in the dynamic model, and calibrated to the static model. | • Static model predicted fewer negative health outcomes prevented and lower averted costs  Net costs:  Static: $43,040  Dynamic]: $-78,800  Net outcomes:  Static: 62  Dynamic: 123  • ICER, in all the scenarios studied, were divergent:  Static model: $700/ major outcome averted  Dynamic model: Dominant  • Dynamic model sensitive to parameters that affect force of infection (e.g., screening duration, contact networks); static model is insensitive to such parameters. | • Dynamic models and static models lead to very different results.  • Static models identified the incorrect optimal age group to target screening. |
| Chrosny et al. (2013)17 | To assess the bias in costs, QALYs and ICER between Markov cohort models and DES. | None described | • Bias in costs and QALYs introduced by Markov cohort models were reduced with shorter time cycles: cost bias fell from 14% to 1% when time cycle changed from one year to one month. Difference in QALY between modelling approaches were <1% irrespective of duration of time cycles.  • Bias in ICER: 2.4%-9.6% when Markov cohort time cycles were one year in duration; 0.6%-5.4% when one month time cycles. | None described although recommends that if ICER in Markov cohort model are close to the willingness-to-pay threshold, either the cycle length should be reduced in the Markov cohort model or a DES should be built. |
| Karnon (2003)16 | To compare between a Markov cohort model and a DES in terms of:  i) processes (i.e., flexibility [how well reality of patient pathways modelled], and analytic inputs [ease and feasibility of constructing the model]); and  ii) outputs (e.g., representation of disease progression, cost-effectiveness). | • Structural difference in how toxicity is incorporated. As three types of drug-induced toxicities exist, these were incorporated as separate health state events with a total of seven toxicity health states to represent all the possible permutations in the Markov cohort model (due to Markovian assumption) whereas the DES model incorporated toxicity as an attribute within a single health state.  • Disease progression in a Markov cohort model was based on transformation of data to constant transition probabilities while a DES incorporated it in its existing format (i.e., survival curve). DES facilitated a more flexible representation of the available data. | • The final analysis took 1 hour and 3 days to run for the Markov cohort model and the DES, respectively.  • Variations in outputs were in the same direction with DES providing higher estimates:  Costs [Markov/ DES]:  New: ₤8740/ ₤9146  Standard: ₤6721/ ₤7115  QALYs [Markov/ DES]:  New: 12.00/ 12.14  Standard: 11.40/ 11.56  • Similar ICER:  Markov: ₤3365/QALY  DES: ₤3483/QALY | • Both models are likely to provide similar recommendation.  • The cost-effectiveness acceptability curve (CEAC) suggests a slightly higher probability that new therapy produces positive net benefits across all willingness-to-pay thresholds in the Markov cohort model compared to the DES. |
| Jahn et al. (2012)15 | To cross-validate a Markov cohort model and DES by examining how closely both modelling approaches describe the natural disease progression and cost-effectiveness. | None described | Cost estimates presented a smaller relative difference compared to QALYs with differences <2.5% between modelling approaches. | The modelling approaches suggested a different set of interventions formed the efficiency frontier. |

*Tsoi B et al, Expert Rev Pharmacoecon Outcomes Res, 2015; [epub ahead of print] 1-13, copyright © 2015, Informa Healthcare. Reproduced with permission of Informa Healthcare*

**CHAPTER 4:**

**Comparison of cost-effectiveness results by decision trees, agent-based and system dynamics models: A case of influenza vaccination**

Tsoi B BSc, MSc, PhD(cand)1,2ŧ, Osgood N PhD3, Tarride J-E BA, MA, PhD1,2,4, Oraji R, BSc, MSc, PhD3, Blackhouse G BA, MBA, MSc1,2, Goeree R BA, MA1,2,4 O’Reilly D BSc, MSc, PhD1,2,4,

1. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

2. PATH Research Institute, St. Joseph’s Healthcare Hamilton, Ontario, Canada

3. Department of Computer Science, University of Saskatchewan, Saskatoon, Canada

4. Centre for Evaluation of Medicines (CEM), St. Joseph’s Healthcare Hamilton, Ontario,

Canada

ŧ **Corresponding author**

Cited as: Tsoi et al. Comparison of cost-effectiveness results by decision tress, agent-based and system dynamics models: A case of influenza vaccination. 2015. *Med Decis Making*: in submission

**Abstract**

Introduction: Decision trees have traditionally been the modelling approach used to assess the economic value of vaccinations. However, this approach fails to capture the complexities in transmission dynamics. Alternative approaches proposed include dynamic models, such as agent-based model (ABM) and system dynamics (SD). This study compares the performance and results generated by decision tree, ABM and SD in assessing the cost-effectiveness of influenza vaccination.

Methods: An existing decision tree comparing intranasal live attenuated vaccine (LAIV) against injectable inactivated influenza vaccine in children was adapted into SD and ABM. Both were constructed in AnyLogic 7.0. Differences in expected prevalence and costs for each strategy along with the incremental cost-effectiveness ratio (ICER) generated by each modelling approach were compared.

Results: Model calibration was successful with all three approaches producing similar estimates when identical model parameters and assumptions were adopted. LAIV was dominant regardless of modelling approach. Scenario analyses revealed that: (1) ICER varied according to the proportion and schedule of vaccination under both dynamic models; (2) economic value of vaccination was sensitive to the network topology in ABM, and; (3) heterogeneity from age-specific parameters, most easily captured by ABM, impacted the infection prevalence and, ultimately, the ICER.

Conclusions: Estimates of costs and effects may differ according to the modelling approach selected and the assumptions employed. ABM, an individual-level approach, was the most flexible as it can capture patient heterogeneity and model individual’s behaviours within their network. SD, typically an aggregate-level model, was limited in capturing patient heterogeneity and required the assumption of random-mixing between individuals. The most rigid amongst these approaches was the decision tree as it inherently relied on a broader set of simplifying assumptions.

**4.1 Introduction**

Decision-analytic health economic modelling represents a set of techniques that informs decisions on the reimbursement and diffusion of healthcare technologies. All models involve an inherent simplification in which real world complex problems are abstracted into simpler components, with the key pieces distilled and the rest omitted. A general recommendation guiding modellers is the notion that models should be kept as simple as possible1-3 with additional complexity added only if there is a justifiable incremental benefit (i.e., more reliable and/or unbiased results)4.

However, given the complexities of many health problems, concerns exist that traditional static approaches (i.e., Markov cohort models and decision trees) may be oversimplifying health processes and, thus, producing biased estimates. There is a growing recognition on the limits of traditional static models as they have been shown to miss crucial effects and mislead findings on what would be deemed most cost-effective5-7. Furthermore, the approach selected can impose constraints on a model’s development and conceptualization, such as what elements are captured, its level of detail and the ease of incorporating data8.

In recent years, growth has been observed in the application of alternative techniques that can better capture certain complexities in healthcare, such as state-transition microsimulation and discrete event simulation, to address problems otherwise difficult with traditional methods9. The potential application of agent-based models (ABM) and system dynamics (SD) have been noted in the health economics literature10 although few have been published. This paper thus aims to introduce the readers to both of these modelling approaches. We explain how these approaches can be used to produce cost-effectiveness estimates and explore when each should be adopted by using an example comparing two vaccines for influenza In addition, we conduct controlled simulations to compare and contrast the performance and results of decision trees, ABM and SD. Such exercises can help better understand the conditions in which each modelling approach may be most useful.

**4.2 Agent-based and system dynamics modelling**

We begin our inquiry by first introducing these two modelling approaches. Both belong within system sciences and emerged as methods to understand dynamic systems. Dynamic systems differ from static ones in that rules specify the changes in the system as a function of its current state, with the relationship between inputs and outputs varying over time. At its foundation, both ABM and SD aim to elucidate the behaviour of complex systems as a whole and to examine the dynamic interrelationships between different components within a system that may span across multiple levels11.

SD, developed in the 1950s, describes a system according to feedback loops, stocks and accumulations. The basic building block is the feedback loop whereby change at one point within a system triggers a cascading series of changes that ripple through the system to either reinforce or “push back against” that original change. Individual feedback loops can have profound effects on system’s behaviour, and complex behaviour may emerge from the interaction of multiple feedback loops. The causal loop diagram qualitatively visualizes the system’s structure and behaviour12 (Figure 1) whereas, stock and flow models quantify it. Movement between stocks (i.e., levels/accumulations) are based on the rate of flow which are specified as a function of the system’s current state. Stocks represent the system’s state or memory, and are often the cause of system inertia while flows dictate the change in the system over time. A stock-and-flow characterization of a system can be readily translated into a set of ordinary differential equations12*.* SD is typically a cohort-level model with the individuals aggregated into stocks assumed to have similar properties (Table 1).

ABM is an individual-level approach focused on autonomous agents and their interaction with each other and the environment (Table 1). As a result of individual agents’ pre-defined rules of interactions, system-level behaviours may emerge13. A tutorial has recently been published on how to construct ABM14. Features of agents may include: reactivity (i.e., perceiving local or global changes in the environment and responding accordingly); proactivity (i.e., goal-seeking to achieve a goal); social ability (i.e., interacting and communicating with each other); and adaptivity (i.e., learning and adapting behaviour based on previous experience)13,15. In addition to the cross-sectional characterization that is also afforded by SD, the granular resolution of ABM allows for rules, interventions or longitudinal output to depend on an agent’s history. This approach can further capture static and dynamic network structures and spatial representation, as well as hierarchical/nested structures of networks.

**4.3 Methods**: **A comparison of decision trees, agent-based models and system dynamics**

Three modelling approaches were compared in our illustrative example. The example was purposefully simple such that it is easily comprehensible to a wider audience of health economists and epidemiologists. A previously-constructed decision tree that evaluated the cost-effectiveness of influenza vaccination (i.e., intranasal live attenuated influenza vaccine (LAIV) vs. injectable trivalent inactivated influenza vaccine (TIV)) amongst children over a year was reconstructed into both a SD and ABM approach under a Ministry of Health perspective16.

**4.3.1 Disease Progression**

At the start of the model, the whole population is at risk of influenza. Over time, two broad outcomes are possible: i) influenza, further separated into four mutually exclusive states: uncomplicated or one of three potential complications (i.e., acute otitis media and/or lower respiratory infections); or ii) no influenza. Children have an increased mortality risk during influenza. Following recovery, children are assumed to be protected from re-infection over the course of the one-year flu season.

In the original decision tree, it was assumed that all children would be immunized at the start of the model and that children may experience vaccine-related adverse events. This assumption was also applied to the calibrated ABM and SD models.

**4.3.2 Models Structure**

Figure 2 (A, B, C) presents the structure of each model. As two vaccines were being compared, the structure and description below applies equally to both arms of the model.

*4.3.2.1 Static model (i.e., decision tree)*

The decision tree assumed a known infection prevalence16 (Figure 2A). Although the original model partitioned the population into three age groups to calculate the weighted cost and effects of the intervention across each group, the calibration model only focused on one age subgroup (i.e., children aged 2 to 5 years).

*4.3.2.2 Dynamic models (i.e., system dynamics and agent-based models)*

SD and ABM differ from decision trees in that the mechanism of infection transmission or, synonymously referred to as infection dynamics, is explicitly described through people’s movement between stocks or health states (Figure 2B, C). Structurally, the epidemiology of disease spread can be described by progression through: susceptibility to influenza, infection, and recovery from or succumb to the disease. A vaccination state further exist. Quantified by a nonlinear system of ordinary differential equations17, the number of patients in each compartment will vary with time according to the state of the system. For further details on the mathematical equations related to infectious disease modelling, we recommend readers to the article by Pradas-Velasco17.

*Properties specific to System Dynamics*

In SD, the population was separated into six possible stocks according to infection and/or vaccination status: 1) Unvaccinated Susceptible; 2) Vaccinated Susceptible; 3) Infected (subscripted into four possible complication states); 4) Alive with Influenza; 5) Death by Influenza; and 6) Recovered (Figure 2B). In terms of implementation, levels in each stock were controlled by the rates of inflows and outflows. Therefore, change over time within each stock equals the difference in total inflow and total outflow over that period.

