OPEN-SOURCE MODELING AND RESOURCE PRIORITIZATION IN HEALTHCARE

Eline Krijkamp

COLOFON

Open-Source Modeling and Resource Prioritization in Healthcare

In writing this thesis Eline Krijkamp and the co-authors have done their best to rely on the most updated sources and the most accurate facts available at the time of writing. Yet as with any human endeavor, mistakes are inevitable. If you spot any error, please highlight them by contacting Eline Krijkamp so that she can do her best to correct them.

Financial support for this thesis was kindly provided by the Society for Medical Decision Making (SMDM) fellowship through a grant by the Gordon and Betty Moore Foundation (GBMF7853) and the Department of Epidemiology of the Erasmus Medical Center, Rotterdam, the Netherlands.

Design and layout by Deksels Design
Cover design by Deksels Design
Printing by Optima Grafische Communicatie (www.ogc.nl)

Open-Source Modeling and Resource Prioritization in Healthcare

Open-source modellering en prioritering van beschikbare middelen in de gezondheidszorg

Proefschrift
ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. dr. A.L. Bredenoord
en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 14 juni 2022 om 13:00 uur

door

Eline Mariëlle Krijkamp
geboren te Haarlem

Erasmus University Rotterdam

Promotiecommissie

Promotor	Prof. dr. M.G.M. Hunink
Overige leden	Prof. dr. M.P.M.H. Rutten-van Mölken Prof. dr. ir. H. Koffijberg Dr. B.S. Ferket
Copromotor	Dr. P. Pechlivanoglou

CONTENTS

01

GENERAL INTRODUCTION

“Learning to choose is hard. Learning to choose well is harder. And learning to choose well in a world of unlimited possibilities is harder still, perhaps too hard.”

Barry Schwartz

GENERAL INTRODUCTION

The overarching goal of this thesis is to expand knowledge on the use of open-source modeling in decision analysis in healthcare and to demonstrate how open-source decision models can contribute to resource prioritization in healthcare.

Decision analysis in healthcare

Decision making is an essential part of everyday clinical practice. These decisions are usually made under conditions of uncertainty and resource constraints. Examples of decisions faced by decision makers include how to use common drug therapies in patients with multimorbidity[1], finding the optimal prehospital transportation strategy for ischemic stroke patients[2,3] or how to prioritize patients during shortages of intensive care unit beds in a pandemic under-capacity scenario.[4]

Similarly, policy makers need to make population-level funding decisions for national health programs (e.g., reimbursement of a therapy[5–7] or national screening program[8–11]). These resource allocation decisions are often constrained by the annual budget of each program.

Decision analysis in healthcare is a quantitative approach to analyze difficult healthcare decisions under conditions of uncertainty. It provides a formal framework to help understand the decision problem. Decision analysis aims to identify the optimal decision and which variables of the problem might have a major impact on that decision.[12]

Healthcare decision analysis is mostly used to guide population-based decisions and support the development of treatment guidelines and recommendations.[5,11,13,14] Although, they can also be used to answer questions on the individual-patient level, for example, to determine the most appropriate treatment for a patient with specific characteristics (e.g., age, sex, tumor type).[15–17]

Decision analysis involves systematic evaluation and communication of the evidence about benefits, harms and costs associated with various possible interventions. More specifically, decision analysis includes steps such as to specify the decision problem and decision rule, to identify the benefits and harms associated with the possible interventions, to identify the preferences of patients and society regarding the outcomes of the intervention, to compile the best available evidence about the trade-offs, to systematically analyze this evidence and finally to present the expected outcomes of the intervention(s) and corresponding recommendations of the analysis to inform decision making.[12,18]

Outcomes of a decision analysis include costs outcomes and health outcomes associated with interventions considered. Health outcomes are often expressed in life years (LY) or quality adjusted life years (QALY). These outcomes are used to guide decision and policy makers in their choices. Examples of decision makers include, among others, individual patients, healthcare professionals, healthcare organizations, and health technology assessment (HTA) agencies.