*Properties specific to Agent-based Model*

In ABM, statecharts describe the connections between health states for each individual (Figure 2C). In this case, individuals were defined by two dimensions: their infection and vaccination status. Transitions in ABM are different from flows in SD. Flows in SD are generally memoryless (i.e., chance a person leaves a stock is independent of the amount of time they have spent in that stock); transitions in ABM are flexible and unrestricted. In AnyLogic 7 software, transitions can be triggered at a certain rate, after a certain time within a state, by a condition being realized or by a message (Table 2).

ABM further differs from SD in that it can explicitly define networks or social structures in which agents exist within their environment. This connectivity is referred to as network topology and is described later.

**4.3.3 Model Parameters**

The dynamic models used same parameter values as those published from the original decision tree (Appendix I)16, unless otherwise specified (Table 3). Probabilities were assigned to each branch within the decision trees; rates, conditions or timeouts to each statechart transition in the ABM; or transition hazards to each flow in the SD. Costs were assigned to the terminal nodes of the decision tree, to the specific transition or health states in ABM or to the stock and flows in SD. Given the model’s perspective, only medical costs associated with vaccination (i.e., acquisition, administration) and influenza (i.e., outpatient, emergency visits, hospitalization, drugs) were included and are expressed as 2010 dollars to remain comparable to the previously published decision tree.

As noted, dynamic models differ from static models in how disease transmission is handled. In dynamic models, disease transmission is calculated endogenously based on a function of key epidemiological parameters: the contact rate, the duration of infection and the probability of transmission upon contact. Infection spread captures both the direct effects of vaccine protection and its indirect effects through herd immunity, with the latter not explicitly captured by decision trees. To model transmission, we assumed 5 contacts per day between individuals. The secondary attack rate for vaccinated children, a parameter necessary in dynamic models, was calibrated based on the known prevalence of influenza in the original decision tree.

**4.3.4 Statistical Analysis**

Both dynamic models were constructed in AnyLogic 7 while the decision tree was built in Microsoft® Excel. The primary outcome measure, the incremental cost-effectiveness ratio (ICER), was defined as the cost per flu case averted and was calculated according to conventional decision rules.

Compared to aggregate-level models, individual-level models permitted the characterization of first-order uncertainty. In the context of stochasticity, 200 independent runs were required to generate stable results for the ABM. In contrast, aggregate-level models can easily account for second-order uncertainty. Monte Carlo simulations (n=1000) were conducted employing identical distributions for both aggregate-level modelling approaches. Similar to most individual-level models, parameter uncertainty remained unexplored in ABM due to computing power constraints. Given that it took approximately 10 hours to run a single simulation, simple extrapolation would suggest that the time needed to conduct second-order uncertainty on a single machine would be roughly 416 days (i.e., 10hours∙1000simulation/24 hours/day).

Given the one year model duration, no costs and effects were discounted.

**4.3.5 Model Validation and Calibration**

To ensure comparability across modelling approaches, a validation exercise verified that all three models employed the same parameters and assumptions. Several simplifying assumptions were necessary to calibrate the ABM and SD model to replicate the results of the reference decision tree. First, the population size was assumed stable with the exception of losses due to influenza-related mortality. This was justifiable given the model’s short time horizon and, in fact, is a common assumption in influenza models17. Secondly, to reflect the original decision tree, a 100% vaccination rate was assumed. Lastly, parameters relating to disease transmission for the dynamic models were calibrated based on the prevalence of influenza assumed in the original decision tree. In a static model, prevalence is assumed stable as the force of infection is not explicitly modelled. For the dynamic models to be comparable to the static model, the secondary attack rate was adjusted to match against the decision tree’s flu prevalence. This is a typical and necessary practice18. Results of all three approaches were therefore expected to be comparable following calibration.

As per good modelling practice, both dynamic models underwent a process of peer-review (N.O, R.O).

**4.3.6 Simulations**

Following model verification, controlled simulations were designed. The first group of analyses, run on both dynamics models, varied the assumptions surrounding the vaccination rates. The second set, focused on different network topologies, were conducted only on ABM. The last simulation incorporated heterogeneity by expanding to model children, ages 2 to 17 (Table4).

*4.3.6.1 Scenario I: Varying vaccination rate (SD and ABM)*

The decision tree is limited in that, although it is possible to explore different rates of vaccination, the infection prevalence must be a known input. In dynamic models, however, prevalence is an output emerging from explicitly modelling the disease transmission. Such models can therefore capture changes to disease transmission due to changes in the contact structure or in disease susceptibility.

For instance, the original model assumed that all children received immunization prior to the model start. However, in reality, not all children will be vaccinated nor will all vaccinations occur at the same time. We varied both aspects in the two scenarios that were explored: (i) immediate vaccination of a proportion of children prior to the model start; and (ii) a vaccination schedule whereby a proportion of children are vaccinated. We assessed the cost-effectiveness of LAIV under existing vaccination penetration rates (16%)19 and a hypothetical campaign that doubled coverage rates. The secondary attack rate for unimmunized children was taken from the literature20.

*4.3.6.2 Scenario II: Network Topologies (ABM only)*

As previously mentioned, spread of infection is explicitly modelled as an output in dynamic models. However, the mechanism of infection spread between SD and ABM relies on different assumptions.

System dynamics, as an aggregate-level model, assumes idealized random-mixing amongst individuals (i.e., each individual within the same compartment has an equal chance per unit time to meet the others)21. Although this requires fewer inputs to model contact between individuals, it may also overlook important patterns and behaviour of disease diffusion. For instance, it offers limited consideration on the impact of persistent connections between those in the population that are common as a result of family, workplace and/or geographical constraints.

By contrast, ABM incorporates the contact structure according to the transmission or social networks. Networks may play an important role in the spread of infectious diseases as, in reality, transmission is triggered by person-to-person interactions. Network topology refers to the network structures forming the environment in which Individuals live or interact. By focusing at the individual-level, the interactions within a network can be more accurately captured, which may be critical in modelling infection spread.

To ensure comparability between the SD and ABM, a random network was selected in the calibration exercises for the ABM, with connections equal to the population size (analogous to caveman network). We revised the network structure to explore the impact of three of the most common types of idealized network topologies: random, small world and scale-free (

Figure3). To ensure comparability, the average number of contacts in the random and small-world network was kept the same. For simplicity, we only assumed static networks (the links or relationships between individuals do not change over time). Below we discuss each of these types of networks22.

*Poisson random network*

Each individual in the network is randomly connected to any other person with a specified probability, irrespective of their spatial position or any other attributes. In other words, any pair of individuals in the network exhibit the same uniform probability of being connected. An assumption was made to limit the average number of connections per individual to 100 agents and different individuals will vary in their number of connections, as defined by a binomial distribution.

*Small world network*

Each agent is connected with the same number of agents. The majority are nearby (local connections) but may also include a small number of distant ones (global connection). In this context, two connected individuals near each other are likely to share similar connections due to their locality. This typically leads to highly clustered connections whereby an infection spreads locally. However, distant links can further capture the transmission phenomenon of disease spread outside constrained regions23.

*Scale free network*

Both random and small world networks display little variation in the number of connections per agent. However, in many observed networks, the number of connections per individuals is not homogeneous. In a scale-free network, the distribution on the number of connection per agent follows a power law (i.e., fraction of agents with k connections is defined by k-γ with γ being the network parameter)22. Some individuals are “hubs” with lots of connections, whereas others are “hermits” with fewer connections. Highly connected agents are disproportionately important in disease transmission as they have a higher risk of becoming infected and transmit the infection more rapidly once infected. Research so far suggests that such networks will allow for an infection to remain endemic within subgroups of the broader population even if the average rate of contact would be insufficient for the infectious disease to be sustained22.

*4.3.6.3 Scenario III: Heterogeneity (decision tree and ABM)*

As decision trees and SD are both aggregate-level models, people within the same group (i.e., health states in the decision tree, stocks in SD) are assumed to be homogeneous. However, heterogeneity may be important. For instance, in the case of this model, age-specific costs and transition probabilities exist (Appendix II).

One common approach in decision trees to handle heterogeneity in age groups is to run the analysis for each age group individually, assuming no interaction between the three age cohorts (i.e., children age 2-5, age 6-9 and age 10-17). Indeed, in the original study, the expected cost and outcomes for each age cohort were weighted by the proportion of children in that particular age group16.

In the case of SD, heterogeneity could be introduced via the creation of new stocks or by (equivalently) subscripting the pre-existing stocks along the dimensions of the particular attribute of interest. However, creation of new stocks has it challenges as stocks have to follow the heterogeneity dimensions in order to be able to distinguish it against others with different attributes. Furthermore, it risks leading to an unmanageable number of stocks as more attributes are considered, which is analogous to state explosion in traditional cohort models. Subscripting, a functionally convenient way to work with duplicated stocks, was employed in this model to distinguish patients across the four possible flu severities. Subscription similarly suffers from a “curse of dimensionality”. As more attributes are consider, the number of stocks exponentiates. To represent the interlinked aspects of heterogeneity, the definition of stocks and flows across a model must be redefined, and often requires a large number of complex and inter-linked differential equations that must be balanced across flows.

By contrast, ABM, incorporates heterogeneity much more simply and is flexible in permitting interactions between children across all age cohorts. To model the differences amongst the three distinct age groups (between the age of 2 to 17), only a single additional parameter was required to be introduced in ABM. Heterogeneity and memory could easily be incorporated, as transitions were defined according to an individual’s attributes or previous states.

**4.4. Results**

**4.1 Validation and calibration (all modelling approaches)**

The validated and calibrated dynamic model produced similar results to the decision tree (Table 5). The higher vaccine costs for LAIV were offset by cost-savings from treating lower numbers of influenza cases compared with TIV.

Larger differences were however observed between the aggregate (i.e., decision trees and SD) and individual-level (ABM) model, which can be explained by stochastic factors inherent to ABM. As ABM simulates individuals – who are subjected to significant stochastic variability - its estimates are based on multiple runs in order to generate stable results. Both aggregate-level models were instead deterministic and generated a single estimate. The computation burden thus differed between these modelling approaches as the simulation could be measured in seconds, minutes and hours for decision trees, system dynamics and agent-based models, respectively.

Parameter uncertainty within the aggregate-level models was characterized (Table 6). As one of the primary differences between static and dynamic models is the inclusion of parameters that directly impact disease transmission (i.e., number of contacts, secondary attack rate and disease duration), these parameters had to be held constant within the dynamic model to ensure model comparability. Under these circumstances, disease transmission remains constant across simulations and only the parameter uncertainty relating to costs is captured. The probabilistic mean estimates were comparable between the modelling approaches (Table 6A) and, likewise, the cost-effectiveness acceptability curve for each modelling approaches were alike (Figure 4A).

If the probabilistic sensitivity analysis was not constrained (in this case, disease duration was also defined by a parameter distribution), the SD model would be able to capture the impact of parameter uncertainty arising from both direct and secondary, indirect effects. Contrary, given the nature of static models, uncertainty in time-related parameters would not capture the potential impacts from indirect effects. To understand this, time has no relationship to the proportion of infected children in the decision tree whereas, in the system dynamics, longer duration of influenza would translate to a longer period of infectiousness and thus, a greater proportion of infected children. Consequently, the proportion of infected children estimated by the complete probabilistic decision tree remained similar whereas, in comparing the complete probabilistic SD model, results differed considerably from the calibrated model (LAIV: 0.172 vs. 0.047; TIV: 0.196 vs 0.104 in the full probabilistic and deterministic estimates, respectively). The wider standard deviation observed in the dynamic SD model compared to the static one further indicates greater parameter uncertainty when incorporating both direct and indirect effects. Despite a similar conclusion when observing the probabilistic estimates, the CEAC curve demonstrates greater parameter uncertainty in the SD model given that both direct and indirect effects are captured (Figure 4B).

**4.4.2 Scenario I: Varied Vaccination Rate and Schedule (SD and ABM)**

A more realistic scenario in which only a proportion of children were vaccinated was explored under both dynamic models.

Compared to a complete (100%) vaccination strategy, the proportion of children infected was estimated to be higher when rates of vaccination were reduced (Table 7). The higher prevalence of influenza also increased the expected costs associated with each treatment strategy. The estimates generated by SD were often higher than ABM and, in fact, the conclusions drawn from the ICER differed between the two models (Table 7). ABM always came to the same conclusion that LAIV would be the dominant strategy, irrespective of the proportion of children vaccinated (i.e., 16%; 32%) and the administration schedule (i.e., immediate; rolling). SD, instead, always suggested that LAIV would be both more costly and more effective than TIV. It was further found that SD was less sensitive to the schedule of vaccination, as results were comparable between scenarios of immediate vaccination and rolling vaccination (Table 7A, B).

Changes to the proportion vaccinated and the vaccination schedule was found to change the number of expected influenza cases. For instance, doubling of the current vaccination target would reduce the number of influenza cases over a year, reflecting the impact of herd immunity (Table 7C). From an economic perspective, the higher upfront costs associated with LAIV vaccines were offset due to cost savings from averted flu cases. The overall incremental cost difference between the two vaccines were larger as rates of vaccination increased (Table 7A, C) and, overall, the ICER for LAIV was lower and more attractive when vaccination rates doubled (Table 7C).

**4.3 Scenario II: Network topologies (ABM only)**

As noted, SD assumes that individuals within a compartment are perfectly mixed (i.e., that transmission is equally likely over every individual in that compartment). ABM, on the other hand, relies on specific person-to-person interactions, defined according to their social network, to propagate the transmission dynamics and, thus, the network topology (e.g., small world, random, scale-free) can impact the mechanism of disease spread.

The random network model yielded the most similar results to those of the calibration exercises (Table 8A). This was not surprising: random-mixing is in fact a special case of a random network whereby everyone is connected to every individual within the model.

In the small-world network, LAIV was found to dominate TIV but with a smaller cost savings (Table 8B). As noted, the small-world network incorporates greater clustering amongst local connections. Thus, despite equal numbers of connections per agent in the small world and random networks, less randomness exists in the former. Clustering, inherent in this network, was found to lead to lower estimates for the prevalence of influenza under both vaccination strategies as disease spread was more localized.

The scale-free network led to the most distinctive results (Table 8C). Prevalence rates for influenza were almost five-fold less than those estimated in the calibration models. To recall, this network structure reflects social behaviours with different individuals presenting different number of connections. Due to stochasticity in the individual-level model, if the index agent (i.e., “patient zero”) is one with few connections, the likelihood of infection spread is lower than if the index agent is one with many connections. As such, some simulations will result in high numbers of infected agents whereas others may generate little flu transmission. These observed clinical outcomes consequently drove the cost-effectiveness estimates. For instance, when transmission was driven by ‘hermits’ with few connections, LAIV became less attractive due to the lower likelihood of influenza spreading to the rest of the population.

Different assumptions regarding the social network structure of an ABM was found to lead to different ICER estimates: LAIV was a dominated strategy under the random and small world networks but, under the scale-free network, the ICER was $1.755/flu case avoided (Table 8).

**4.4.4 Scenario IV: Patient Heterogeneity (decision tree and ABM)**

We further investigated the sensitivity of our model to heterogeneity, specifically with regards to age. When the model was expanded to include all children, aged two to seventeen, the proportion of infected children estimated by the two modelling approaches varied considerably. In the decision tree, each age group was modelled individually, with the estimates then weighted under the assumption that cohorts do not interact with each other and are, thereby, independent. In the ABM, interactions between children across all age-group can be permitted (i.e., free-mixing whereby there exist non-independence in the age cohorts). Thus, the decision tree and the agent-based model that assumed no interactions between age cohorts led to comparable results (Table 9A, B). However, when age cohorts were permitted to interact with each other, the prevalence of infection in the LAIV strategy reduced by approximately two fold whereas the prevalence of infection in the TIV strategy remained comparable. Thus, the ICER shifted, with LAIV becoming the dominant strategy (Table 9C).

**4.5. Discussion**

When selecting amongst the many modelling approaches that are available, common questions often arise: which approach is most efficient? Which best represents reality? To what degree does added granularity from individual-level models provide practical benefits? By understanding the situations whereby these approaches produce either concordant or discordant results compared to each other, this can provide knowledge on when each approach should be used.

In this paper, we conducted controlled simulations to compare the performance and results of a decision tree, a system dynamics and an agent-based model in the context of influenza vaccination to better understand the trade-offs specific to each modeling approach. Although the scenarios may be considered highly stylized, it helped illustrate the different assumptions associated with each approach and their capabilities in capturing different features of the decision problem.

In the decision tree, prevalence of influenza is a known model input in which clinical outcomes and cost-effectiveness are then calculated. As a result, probabilistic sensitivity analysis does not explicitly capture parameter uncertainty arising from secondary effects due to disease transmission, leading to a higher probability that LAIV was a cost-effective strategy across most willingness-to-pay thresholds (Figure 3). By contrast, both SD and ABM model, disease transmission is a function of several parameters including infectiousness of the disease, population contact structure and the nature of vaccination. Consequently, probabilistic sensitivity analysis within a dynamic model captures not only parameter uncertainty in the direct effects but also uncertainty relating to secondary, indirect effects18,24. Indeed, static models have been previously described as a specific case of dynamic modelling whereby the probability of disease exposure remains constant over time (e.g., when an infectious agent has a long latency period).

As disease transmission is modelled in a dynamic model, it is possible to characterize the consequences of vaccination in terms of both: (i) direct effects from vaccinating individuals, and (ii) indirect effects on the unvaccinated population due to herd immunity effects. Dynamic models can therefore flexibly model disease transmission and would be insightful to address public policy questions regarding vaccination campaigns aimed to target certain rates of immunization (Scenario I, Table 6).

Although both SD and ABM can offer insight into the mechanism of infection dynamics, their underlying assumptions do differ. As an aggregate-level model, system dynamic models aggregate individuals into groups (i.e., stocks) that share similar characteristics. Due to this, SD is unable to model specific interactions or specific social constructs through which an infection is spreading. Transmission is calculated based on the mean rates of exposure with individuals that belong within the same stock, which is assumed perfectly well-mixed25. This simplifying assumption may reduce the accuracy of estimates of the prevalence of infection. Although this assumption can be relaxed through stratification and incorporating preferential mixing between distinct groups according to mixing matrices, its representation and implementation can be cumbersome, complicated and lacks the capacity to examine or take advantage of particular characteristics of network structure in the intervention scenarios. Furthermore, although it provides a solution to random mixing across an entire population, the assumption of random mixing still applies with a given category.

In contrast, ABM offers greater flexibility. It can also account for the contact structure and pathways of infection spread across the associated social networks since transmission is modelled based on person-to-person interactions rather than an idealized assumption of perfectly random-mixing. ABM opens the opportunity to examine intervention strategies that explicitly make use of information on, or even the shape, of network position (e.g., target strategies on people with central positions in the network for enhanced monitoring, behavior change, or prophylaxis). We investigated three types of network topologies and found that differences in network structure can alter the overall infection spread (Scenario II, Table 7). However, modelling a network demands considerable details (e.g., structure of network, contact/behavioural patterns) that can be conveniently omitted in aggregate-level models. Indeed, for simplicity, the networks studied here were assumed static as links between individuals did not change over time (i.e., no lost or new connections). However, current research suggests that dynamic networks may be of importance in modelling the spread of some pathogens26 and readers interested in modelling dynamic networks are recommended the following papers22,26.

From our experience, ABM can also easily capture additional components of heterogeneity more effectively and accurately. To model flu transmission beyond the ages of 2 to 5, only one additional variable was introduced, representing an individual’s age in order to incorporate age-specific costs and transition probabilities (Scenario III, Table 8). Statecharts have a capacity to easily factor dimensions that characterize an agent which can often significantly reduce the complexity and enhance transparency of a model. This is in contrast to the combinatorial explosion that is characteristic amongst aggregate-level models. Representing heterogeneity within SD is awkward, inflexible and scales poorly as the count of heterogeneity dimensions rise27. Furthermore, whereas the decision tree relied on a restrictive assumption of no interaction between age cohorts, this was easily permitted in ABM and was found to have a significant impact on the modelled results.

So far, we have discussed the technical features specific to each modelling approach. However, practical considerations may also exist when selecting a modelling approach. Firstly, few health economists have been trained on dynamic modelling as the focus thus far has been on static models, such as decision trees and Markov models. Certainly ABM and SD may demand a higher level of training and greater effort to pursue it and thus, a greater time investment may be required when constructing these models. In terms of simulation time and computer memory demands, this is not dependent on population size in aggregate-level models; whereas, in individual-based models, this grows at least linearly with size of the population modelled4,21 if absent of any sophisticated computing techniques. Thus, if there is a need to simulate large populations, the performance of ABM would be of considerable concern, and begins to require recourse to distributed computing. Furthermore, greater flexibility does come at a cost in terms of computation demands (e.g., time/speed to run analyses, data requirements). This is especially noticeable in the conduct of sensitivity analysis, as individual-level models are far less nimble than aggregate-level models28. The time taken to conduct probabilistic sensitivity analysis in SD and decision tree for our case could be measured in minutes or seconds, respectively. However, a single analysis of our ABM requires approximately 10 hours to complete due to the need to run Monte Carlo simulations to account for stochastics. As such, probabilistic sensitivity analysis for ABM was not explored in this paper given its infeasibility. One option being proposed in individual-level models to support more efficient conduct of probabilistic sensitivity analysis is Latin hypercube sampling. Latin hypercube sampling is a stratified sampling technique that randomly generates a plausible sample collection for multiple parameters that is guaranteed to cover the entire sample space of each parameter29 and it may be worthwhile to explore this variation reduction technique in the context of conducting probabilistic sensitivity analysis for ABMs.

Despite the early infancy of applying alternative modelling approaches in cost-effectiveness analysis, it is important to gain a better understanding of the circumstances in which to use each approach given their differences in computational demand and model flexibility. The choice of a modelling approach will remain an important step in the process of model development. Although few would disagree that these alternative approaches may better capture certain complexities, a remaining concern is whether the extra complexity is indeed required. Furthermore, with the current growing interest on hybrid system science techniques and the forays into mixing systems science with decision-analytic approaches, the choice of modelling approaches extends to how to mix such techniques effectively. This paper has shown how each of these three different modelling approaches may be similar and different in the context of infectious disease modelling. This work is particularly timely given recent recommendations stating the need to construct dynamic models in the area of infectious disease modelling to support policy-making in many agencies, such as the National Institute of Clinical Excellence, the Canadian National Advisory Committee on Immunization alongside joint good research guidelines from the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) and the Society of Medical Decision Making (SMDM)30.

We acknowledge that the scenarios provided in this paper are hypothetical. They may not be realistic and are intended to provide an illustrative example to better understand the trade-offs in the advantages and disadvantages of different modelling approaches. However, this study shows that the estimates on expected costs and effects may differ according to the choice of the modelling approach selected and, more importantly, according to the assumptions made in relation to that particular modelling approach. The calibrated models all reached the same conclusion: LAIV vaccination would be dominant in children between the ages of 2 to 5 years whom are all vaccinated before the start of the model and whom follow perfectly randomly mixing. However, different approaches can support easier relaxation of these assumptions, which may lead to considerably different conclusions. Based on this paper, we remark on the following trade-offs: 1) ABM was found to be the most flexible by reflecting the disease course naturally, capturing first-order uncertainty and patient heterogeneity, incorporating agent behaviours within their network structure but at the expense of a longer simulation time; 2) SD, an aggregate-level model, was more restricted as, although it retained the epidemiological structure of disease transmission, it relied on the simplifying assumption of random-mixing between individuals and homogeneity within each stock and; 3) decision trees were the most restrictive amongst these approaches as both the proportion vaccinated and the natural disease transmission were fixed inputs in the model but required the shortest simulation time. Future research should now focus on applying alternative modelling approaches to other areas of infectious disease prediction.

**References**

1. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy* 2004; 9(2), 110-118.

2. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics* 2000; 17(5), 461-477.

3. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008; 26(2), 131-148.

4. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manag Sci* 1998; 1(2), 133-140.

5. Armstrong GL, Billah K, Rein DB, Hicks KA, Wirth KE, Bell BP. The economics of routine childhood hepatitis A immunization in the United States: the impact of herd immunity. *Pediatrics* 2007; 119(1), e22-e29.