Decision-analytic models

Decision analysis in healthcare is most commonly performed using mathematical models to describe decision problems. We call these models decision-analytic models. Like all models, a decision-analytic model is a simplified reflection of reality. However, these simplified models enable us to understand complex scenarios that have not been or cannot be observed in real life. These decision-analytic models are built because they can integrate scientific knowledge from various sources (e.g., clinical trials and observational studies) into the evaluation of real life clinical or policy decision problems.

While randomized clinical trials (RCTs) are the state-of-the-art design for gathering evidence on relative effectiveness of an intervention, they have several limitations[19,20] (e.g., relatively short follow-up time[21,22], limited strategies[22], and a population that is not reflective of the real-world population[23–28]). Thus, the information provided by RCTs is often not sufficient to answer a decision problem where long-term costs and health consequences need to be considered. In those situations, decision-analytic models can be used to simulate patients' trajectories in a long-term perspective using evidence from a trial.[29,30] For a practical example, see the SPRINT trail results[31] and corresponding cost-effectiveness analysis[32]. Furthermore, decision-analytic models synthesize available evidence from different sources. This way trial results can complement each other. Decision analytic-models can also be used to generalize relative effectiveness estimates from trials to other settings (e.g., countries or populations) by parameterizing the model with setting-specific data.[33] In addition, modeling can be done to assess, address and communicate uncertainty around the decision problem. More recently, decision-analytic models have been proposed as a tool to inform study designs by quantifying the added value of collecting further information.[34–36]

Decision-analytic models can evaluate both clinical and economic aspects of the decision and are considered the most comprehensive and suitable approach to inform health policy decision making.[37,38] Characteristic examples include the routine use of these methods in HTA[39], screening and prevention policy making[40–42], and cost-effectiveness evaluation of vaccination for infectious diseases[33,43–45]. The World Health Organization[46] and The International Network of Agencies for Health Technology Assessment (INAHTA) provide an overview of national agencies that use the results of such HTA-evaluations to inform their own ministries of health. Although decision-analytic models are an essential tool to inform stakeholders about healthcare decisions to optimize resource allocation, it is important to realize that it is only a method to help make more informed decisions and it represents one of the many aspects taken into consideration in the decision making process.[37]

Each decision problem is characterized by specific transitions, events, rewards, trade-offs and constraints. To reflect each of those components in the model the modeler can use different modeling techniques like decision trees[47,48], state-transition models[48,49], discrete event simulations[50–53] or dynamic transmission models[44,54]. A useful taxonomy of these model structures is provided by Brennan et al.[55] Guidance to select the appropriate approach is provided by Barton at al.[48] Both the technique selected and the elements incorporated in the model should be sufficiently complex to be realistic, but not unnecessarily complex.[56] The reliability of any of these models depends on how well the model structure, quality of the input data and modeling assumptions reflect the decision problem of interest.

Uncertainty and decision making

To build a reliable decision-analytic model is not always straightforward. Throughout the development process, modelers and stakeholders have to make many choices and assumptions regarding both model structure and model inputs.[48] For example, they have to decide which events and costs are most relevant to include in the model and which data reflect best the treatment effect of the intervention for the population of interest.

Two major components of uncertainty in decision analysis are model structure uncertainty and parameter uncertainty.[12,57] Structural uncertainty is uncertainty related to the assumptions and judgments inherent in constructing the decision-analytic model.[57] Parameter uncertainty is uncertainty in estimation of the parameter of interest and reflects that we are dealing with imperfect information about model input parameters.[57] These model input parameters (e.g., the probability of getting sick or the risk of side effects of a drug) are rarely known exactly because they are informed by (clinical) studies in which a subset of the population is followed to infer or generalize observation from this subset to the larger population. Although this population subset is aimed to be an unbiased representation of the larger population, there remains uncertainty about the true value in the entire population.