6. Kim S-Y, Goldie SJ, Salomon JA. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. *Med Decis Making* 2010; 30(5), E1-E28.

7. Trotter CL, Edmunds WJ. Reassessing the cost-effectiveness of meningococcal serogroup C conjugate (MCC) vaccines using a transmission dynamic model. *Med Decis Making* 2006; 26(1), 38-47.

8. Meadows D, Robinson J. The Electronic Oracle. Computer Models and Social Decisions. John Wiley & Sons, Chichester, UK; 1985.

9. Cooper K, Davies R, Roderick P, Chase D, Raftery J. The development of a simulation model of the treatment of coronary heart disease. *Health Care Manag Sci* 2002; 5(4), 259-267.

10. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview a report of the ISPOR-SMDM modeling good research practices task force–1. *Med Decis Making* 2012; 32(5), 667-677

11. Lich KH, Ginexi EM, Osgood ND et al. A call to address complexity in prevention science research. *Prev Sci* 2013; 14(3), 279-289.

12. Sterman JD. Business Dynamics: System Thinking and Modeling for a Complex World. Irwin/McGraw-Hill, Boston,USA; 2000.

13. North MJ, Macal CM. Managing business complexity: discovering strategic solutions with agent-based modeling and simulation. Oxford University Press, New York, USA; 2007.

14. Chhatwal J, He T. Economic Evaluations with Agent-Based Modelling: An Introduction. *Pharmacoeconomics* 2015, 1-11. [Epub ahead of print]

15. Wooldridge M. An introduction to multiagent systems. John Wiley & Sons, Chichester, UK; 2002.

16. Tarride JE, Burke N, Von Keyserlingk C, O'Reilly D, Xie F, Goeree R. Cost-effectiveness analysis of intranasal live attenuated vaccine (LAIV) versus injectable inactivated influenza vaccine (TIV) for Canadian children and adolescents. *Clinicoecon Outcomes Res* 2012; 4, 287-298.

17. Pradas-Velasco R, Antoñanzas-Villar F, Martinez-Zárate M.P. Dynamic Modelling of Infectious Diseases. *Pharmacoeconomics* 2008; 26(1), 45-56.

18. Welte R, Postma M, Leidl R, Kretzschmar M. Costs and effects of Chlamydial screening: dynamics versus static modelling. *Sex Transm Dis* 2005; 32(8), 474-483.

19. Johansen H, Sambell C, Zhao W. Flu shots--national and provincial/territorial trends. Sta*tistics Canada, Canadian Centre for Health Information* 2006; 17(2), 43-48.

20. Smith A, Coles S, Johnson S, Saldana L, lhekweazu C, O'Moore É. An outbreak of influenza A (H1N1) in a boarding school in South East England, May-June 2009. *Euro Surveill* 2009; 14(27), 2335-2346.

21. Tian Y, Osgood N. Comparison between individual-based and aggregate models in the context of tuberculosis transmission. Washington DC, USA: Proceeding at *System Dynamics Winter Conference.* 2011.

22. Keeling MJ, Eames KT. Networks and epidemic models. *J R Soc Interface* 2005; 2(4), 295-307.

23. Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. *Nature* 1998; 393(6684), 440-442.

24. Lugner AK, Mylius SD, Wallinga J. Dynamic versus static models in cost-effectiveness analyses of anti-viral drug therapy to mitigate an influenza pandemic. *Health Econ* 2010; 19(5), 518-531.

25. Rahmandad H, Sterman J. Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Manage Sci* 2008; 54(5), 998-1014.

26. Morris M, Kretzschmar M. Concurrent partnerships and transmission dynamics in networks. *Social Netw* 1995; 17(3), 299-318.

27. Osgood N. Representing heterogeneity in complex feedback system modeling: Computational resource and error scaling. Oxford, UK: Paper presented at: 22*nd International Conference of the System Dynamics Society*. 2004.

28. Kuehne FC, Chancellor J, Mollon P, Weinstein M, Powderly W. Microsimulation or cohort modelling? A comparative case study in HIV infection. Copenhagen, Denmark: Paper presented at: iHEA 2007 6th World Congress: Explorations in Health Economics Paper. 2007.

29. Calvert SC, Taale H, Snelder M, Hoogendoorn SP. Application of advanced sampling for efficient probabilistic traffic modelling. *Transportation Research Part C: Emerging Technologies* 2014; 49, 87-102.

30. Pitman R, Fisman D, Zaric GS, Postma M, Kretzschmar M, Edmunds J, Brisson M. Dynamic transmission modeling: a Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–5. *Med Decis Making* 2012; 32(5), 712-721.

31. Luce BR, Zangwill KM, Palmer CS, Mendelman PM, Yan L, Wolff MC, Cho I, Marcy SM, Iacuzio D, Belshe RB. Cost-effectiveness analysis of an intranasal influenza vaccine for the prevention of influenza in healthy children. *Pediatrics* 2001; 108(2):e24.

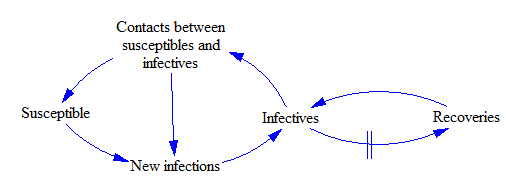
32. Belshe RB, Edwards KM, Vesikari T, Black SV, Walker RE, Hultquist M, Kemble G, Connor EM. Live attenuated versus inactivated influenza vaccine in infants and young children. *N Engl J Med* 2007; 356(7), 685-696.

33. Sebastian R, Skowronski D, Chong M, Dhaliwal J, Brownstein J. Age-related trends in the timeliness and prediction of medical visits, hospitalizations and deaths due to pneumonia and influenza, British Columbia, Canada, 1998–2004. *Vaccine* 2008; 26(10), 1397-1403.

34. Sander B, Kwong JC, Bauch CT, Maetzel A, McGeer A, Raboud JM, Krahn M. Economic appraisal of Ontario's Universal Influenza Immunization Program: a cost-utility analysis. *PLoS Med* 2010; 7(4), e1000256.

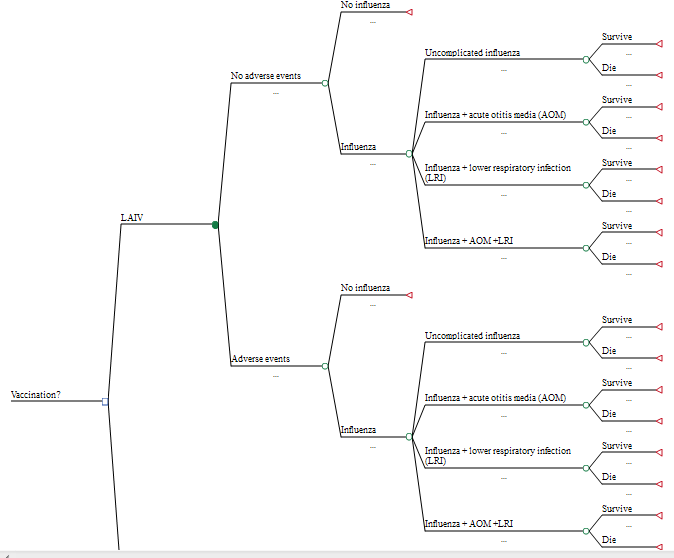
35. Loughlin J, Poulios N, Napalkov P, Wegmüller Y, Monto AS. A study of influenza and influenza-related complications among children in a large US health insurance plan database. *Pharmacoeconomics* 2003; 21(4), 273-283.

**Figure 1:** Example of a causal-loop diagram describing the transmission of a hypothetical infectious disease based on the susceptible-infectious-recovered (SIR) structure. The blue arrows are links, describing the relationship between two variables within the system. Two vertical lines on an arrow indicates a delay in the system.

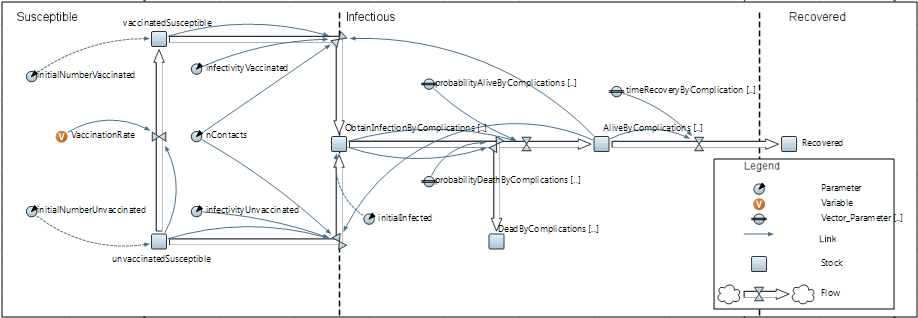


**Figure 2:** Model structure for immunization. Only one arm of the model is presented as the second arm is a mirrored version.

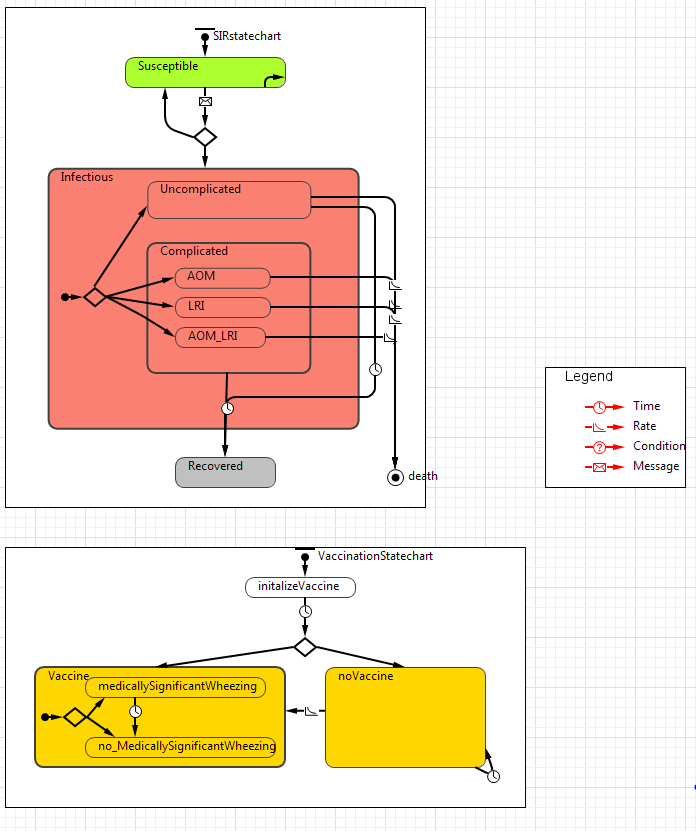
(A) Decision Tree



(B) System Dynamics Model



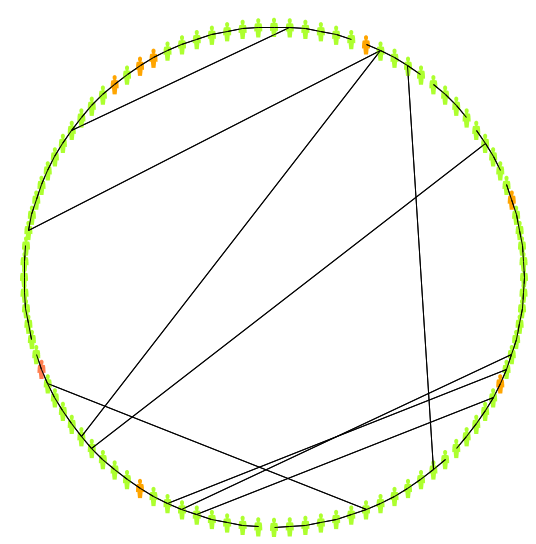
(C) Agent-based Model



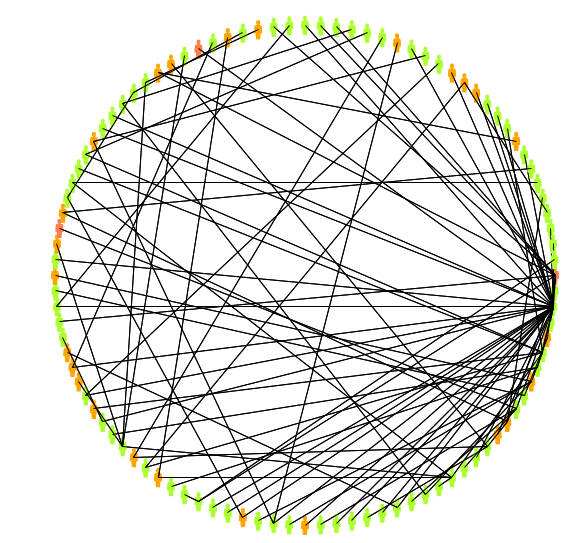
**Abbreviation**: AOM: acute otitis media; LAIV, intranasal live attenuated influenza vaccine; LRI, lower respiratory infection; MSW, medically significant wheezing; TIV, injectable trivalent inactivated influenza vaccine

**Figure 3:** Diagrammatic representation of three common network topologies in ABM. Agents are people, laid out around a circle. A connection between a pair of agents is represented by a line linking those agents. (A) Poisson random network is based on random connections between agents, in this case, each agent has on average two random connections; (B) Small world assumes exactly two connections per agent with the effect of clustering observable (agents are more likely to be connected to those closer to them than to those further away; (C) Scale free has greater variability amongst agents in terms of the number of connections they each have. Rather, some agents are considered hubs with many connections, while a larger proportion are hermits with fewer connections.

(A) Random (B) Small World

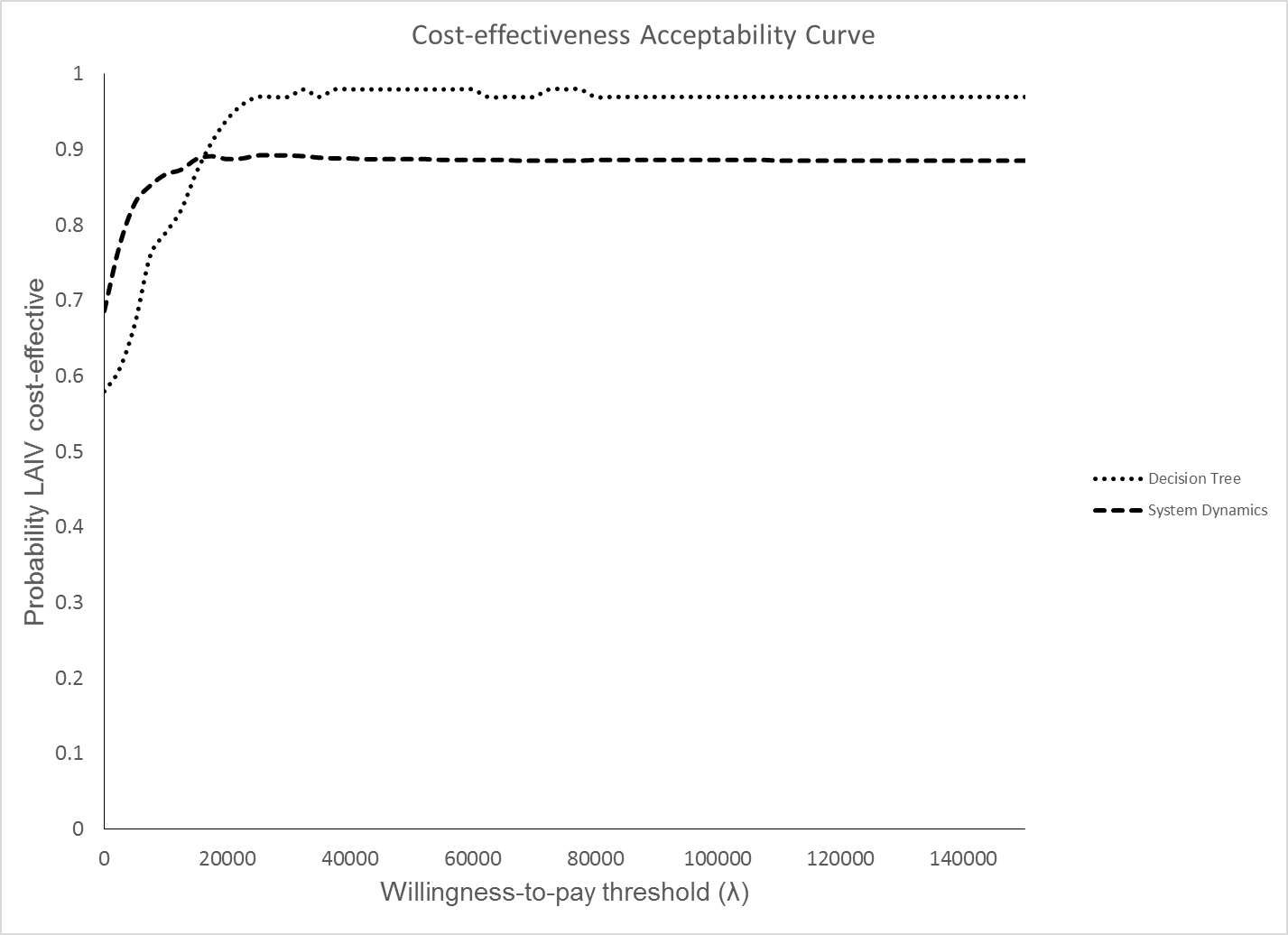
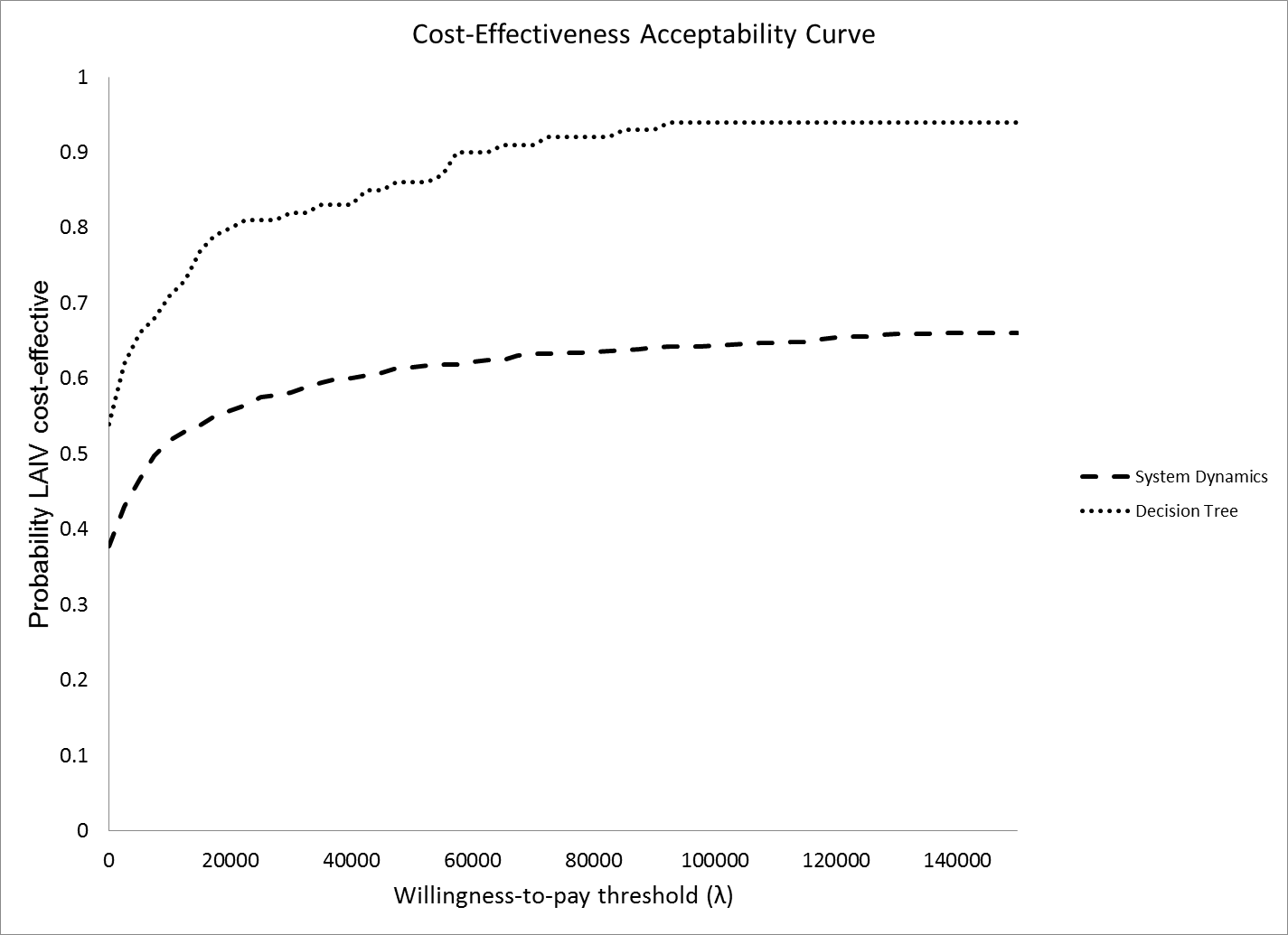
 

(C) Scale free



**Figure 4:** Cost-effectiveness acceptability curve from the two aggregate-level models: decision tree (dotted line); system dynamics model (dash line) (A) The constrained probabilistic model only captures parameter uncertainty arising from direct effects. (B) The complete probabilistic model allows all parameters to be defined by a distribution, with the dynamic model capturing the impact of parameter uncertainty arising from direct and indirect effects, whereas the static model only captures direct effects.

1. Constrained probabilistic model (B) Complete probabilistic model

**Table 1:** Comparison of the main characteristics of agent-based and system dynamics model

|  |  |  |
| --- | --- | --- |
|  | **System Dynamics** | **Agent-based model** |
| Focus | System-oriented | Individual-oriented |
| Population resolution | Typically at an aggregate-level whereby entities in the same stock are assumed homogeneous (e.g., working with averages). Instances may exist where it is applied as an individual-level approach. | Individual level whereby autonomous agents can make decisions and interact with others and its environment. Heterogeneity in agent’s behavior and attributes can be easily captured. |
| Scalability | Good scaling to a population-level although poor for heterogeneity | Poor scaling to a population-level although good for heterogeneity |
| Interactivity | Dynamic - driver based on concept of the ‘feedback loop’ | Dynamic – driver is ‘agent’s decisions and interactions’ |
| Local interactions | Random-mixing between individuals (equal probability of interaction amongst individuals in the same stock) | Complex Interaction at a social level. |
| Resource constraints | Yes | Yes |
| Dimension of time | Typically continuous, can be made discrete | Discrete or continuous |
| Mode of behaviour characterization | Stock and flow (ordinary differential equations) | Various (state charts, events, update rules) |
| Deterministic vs. stochastic | Deterministic | Stochastic |

**Table 2:** Description of common transitions in agent-based modelling. An example relating to the context of this model is provided.

|  |  |  |
| --- | --- | --- |
| **Transition** | **Example** | **Analogous to:** |
| Rate | Upon vaccination, children have a hazard rate of 0.021 of developing a vaccination-related adverse events. | Transition Probabilities in Markov cohort models |
| Timeouts | Upon contracting uncomplicated influenza, a child will recover after 2.2 days. | Time-to-event in discrete event simulation |
| Condition | Upon vaccination, children less than 10 years of age have a 2.1% probability for developing an adverse event whereas children older than the age of 10 have 0 probability of developing an adverse event. | Conditional probabilities in Markov microsimulation (often used to incorporate memory or heterogeneity) |
| Message | Upon coming in contact with an infectious individual, a message is sent to the original agent that they have been exposed to someone with the influenza virus. | None. Message transitions are a mean of asynchronous communication between components of the model (such that one piece of the mechanism can depend on an action occurring in another part of the model) |

**Table 3:** Model parameters that differ by modelling approach

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter Name | Static model value | | Dynamic model value | |
| Mean | Standard deviation (distribution) | Mean | Standard deviation (distribution) |
| Probability Infected (LAIV) | 0.0489 |  | Model Output | |
| Probability Infected (TIV) | 0.1046 |  | Model Output | |
| Number of contacts per Day | Not modelled | | 5 | 2 (gamma) |
| Secondary attack rate (LAIV) | Not modelled | | 0.033 | N/A |
| Secondary attack rate (TIV) | Not modelled | | 0.034 | N/A |
| Secondary attack rate (unvaccinated) | Not modelled | | 0.054 | N/A |