Several papers and guidelines exist to help modelers address uncertainty in decision analysis. These documents provide best practices to produce credible standardized (economic) information that is relevant to support decision makers in their choices. The papers from the joint effort between the Society of Medical Decision Making (SMDM) and the Professional Society for Health Economics and Outcomes Research (ISPOR), better known as SMDM-ISPOR Task Force papers, provide a clear methodological structure regarding method selection.[49,53,57–63] National agencies, like the National Institute for Health and Clinical Excellence (NICE) in England, the Canadian Agency for Drugs and Technologies in Health (CADTH) and Care Institute Netherlands (ZIN), provide generic country-specific research guidelines about how to perform economic evaluations in healthcare.[38,64,65]

These national guidelines also aim to provide directions about how to report all the steps from the research question until the final conclusion. For example, the "Report Documentation" chapter from the ZIN guidelines gives instructions on how to report on topics like input parameters (e.g., effectiveness data, costs and quality of life data), base case analysis, sensitivity analysis and also recommendations on how to graphically present these results.[64] Another great resource is the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.[66] They provide recommendations of items to include when reporting evaluations of health interventions in the form of a checklist, called the CHEERS checklist. These guidelines contribute to clear and comprehensive reports of the decision analysis that decision makers can use to inform their decision.

For decision uncertainty to be represented properly, both structural and parameter uncertainty should be considered.[57,65,67] For structural uncertainty, NICE and Briggs et al. state that structural assumptions should be clearly documented and the evidence and rationale to support them provided.[57,65] In case uncertainties in structural assumptions are identified during model development, those should be tested in uncertainty analysis.[57,65] This would involve separate

analyses of a representative range of plausible scenarios to explore the impact of structural uncertainty on cost-effectiveness estimates.[65] However, performing a structural uncertainty analysis is not frequently done and can be challenging. Strong et al. state that quantifying structural uncertainty is difficult since it requires judgements about a model's ability to represent a complex real life decision problem faithfully.[67] To help modelers, they provide some guidance on how to assess structural uncertainty in their manuscript.[67] A practical example is provided by Incerti et al.[68]

Parameter uncertainty should be assessed via sensitivity analysis.[57,65] The preferred method for sensitivity analysis is probabilistic sensitivity analysis (PSA).[64,65,69] PSA enables the uncertainty associated with parameters to be simultaneously reflected in the model outcomes. This parameter uncertainty of the input parameters itself is reflected by probability distributions that best describe the available evidence.[57,65] ZIN recommends to use these PSA results in turn for a value of information analysis to quantify the consequences of the parameter uncertainty.[64]

Alternative sensitivity analysis methods are univariate and multivariate sensitivity analyses. In a univariate sensitivity analysis, parameters values are manually varied over a range of plausible values to test the sensitivity of the model results to specific values of those parameters. A multivariate sensitivity analyses works similarly, but for a combination of parameter values. It can therefore be used to explore combined uncertainty of these parameters. ZIN recommends to use univariate and multivariate sensitivity analyses to provide insight in the relative influences of input parameters on the model outcomes and in the consequences of fixed values in the model (e.g., discount rates and prices).[64]

Model review

An important component of decision analysis in healthcare is the model review process. The aim of the review process is to engender confidence in the model and conclusions. Reviewers are asked to evaluate any assumptions and identify potential limitations that may have impact on the model's results. HTA-agencies that use decision-analytic methods to guide their decisions, such as NICE, have special teams to carry out the model reviewing process.[70] Scientific journals make use of peer-review. The peer-reviewing process, in which other researchers in the field are asked to validate the work, is described on the Elsevier website.[71]

To gain trust in the accuracy and lack of bias of a model, reviewers, decision makers and other stakeholders want to be able to fully review the model structure, input data, and assumptions. This necessitates that these models are transparent and are available to permit adequate review.[72] Surprisingly, national guidelines do not provide specific recommendations on model access, software use or model documentation.[64,65] Furthermore, many scientific journals do not require access to the model files.[73] As a result, HTA reviewers see many different model source files, some more structured than others, and journal publications often come with a technical appendix only.[39,42] These files provide limited transparency about the model structure and assumptions. Therefore, decision-analytic models are often considered to be "black boxes", a weakness of the field, which begs for more transparency.

Improving transparency in decision analysis in healthcare

Increasing transparency of decision-analytic models is desired to achieve an easy review process of the model and to facilitate reproducibility of the model results. This can result in better understanding of the usefulness of the model results to inform decision making and more chances of model reuse.

One way to increase transparency would be to share model source files. An excellent initiative for this is the Global Health Cost Effectiveness Analysis (GH CEA) Registry[74] with its recently launched Open-Source Model Clearinghouse.[68] This Open-Source Model Clearinghouse provides an open-source model repository to share

results and model source files of cost-effectiveness analyses.[74] Publication of the source files to public repositories, (e.g., GitHub[76]) can also facilitate code file sharing. The publication of source files does not make models transparent by default. The source files have to be well structured and accompanied with detailed explanation in order for others to understand it.

There are diverse software programs[77] that can be used to develop decision-analytic models, for example specialized commercial packages, like TreeAge[78,79] or spreadsheet calculators, mainly Microsoft Excel[80]. The source files of these traditionally used software programs have their limitations. Files from TreeAge[78], need in addition to the source file a well-documented script to guide users through the point-and-click interface to obtain the results. Excel models are often structured over a large number of spreadsheets that make use of cross-linking between the sheets, which makes them cumbersome to use and review from a technical point of view.[81] Another limitation is the required paid license to open these files.

Scripts from open-source programming languages, such as R[82] or Python[83], have the potential to overcome many of the challenges that specialized software programs and spreadsheets have. However, since no clear guidance for the implementation of decision-analytic models in these programming languages exists, it is challenging to realize the full potential of open-source programming languages in healthcare. Therefore, it is of high importance to explore how to use open-source software to construct decision-analytic models for healthcare decision making.

Desired goal

Greater transparency of decision-analytic models is needed to increase model credibility, to achieve reproducibility of model results, and to reduce resource waste by avoiding duplicative efforts. The “black box” culture of the decision-analytic model should become a “transparent package” of well-structured model components, including the model itself, structural assumptions, the input parameters, and documentation.

Ideally, this “transparent package” comes with different levels of depth. The first aims to increase model credibly. A clear report about the full model can be used to evaluate clinical realism (i.e., the ability of the model to correctly reflect the decision problem). Form this documentation it should be clear that the model can appropriately inform the decision and that the used software and available data synthesis methods were not a bottleneck for choices about for the modeling approach. In addition, this report needs to provide clear results and interpretation of the decision uncertainty.

The second level would be to facilitate reproducibility of the results. Replication of results is one of the foundations of academic sciences and increases trust in the model results. Replication of the results requires preferably access to the source code, or very detailed documentation about how to replicate the model.

Finally, the third level, would be to facilitate reusability of the model. This avoids replication efforts when new clinical evidence or an alternative intervention becomes available. Updating the existing models, based on new evidence, can be achieved by having clearly documented decision-analytic model with guidance on how to modify model parameters.

Achieving these levels of transparency will contribute to better and more efficiently informed decision making in healthcare. Open-source programing languages have the potential to facilitate transparency in decision-analytic models. In this thesis, I will explore how open-source software can be of benefit to health decision making and will establish a practical approach to construct advanced open-source decision-analytic models to inform healthcare decisions.

Open-source modeling

Open-source software is software that is freely available and gives users the right to study, modify or distribute the software for any purpose.[84] Both the source code, which are useful for programmers, and the compiled code, which are used by end-users, needs to be freely distributed for software to be classified as open-source software. Any of the program code altered by either programmers or end-users need to be redistributed under the same conditions as the license for the original version of the software.

An example of open-source software is the "R project for statistical computing", often abbreviated to R. R is a free software environment and programming language for statistical computing and graphics.[82] While the benefits of open-source software in modeling apply to other software packages as well, such as Python, this thesis discusses the use of R for open-source decision-analytic modeling.

Advantages of R

Developing decision-analytic models in R offers several advantages over the currently available special-purpose software programs;