**Table 4:** Summary of scenarios studied

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Scenario | Modelling Approaches Compared | | | Assumptions | | |
| Decision Tree | System Dynamics | Agent-based Model | Vaccination | Network Structure | Age Range modelled (years) |
| Calibration | X | X | X | 100% | Random | 2-5 |
| I. Varying vaccination rate |  | X | X | 16%, 32% | Random | 2-5 |
| II. Network topology |  |  | X | 100% | Random  Small-world  Scale-free | 2-5 |
| III. Heterogeneity | X |  | X | 100% | Random | 2-17 |

**Table 5:** Expected costs, effects and ICER from all three modelling approaches following calibration. Assumptions include 100% vaccination prior to start of model with random-mixing amongst the children (ages 2 to 5 years).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Decision Tree** | | **System-dynamic model** | | **Agent-based model**  **(st dev)** | |
| **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** |
| **Number of flu cases (per 1000)** | 49 | 105 | 49 | 104 | 46  (154) | 119 (291) |
| **Expected Costs ($)** | 44.58 | 47.18 | 44.51 | 47.11 | 44.08  (27.94) | 49.63 (52.70) |
| **Number of flu cases averted (per 1000)** | 56 |  | 55 |  | 73 |  |
| **Incremental costs** | -2.61 | -2.6 | -5.55 |
| **ICER ($/flu cases averted)** | **Dominant** | **Dominant** | **Dominant** |

**Table 6:** Probabilistic estimates from the cohort-level models for children vaccinated between the ages of 2 to 5. (A) The constrained probabilistic model only captures parameter uncertainty arising from direct effects. Any model parameters that have an impact on flu transmission are held constant. (B) The complete probabilistic model allows all parameters to be defined by a distribution. Specifically, under this scenario, time-related parameters (e.g., duration of influenza) was assumed to be defined by a parameter distribution. The dynamic model therefore captures the impact of parameter uncertainty arising from direct and indirect effects.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1. **Constrained probabilistic model** | | | | 1. **Complete probabilistic model** | | | |
| **Decision Tree**  **(st dev)** | | **System-dynamic model**  **(st dev)** | | **Decision Tree**  **(st dev)** | | **System-dynamic model**  **(st dev)** | |
| **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** |
| **Number of flu cases (per 1000)** | 49 | 105 | 47 | 104 | 49  (8) | 105  (9) | 172  (186) | 196  (195) |
| **Expected Costs ($)** | 43.66  (6.69) | 45.52  (8.03) | 45.25  (6.85) | 48.72  (8.99) | 44.58  (6.87) | 47.18  (8.11) | 86.12  (56.68) | 84.98  (59.43) |
| **Number of flu cases averted (per 1000)** | 56 |  | 57 |  | 56 |  | 24 |  |
| **Incremental costs** | -1.86 | -3.47 | -2.60 | 1.14 |
| **ICER ($/flu cases averted)** | **Dominant** | **Dominant** | **Dominant** | **0.0475** |

**Table 7:** Expected results under a varied vaccination scenario for children age 2 to 5. Assumes random mixing amongst children.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **(A) 16% (immediate- children vaccinated prior to start of model)** | | | | **(B) 32% (immediately prior to start of model)** | | | | **(C) 16% (rolling- real-world vaccination schedule)** | | | |
| **System Dynamics** | | **Agent-based model**  **(st dev)** | | **System Dynamics** | | **Agent-based model**  **(st dev)** | | **System Dynamics** | | **Agent-based model**  **(st dev)** | |
| **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** |
| **Number of flu cases (per 1000)** | 637 | 642 | 565  (495) | 585  (492) | 567 | 579 | 459  (496) | 495  (498) | 641 | 646 | 521  (499) | 551  (497) |
| **Expected Costs ($)** | 120.64 | 120.50 | 107.80  (89.53) | 110.25  (88.78) | 113.71 | 113.61 | 94.40  (89.55) | 98.29  (89.87) | 120.26 | 119.34 | 97.93  (87.17) | 102.48  (87.38) |
| **Number of flu cases averted (per 1000)** | 5 |  | 20 |  | 12 |  | 36 |  | 5 |  | 30 |  |
| **Incremental costs** | 0.14 | -2.45 | 0.10 | -3.89 | 0.92 | -4.55 |
| **ICER ($/flu cases averted)** | **0.028** | **Dominant** | **0.008** | **Dominant** | **0.184** | **Dominant** |

**Table 8:** Expected results of ABM across varying network topology for children vaccinated between the ages of 2 to 5. Assumes 100% vaccination prior to start of model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1. **Random**   **(st dev)** | | **(B) Small world**  **(st dev)** | | **(C) Scale-free**  **(st dev)** | |
| **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** |
| **Number of flu cases (per 1000)** | 54  (162) | 120  (287) | 39  (114) | 80  (194) | 9  (22) | 13  (50) |
| **Expected Costs ($)** | 45.60  (29.45) | 49.78  (51.67) | 42.81  (20.70) | 42.93  (35.29) | 37.37  (4.51) | 30.35  (9.04) |
| **Number of flu cases averted (per 1000)** | 66 |  | 41 |  | 4 |  |
| **Incremental costs** | -4.18 | -0.12 | 7.02 |
| **ICER ($/flu cases averted)** | **Dominant** | **Dominant** | **1.755** |

**Table 9:** Deterministic results of decision tree and ABM when incorporating heterogeneity by expanding the ages of patients modelled. Assumptions include 100% vaccination prior to start of model with random-mixing amongst children.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **(A) Decision Tree:**  **Random mixing within age cohort. No interaction outside of age cohort.** | | **(B) ABM**  **Random mixing within age cohort. No interaction outside of age cohort.**  **(st dev)** | | **(C) ABM**  **Random mixing between age cohort**  **(st dev)** | |
| **LAIV** | **TIV** | **LAIV** | **TIV** | **LAIV** | **TIV** |
| **Number of flu cases (per 1000)** | 49 | 105 | 53  (91) | 99  (181) | 28  (88) | 68  (212) |
| **Expected Costs ($)** | 39.97 | 38.74 | 32.24  (14.53) | 31.95  (28.16) | 27.92  (14.52) | 26.91  (34.12) |
| **Number of flu cases averted (per 1000)** | 56 |  | 46 |  | 40 |  |
| **Incremental costs** | 1.23 | 0.29 | 1.01 |
| **ICER ($/flu cases averted)** | **0.022** | **0.006** | **0.025** |

**Appendix 4.I:** Model Parameters (for children, aged 2-5)

(A) Clinical-related parameters

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **LAIV** | | | | **TIV** | | | | **Data sources** |
| **Mean** | **Distribution** | **Alpha** | **Beta** | **Mean** | **Distribution** | **Alpha** | **Beta** |
| **Clinical probabilities** | | | | | | | | | |
| Probability of developing influenza\* | 0.0490 | -- |  |  | 0.1050 | -- |  |  | Luce et al, 200831 |
| Proportion of influenza cases that are uncomplicated | 0.67 | Beta | 24.15 | 712.51 | 0.67 | Beta | 92.92 | 1232.43 | Tarride et al, 201216 |
| Proportion of influenza cases with AOM | 0.12 | Beta | 24.85 | 4207.33 | 0.12 | Beta | 98.73 | 7764.04 | Tarride et al, 201216 |
| Proportion of influenza cases with LRI | 0.14 | Beta | 24.82 | 3599.04 | 0.14 | Beta | 98.52 | 6626.57 | Tarride et al, 201216 |
| Proportion of influenza cases with AOM + LRI | 0.07 | Beta | 24.91 | 7248.82 | 0.07 | Beta | -- | -- | Tarride et al, 201216 |
| Probability of death for uncomplicated influenza | 1.034 e-6 | -- | -- | -- |  | | | | Luce et al, 200831 |
| Probability of death for influenza + AOM | 1.034 e-6 | -- | -- | -- |  | | | | Luce et al, 200831 |
| Probability of death for influenza + LRI | 2.473e-6 | -- | -- | -- |  | | | | Luce et al, 200831 |
| Probability of death for influenza + AOM + LRI | 2.473e-6 | -- | -- | -- |  | | | | Luce et al, 200831 |
| **Infectious-related parameters\*\*** | | | | | | | | | |
| Secondary attack rate (vaccinated) | 0.0330 | -- | -- | -- | 0.0345 | -- | -- | -- | Calibration |
| Secondary attack rate (unvaccinated) | 0.054 | -- | -- | -- |  | | | | 20 |
| Number of contacts per susceptible (per day) | 5 | -- | -- | -- |  | | | | Estimate |
| **Time-dependent parameters (days)** | | | | | | | | | |
| Average duration for MSW | 12.79 | Normal (9.423) | | |  | | | | Belshe, 200732 |
| Average duration for uncomplicated influenza | 4.05 | Normal (2.236) | | |  | | | | Belshe, 200732 |
| Average duration for complicated influenza (i.e. AOM &/ LRI) | 11.06 | Normal (7.803) | | |  | | | | Belshe, 200732 |

\*Parameter only applicable to static models

\*\*Parameter only applicable to dynamic models

**Abbreviation**: AOM: acute otitis media; LAIV, intranasal live attenuated influenza vaccine; LRI, lower respiratory infection; MSW, medically significant wheezing; TIV, injectable trivalent inactivated influenza vaccine

(B) Costing-related parameters

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **LAIV** | | | | **TIV** | | | | | | | **Data sources** |
| **Mean** | **Distribution** | **Alpha** | **Beta** | **Mean** | | **Distribution** | | **Alpha** | | **Beta** |
| **Resource utilization** | | | | | | | | | | | | |
| Proportion requiring two doses of vaccine | 0.63 | Beta | 8.62 | 5.06 |  | | | | | | | Tarride et al, 201216 |
| Probability of hospitalization due to MSW | 0.0425 | -- | -- | -- | 0.0714 | | -- | | -- | | -- | Luce et al, 200831 |
| Probability of hospitalization due to uncomplicated influenza | 7.778e-4 | -- | -- | -- | 7.873e-4 | | -- | | -- | | -- | Sebastian et al, 200833 |
| Probability of hospitalization due to influenza + AOM | 0.0043 | -- | -- | -- | 0.0044 | | -- | | -- | | -- | Sebastian et al, 200833 |
| Probability of hospitalization due to influenza + LRI | 8.716e-4 | -- | -- | -- | 8.805e-4 | | -- | | -- | | -- | Sebastian et al, 200833 |
| Probability of hospitalization due to influenza + AOM + LRI | 0.0017 | -- | -- | -- | 0.0018 | | -- | | -- | | -- | Sebastian et al, 200833 |
| Ratio of ER visit to hospitalization due to MSW | 1.81 | -- | -- | -- |  | | | | | | | Luce et al, 200831 |
| Ratio of ER visit to GP visits due to any forms of influenza | 11.3 | Normal (2.26) | | |  | | | | | | | Luce et al, 200831 |
| Average number of GP visits due to MSW | 2.4 | -- | -- | -- |  | | | | | | | Tarride et al, 201216 |
| Average number of GP visits due to uncomplicated influenza | 2.1 | -- | -- | -- |  | | | | | | | Tarride et al, 201216 |
| Average number of GP visits due to influenza + AOM | 2.3 | -- | -- | -- |  | | | | | | | Tarride et al, 201216 |
| Average number of GP visits due to influenza + LRI | 2.3 | -- | -- | -- |  | | | | | | | Tarride et al, 201216 |
| Average number of GP visits due to influenza + AOM +LRI | 2.1 | -- | -- | -- |  | | | | | | | Tarride et al, 201216 |
| Probability of blood test given uncomplicated influenza | 0.14 | Beta | 21.31 | 128.75 |  | | | | | | | Luce et al, 200831 |
| Probability of blood test given complicated influenza | 0.18 | Beta | 20.40 | 94.84 |  | | | | | | | Luce et al, 200831 |
| Probability of X-ray given MSW | 0.03 | Beta | 16.68 | 35.45 |  | | | | | | | Luce et al, 200831 |
| Probability of X-ray given uncomplicated influenza | 0.02 | Beta | 24.48 | 1199.52 |  | | | | | | | Luce et al, 200831 |
| Probability of X-ray given complicated influenza | 0.08 | Beta | 22.97 | 271.54 |  | | | | | | | Luce et al, 200831 |
| Probability of viral culture given any forms of influenza | 1 | -- | -- | -- |  | | | | | | | Luce et al, 200831 |
| Probability of pulmonary test given uncomplicated influenza | 0.006 | Beta | 2484 | 4115.82 |  | | | | | | | Luce et al, 200831 |
| Probability of pulmonary test given complicated influenza | 0.04 | Beta | 24.06 | 644.38 |  | | | | | | | Luce et al, 200831 |
| Probability of antibiotics given uncomplicated influenza | 0.23 | Beta | 18.99 | 63.23 |  | | | | | | | Luce et al, 200831 |
| Probability of antibiotics given complicated influenza | 0.61 | Beta | 9.14 | 5.84 |  | | | | | | | Luce et al, 200831 |
| Probability of analgesic given uncomplicated influenza | 0.01 | Beta | 24.77 | 2727.01 |  | | | | | | | Luce et al, 200831 |
| Probability of analgesic given complicated influenza | 0.02 | Beta | 24.56 | 1420.03 |  | | | | | | | Luce et al, 200831 |
| Probability of bronchodilators given MSW | 0.96 | Beta | 13.79 | 0.57 |  | | | | | | | Luce et al, 200831 |
| Probability of bronchodilators given influenza with LRI | 0.53 | Beta | 11.22 | 9.95 |  | | | | | | | Luce et al, 200831 |
| **Costing parameters** | | | | | | | | | | | | |
| Cost of vaccine | 14 | Gamma | 25 | 0.56 | 6.79 | Gamma | | 25 | | 0.2716 | | Tarride et al, 201216 |
| Cost of vaccine administration | 3.59 | Gamma | 25 | 0.14 |  | | | | | | | Sander et al, 201034 |
| Cost of hospitalization | 3641.59 | Gamma | 1 | 3641.59 |  | | | | | | | CIHI; OHIP schedule of benefits and fees |
| Cost of ER visit due to MSW | 221.39 | Gamma | 25 | 8.86 |  | | | | | | | OCCI; OHIP schedule of benefits and fees |
| Cost of ER visit due to Influenza | 232.77 | Gamma | 25 | 9.31 |  | | | | | | | CIHI; OHIP schedule of benefits and fees |
| Cost of GP visit | 42.35 | Gamma | 25 | 1.694 |  | | | | | | | OHIP schedule of benefits and fees |
| Cost of blood test | 8.30 | Gamma | 25 | 0.33 |  | | | | | | | OHIP Scheduled of Benefits for Laboratory Services |
| Cost of x-ray | 66.83 | Gamma | 25 | 2.67 |  | | | | | | | St Joseph’s Health Care; OHIP |
| Cost of viral culture | 15.51 | Gamma | 25 | 0.62 |  | | | | | | | OHIP Scheduled of Benefits for Laboratory Services |
| Cost of pulmonary test | 70.70 | Gamma | 25 | 2.82 |  | | | | | | | OHIP Schedule of Benefits |
| Cost of antivirals | 40.56 | -- | -- | -- |  | | | | | | | Calculated from 4 local pharmacies. |
| Cost of antibiotics | 25.70 | Gamma | 25 | 1.03 |  | | | | | | | ODB |
| Cost of analgesic | 8.00 | Gamma | 25 | 0.32 |  | | | | | | | Calculated from 4 local pharmacies. |
| Cost of nebulized bronchodilator for MSW | 13.65 | Gamma | 25 | 0.55 |  | | | | | | | ODB |
| Cost of nebulized bronchodilator for Influenza | 11.57 | Gamma | 25 | 0.46 |  | | | | | | | ODB |
| Cost of nebulizer kit | 21.99 | Gamma | 25 | 0.88 |  | | | | | | | ODB |

\*Parameter only applicable to static models

\*\*Parameter only applicable to dynamic models

**Abbreviation**: AOM: acute otitis media; LAIV, intranasal live attenuated influenza vaccine; LRI, lower respiratory infection; MSW, medically significant wheezing; TIV, injectable trivalent inactivated influenza vaccine

**Appendix 4.II:** Model Parameters that Differ Across Age Groups (specific to the decision tree and agent-based model)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | LAIV | | | TIV | | | Data sources |
| 2-5 years old | 6-9 years old | 10-17 years old | 2-5 years old | 6-9 years old | 10-17 years old |
| **Infectious-related parameters** | | | | | | | |
| Secondary attack rate (vaccinated) | 0.0306 | 0.0323 | 0.0342 | 0.0327 | 0.0345 | 0.0369 | Calibrated |
| **Resource utilization** | | | | | | | |
| Probability of hospitalization due to uncomplicated influenza | 7.778e-4 | 0.0017 | 1.635e-4 | 7.873e-4 | 7.963e-4 | 7.645e-5 | Sebastian, 200833 |
| Probability of hospitalization due to influenza + AOM | 0.0043 | 0.13 | 0.0020 | 0.0044 | 0.0064 | 0.0010 | Sebastian, 200833 |
| Probability of hospitalization due to influenza + LRI | 8.716e-4 | 0.0094 | 9.434e-4 | 8.805e-4 | 0.0044 | 4.410e-4 | Sebastian, 200833 |
| Probability of hospitalization due to influenza + AOM + LRI | 0.0017 | 0.0204 | 0.0020 | 0.0018 | 0.0096 | 9.556e-4 | Sebastian, 200833 |
| Probability of blood test given uncomplicated influenza | 0.14 | 0.09 | 0.09 |  | | | Loughlin, 200335 |
| Probability of blood test given complicated influenza | 0.18 | 0.20 | 0.20 |  | | | Loughlin, 200335 |
| Probability of X-ray given uncomplicated influenza | 0.020 | 0.023 | 0.023 |  | | | Loughlin, 200335 |
| Probability of X-ray given complicated influenza | 0.08 | 0.11 | 0.11 |  | | | Loughlin, 200335 |
| Probability of pulmonary test given uncomplicated influenza | 0.006 | 0.008 | 0.008 |  | | | Loughlin, 200335 |
| Probability of pulmonary test given complicated influenza | 0.036 | 0.042 | 0.042 |  | | | Loughlin, 200335 |
| Probability of antibiotics given uncomplicated influenza | 0.23 | 0.27 | 0.27 |  | | | Loughlin, 200335 |
| Probability of antibiotics given complicated influenza | 0.61 | 0.63 | 0.63 |  | | | Loughlin, 200335 |
| Probability of analgesic given uncomplicated influenza | 0.009 | 0.011 | 0.011 |  | | | Loughlin, 200335 |
| Probability of analgesic given complicated influenza | 0.02 | 0.01 | 0.01 |  | | | Loughlin, 200335 |

**Abbreviation**: AOM: acute otitis media; LAIV, intranasal live attenuated influenza vaccine; LRI, lower respiratory infection; MSW, medically significant wheezing; TIV, injectable trivalent inactivated influenza vaccine

**CHAPTER 5:**

**Discussion and Conclusion**

**5.1 Summary and Major Contributions**

Modelling is a methodology to characterize, understand and study a system. Despite the wide range of reasons for which to model, decision-analytic models all share a common overarching purpose: rather than being a substitute for critical thought, such models are tools to improve judgement and intuition with the intent of helping decision-makers reach a more evidence-informed and rational decision regarding the consequences of their actions under conditions of uncertainty.

In particular, health economic models provide an account of the expected costs and outcomes associated with alternative health interventions. These models play an important role in allocating society’s scarce resources in order to meet the public health concerns of preventing disease, prolonging life and promoting health to the whole community. A good quality economic model is one that clearly and accurately captures all relevant options in order to provide an opportunity to explore the implications of different decisions under varying “what-if” conditions1. Furthermore, to serve its purpose, models must produce accurate and valid predictions in addition to capturing the substantial variation and uncertainty in the factors that may influence costs and clinical effects (e.g., sensitivity analysis)2-4.

With the growth of decision-analytic health economic modelling in recent years, greater focus has been placed on defining good practice guidelines to support higher-quality model development. Indeed, organizations and individuals who work with decision-analytic models – whether by constructing them to address a particular problem or by interpreting their results to support policy decisions –have developed systematic framework within which to operate and critically appraise such work. Much has been done in developing good research guidelines, such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Society for Medical Decision Making (SMDM) joint task force groups3, HTA-agency led initiatives5 and the work of individual groups6,7. These guidelines span across the modelling process, including: model conceptualization3,5,6,8; the incorporation of costs and effects2,3,5-7; model construction; interpretation of results; model validation3,4; and transferability into other jurisdiction9.

Despite the extensive lengths and efforts devoted to guideline development, one area that has received limited attention by academics and policy-makers is how to choose an appropriate modelling approach. This is particularly concerning given the growing range of ‘alternative’ approaches to modelling that have been adopted from other disciplines and applied to the health context. As mentioned previously, every approach to modelling is based on a unique set of assumptions on how a system should be captured10. Yet, there is often limited understanding and acknowledgement of what these assumptions are and how these assumptions impact the modelling process (e.g., practicalities to modelling, validity of the model, reliability of its results). Improved knowledge on this matter may lead to a better ability by modellers, who construct these decision-analytic models, and decision-makers, who apply these models to support health care policy decisions, in distinguishing one approach from another and in properly applying an appropriate and unbiased method for a given decision problem.

Each chapter in this thesis was aimed at investigating the assumptions specific to each approach to modelling and to better understand how this may impact the overall modelling process in order to ultimately answer, if, and under what conditions, each approach may be most appropriate. The inquiry addressed this topic from both a theoretical standpoint through decision frameworks and an empirical perspective by gathering and conducting cross-validation exercises. This final chapter offers a summary of the thesis findings and the main contributions of each study, and of the overall thesis, in addressing the question on how to select a modelling approach. It is further accompanied by a discussion of the implications of the results and potential areas of future research.

The first systematic literature review identified eight frameworks and taxonomies that aimed to differentiate between model types and to assist in the selection of a modelling approach11. This represents the first systematic literature review that has been conducted on this topic. An important finding from this review was that, although most existing frameworks were unique in the approaches covered and the combination of selection criteria considered, there was overall poor agreement between frameworks with respect to how they described the features specific to each modelling approach. Only the structural criteria were included in the comparison as these, in theory, should remain consistent irrespective of the context of the decision problem. Yet, this did not hold true as frameworks differed in how they classified the structural elements of each approach. This is particularly important as it meant that different frameworks and taxonomies may provide inconsistent guidance on what they consider to be the most suitable modelling approach for a given decision problem. An alternative perspective to interpreting these findings is that perhaps more than one modelling approach may be appropriate for a particular decision problem.

To address the topic of whether different approaches to modelling are in fact necessary, this topic was explored further but from an empirical perspective. Highly-dependent cross-validation exercises are an ideal means to address this question since, beyond the modelling approaches being different, the assumptions and data parameters are nearly identical between the models being compared. Our systematic literature review on empirical cross-validated studies identified consistent findings12 with respect to the impact some of the more commonly studied factors (i.e., population resolution and interactivity) had on the choice of an approach to modelling and the overall modelling process.

Amongst the studies that explored the structural element of population resolution, aggregate-level and individual-level models both produced similar results. However, a practical implication was observed in terms of a trade-off between face validity and feasibility. Individual-level models could more easily and compactly reflect features of a disease (e.g., memory, heterogeneity, time-dependencies) while keeping data in its existing form. Aggregate-level models, on the other hand, albeit still capable of capturing such complex features, required the introduction of health states (i.e., state explosion). In contrast, aggregate-level models were associated with shorter time requirements in terms of validating, computing and conducting probabilistic sensitivity analysis. These findings are important as they lend support that, when deciding the level of resolution in which to model a system, it is the practical concerns relating to the modelling process that emerge as having a greater impact on the modelling approach.

When it came to the element of interactivity, this factor - on its own - had a considerable impact on model validity. If indirect effects were marginal within a system, both the results of the dynamic and static models converged. Examples of such instances include when the force of infection is constant or when there is a low likelihood of disease transmission. However, if indirect effects were significant, static and dynamic models were often found to produce divergent results that could lead to conflicting policy recommendations. A similar trade-off between face validity and feasibility was also observed. Static models generally required less data and were quicker to construct while dynamic models offered greater face validity as they more easily captured relevant and important features of the decision problem.

Given that consistent conclusions could be drawn across heterogeneous decision problems, this provided credence on the potential impact of each factor on model selection and on the subsequent modelling process. However, existing cross-validation studies12 have only involved a comparison between a traditional approach (i.e., Markov and decision tree) and alternatives such as microsimulation, discrete event simulation and compartmental models. Few to none of the exercises were found to involve system dynamics (SD) or agent-based models (ABM). As such, it was uncertain whether these findings regarding the factor’s impact would hold when using these approaches. To address this methodological gap, a pre-existing decision tree on childhood influenza vaccination13 was replicated into SD and ABM to explore both the factors of population resolution and interactivity. Each model was calibrated using identical parameters and assumptions. From scenario analyses, it was possible to explore the assumptions underlying these three modelling approaches. This work represents the first time a cross-validation exercise has studied ABM and is furthermore the first time, to our knowledge, that ABM is employed empirically in health economic evaluations rather than presented as hypothetical examples.

Upon calibration of the three modelling approaches, the deterministic point-estimates that each produced were found to be similar. Intranasal live attenuated vaccination was the dominant vaccination strategy for young children. The higher unit cost of intranasal live attenuated vaccination compared to injectable inactive influenza vaccine was offset by preventing more flu cases that led to a decrease in health care resource utilization and thus lower expected costs. However, this may represent an idealized circumstance.

The impact of the factors triangulated with the findings of the systematic review despite studying different modelling approaches. Aggregate-level models (i.e., decision tree and SD) generally required less time to construct and run; whereas, this was not the case for individual-level model (i.e., ABM). Indeed, probabilistic sensitivity analyses were not conducted in the ABM as the number of permutations required was considered unfeasible. Nevertheless, aggregate-level models required a degree of simplification, in part, due to its restrictive assumptions concerning homogeneity within each component (i.e., nodes in the decision tree, stocks in the SD). While the nature of the interactions between individuals could be easily modelled in an ABM, aggregate-level SD relied on an idealized assumption of random-mixing. The only potential solution would be to add new components to the cohort-level models in order to overcome some of these limitations but not without the cost of combinatorial explosion. Although similar conclusions were reached when heterogeneity was incorporated in both the decision tree and the ABM, under the former, an entire model had to be duplicated with restricted assumptions on the nature of the interaction whereas only one parameter had to be introduced in the latter with flexibility in describing the nature of interaction.

These approaches further differ in their ability to describe changes to the infection rates (i.e., clinical attack rate) over time. Disease transmission must be a known, pre-specified input to the decision tree[[1]](#footnote-1) (i.e., static model) whereas, in ABM and SD (i.e., dynamic models), transmission is handled as an output according to a function of infectivity and the underlying contact structure. This is far more representative of the underlying epidemiology of infectious diseases. The findings from simulation studies suggest that dynamic models offered greater flexibility by explicitly modelling the mechanism of disease transmission and this may impact the modelled results, especially if details regarding transmission are unknown. This supports the need for dynamic models, especially when exploring possible public health strategies that target or impact parameters relating to transmission (e.g., increasing the rate of vaccination).

ABM, as an individual-level dynamic model, offered the most flexibility and strongest face validity. Indeed, this granularity permitted the explicit modelling of person-to-person interactions within a network structure and potentially more realistic contact structures. This may impact the cost-effectiveness results and generate incremental cost-effectiveness ratios that fall into different quadrants of the incremental cost-effectiveness plane. Thus, in infectious disease modelling, if the nature of disease transmission is highly dependent on individuals’ or the infectious agent’s behaviours and/or the contact structure, ABM may be the most appropriate approach.

From this thesis, a key theme that emerged on the topic of model selection was the concept of trade-offs. Indeed, when selecting between the factors that describe models (e.g., cohort vs. aggregate, static vs. dynamic), modellers must be aware of the trade-offs in terms of balancing the ability of a model to describe the natural history of a condition (e.g., treatment pathway) and the pragmatic consideration of being able to put a model together. This has been referred to as the face validity-feasibility trade-off(14 and highlights the fine balance that characterizes any modelling endeavours. A balance must be struck between building a model that is reflective of the intricacies and complexities of reality while distilling it down to its core fundamental pieces.

**5.2 Future Areas of Research**

Research from this thesis has identified areas in which further work may be warranted. In particular, an obstacle identified from the two systematic reviews (Chapters 2 and 3) was the lack of standardization in the definition and use of terminology. For instance, Markov microsimulation has been referred synonymously to “individual sampling model”15, “patient-level simulation”16,17, “Monte Carlo Markov models18, “Monte Carlo simulation/microsimulation”19 and “First-order Markov model”20. This is likely problematic as it can hamper communication and progress on methodological work since a universal language does not exist in which to describe models and their features. This may have arisen as a consequence of the technical expertise of modellers who come from a broad range of disciplines (e.g., operations research, infectious disease epidemiologist, statistics, economists) and who have brought along the terminology native to their respective disciplines. As the field of decision-analytic modelling grows, it is increasingly necessary to bring together individuals from these different disciplines to develop a common and unified language.

Given the consistency identified so far in the handful of empirical cross-validation exercises (Chapters 2 and 3), value remains in conducting further empirical work to continue one’s understanding of the trade-offs between modelling approaches, especially in some of the less-studied approaches (e.g., agent-based models, system dynamics) and in some of the less-studied factors (i.e., handling of time, resource constraints). It would be of interest to ascertain whether the trade-off between face validity and feasibility that was noted for both factors of population resolution and interactivity would hold true in the context of the remaining factors. Additional comparative evidence would further offer insights to better understand the benefits and limits associated with each approach. This, in turn, could be invaluable towards informing the selection of a modelling approach for decision-analytic evaluations which can ultimately impact the underlying validity of a model and the policy it seeks to inform.

A better understanding of the features of various models and the extent to which they may impact the validity of the modelled results would likely improve how models are presently selected.Cross-validation may be the key towards developing an evidence-based framework by highlighting the circumstances and the extent to which modelling approaches may produce similar or diverging results. This would be useful as it could provide empirical insight that may help researchers in selecting the appropriate modelling approach rather than relying on existing frameworks that are based on heuristics and share a high level of disagreement.

The trade-off between face validity and feasibility is closely related to a debate between simplicity and complexity. This is an interesting area of debate as there is no agreement amongst modellers regarding this topic. Some argue that a model should be as parsimonious as possible; when selecting a model, it is important to select one that can capture the necessary level of detail for a particular system such that effects vital to understanding a policy question are captured but, equally important, any unnecessary data are omitted21,22. The incremental informational value of adding additional data and complexity to a model must be balanced with the added time and effort required to develop and construct such a model22. It is evident that simpler models are easier to parameterize, adapt and interpret. However, some contend that simpler models are designed to address a limited range of questions and can be misleading if used to address broader issues18. In contrast, more complex models may offer greater flexibility and increased efficiency but may potentially require supplementary analytic knowledge, greater time commitment and more extensive data needs. Can complexity be quantified in such a manner that an optimal point is identified in which a model is sufficiently comprehensive and valid; yet simplistic enough that it is feasible? This debate remains unresolved.

**5.3 Main concluding points of the thesis**

This thesis highlights that, rather than simply defining the ‘best’ modelling approach, there is greater value to instead establish the most appropriate circumstances in which to use one approach over another and to understand the strengths and limitations specific to each approach. Despite several frameworks and taxonomies that have been published to differentiate between the various modelling approaches, a systematic literature review found that, in general, there is a lack of consistency between these frameworks in how different modelling approaches are distinguished and, consequently, inconsistent recommendations11. There is no evidence that existing frameworks have taken the insights gained from cross-validation studies to support its recommendation. This is of particular concern given that, despite the rapid evolution in advanced techniques for health economic modelling, certain aspects of its application remain poorly standardized. Inappropriate use of modelling approaches that are unable to adequately capture a system may lead to incorrect results and even erroneous conclusions with the potential consequences of generating poor advice and a waste of scarce resources.

The development of an evidence-based framework to guide the selection of a modelling approach may therefore be the next logical step forward. Cross-validation exercises have so far highlighted some of the circumstances and the extent to which modelling approaches may produce similar or diverging results across a multitude of disease areas12. Such exercises may prove useful as they provide real-world insight in which to help researchers in their task of selecting an appropriate modelling approach. By better understanding the features of various approaches to modelling and the extent to which it may impact the validity of the results, this may improve the decision as to which modelling approach should be selected for different decision problems.

**5.4 References**

1. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation*.* Oxford University Press, Oxford, UK; 2006.

2. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health.* 2012; 15(6), 835-842.

3. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Med Decis Making.* 2012; 32(5), 667-677.

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. *Med Decis Making.* 2012; 32(5), 733-743.

5. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies3rd Ed*.* Ottawa, Canada; 2006.

6. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Int J Technol Assess Health Care.* 2013; 29(2), 117-122.

7. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics.* 2006; 24(4), 355-371.

8. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a Model A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. *Med Decis Making.* 2012; 32(5), 678-689.

9. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, Reed SD, Rutten F, Sculpher M, Severens J. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health.* 2009; 12(4), 409-418.

10. Meadows DH, Robinson JM. The Electronic Oracle: Computer Models and Social Decisions. John Wiley & Sons, Chichester, UK; 1985.

11. Tsoi B, O'Reilly D, Jegathisawarn J, Tarride J-E, Blackhouse G, Goeree R. Systematic narrative review of decision frameworks to select the appropriate modelling approaches for health economic evaluations. *BMC Res Notes.* 2015;[in submission].

12. Tsoi B, Goeree R, Jegathisawarn J, Tarride J-E, Blackhouse G, O’Reilly D. Do different decision-analytic modelling approaches produce different cost-effectiveness results? Insights from a systematic review of existing cross-validation studies. *Exp Rev Pharmacoecon Outcomes Res.* 2015; [epub ahead of print].

13. Tarride J-E, Burke N, Von Keyserlingk C, O’Reilly D, Xie F, Goeree R. Cost-effectiveness analysis of intranasal live attenuated vaccine (LAIV) versus injectable inactivated influenza vaccine (TIV) for Canadian children and adolescents. *Clinicoecon Outcomes Res* 2012; 4, 287-298.

14. Kuehne FC, Chancellor J, Mollon P. Microsimulation or cohort modelling? A comparative case study in HIV infection. iHEA 2007 6th World Congress: Explorations in Health Economics Paper. 2007. Available from: <http://ssrn.com/abstract=994349>.

15. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J Health Serv Res Policy.* 2004; 9(2), 110-118.

16. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006; 15(12), 1295-1310.

17. Chick SE. Taxonomy of model structure for health economics. Assessed June 13, 2013. Link: http://www2.wmin.ac.uk/hscmg/qmmhealth2007/talks/Chick\_S.-.IMAHealth2007.-.Keynote.pdf

18. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment. *Pharmacoeconomics* 2008; 26(2), 131-148.

19. Kim S-Y, Goldie SJ. Cost-effectiveness analyses of vaccination programmes. *Pharmacoeconomics* 2008; 26(3), 191-215.

20. Heeg BM, Damen J, Buskens E, Caleo S, de CF, Van Hout BA. Modelling approaches: the case of schizophrenia. *Pharmacoeconomics* 2008; 26(8), 633-648.

21. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. A suggested framework and example of application. *Pharmacoeconomics* 2000; 17(5), 461-477.

22. Jit M, Levin C, Brisson M, Levin A, Resch S, Berkhof J, Kim J, Hutubessy R. Economic analyses to support decisions about HPV vaccination in low-and middle-income countries: a consensus report and guide for analysts. *BMC Med.* 2013; 11(1), 23.

1. Knowing the disease prevalence beforehand may not always be feasible. Decision-analytic models are often used as a prediction tools to ascertain which intervention should be funded without actually having the interventions introduced already into the system. However, to know the disease prevalence under a particular treatment strategy would either require that it’s already in the system or the need for estimation/extrapolation to generate such a value. [↑](#footnote-ref-1)