- R is free and therefore available to everyone. This makes it easy to share any analyses performed in R.
- R comes as a blank slate. In R, decision-analytic models do not have to follow any predefined model or analysis structure dictated by the software. This makes R more flexible compared with specialized software programs.
- R comes with a wide variety of statistical techniques. This allows for a smooth integration of primary data analysis and synthesis with the decision-analytic model. This also facilitates the ability to use cutting-edge methods (e.g., model calibration[85] and value of information[86,87]) that rely heavily on statistical and mathematical techniques.
- R has a high processing speed which makes the software valuable for the implementation of computationally demanding analyses.[77]
- The graphical techniques available in R facilitate making well-designed publication-quality graphs, figures and tables to communicate results.
- R is all script based. This makes it possible to order all the steps in an analysis in a sequential way. Reviewing the source code is therefore easier since users can follow the modeling steps in chronological order.
- The functionalities of R can be easily extended via packages. Many of these packages are available on the Comprehensive R Archive Network (CRAN)[88] and GitHub[76], but can be hosted anywhere on the web.[82]
- R has a large community of users that help each other via several platforms like R-blog[89], Reddit90 and Stackoverflow[91].
- R can be accompanied with Rstudio[92], an (often free) integrated development environment for R. Rstudio comes with a markup language to create formatted text, called R-Markdown, which makes it possible to generate comprehensive (model) documentation as a hardcopy or webpage.

Challenges to implement open-source modeling

Despite the many advantages of R, and the first R-based decision-analytic model published already in 2005[93], applications of decision-analytic models in R remain isolated. The flexibility of R, which is a strong benefit, might potentially be one of the contributing factors for its limited uptake. In fact, an R file comes as a blank sheet and leaves coding the model entirely up to the modelers. Many decision-analytic modelers are not statisticians or computer programmers by training, and their limited experience with programming languages may represent the barrier. The lack of clear guidance on how to implement a decision-analytic model in R, and the lack of templates and example models, make the learning curve of R too steep for many decision modelers in the medical field. This thesis aims to outline how R can be used for healthcare decision analysis and to empower modelers to construct R-based decision-analytic models.

Resource prioritization in healthcare

It is essential to identify optimal strategies for prioritization in healthcare to appropriately guide policy makers in the allocation of scarce healthcare resources to improve health. The allocation issue in healthcare exists, because resources are limited[94,95] and the care demand exceeds supply[96,97]. Similar to the well-known healthcare resources money and staff, resources to fund new evidence collection via research are scarce as well. Decision-analytic methods can facilitate prioritization of both healthcare and research resources. The use of R is also an important facilitator because R allows application of recently developed cutting-edge methods in a computationally less expensive way.

Healthcare resources

Healthcare resources include all materials, personnel, facilities, funds, and anything else that is needed to provide healthcare services.[98] The use of resources in one setting, means that they cannot be used in another. Therefore, it is often preferred to use these resources in the most optimal way (i.e., how we can get the "best value for money"). The optimal resource use can be defined as the strategy that maximizes one reward (e.g., most cases prevented or longest expected survival after treatment) or the one that optimizes over multiple parameters such as health outcomes and costs. Decision-analytic models can help to identify the optimal strategy via a comparative analysis of the expected benefit of alternative courses of action.

For comparison reasons, decision-analytic models that evaluate cost-effectiveness often combine resource spending (costs) and monetized health outcomes (effects) in a summary measure called the net monetary benefit.[55] The net monetary benefit illustrates the net financial value of an intervention. In order to monetize health outcomes, a society's willingness to pay for a health unit is needed.[99] This willingness to pay is the amount that society would be willing to pay to achieve an additional unit of health. Under the objective to maximize health for a given budget, the optimal decision is the intervention that generates the highest net monetary benefit on expectation.[100,101]

Although the optimal strategy is identified as the strategy with the highest expected net monetary benefit[100], that does not mean that the distribution of the net monetary benefit values is irrelevant information.[102] It could reflect high decision uncertainty, that is, high changes of making a wrong decision. The uncertainty around the net monetary benefit should be used to determine whether the current evidence is sufficient to determine the optimal intervention or whether more evidence should be acquired in further research before making the decision.[100] Of note, the collection of this additional evidence may be costly, both in terms of research resources as well as forgone health benefits, due to delayed implementation of a potentially valuable intervention.[103]

Research resources

Apart from healthcare resources, research resources are also scarce. Research resources can be defined as all facilities, materials, labor, study participants, funding and all other sources that are required to perform a (clinical) study. Researchers compete to get a share of these limited available resources. Many academic scientists rely on external grants to get their research proposals funded.[104] These researchers are not only competing for funding, they also compete for the participants who can be included in a research study.[105] Therefore, the research resources invested should be justified by the value of the collected evidence.

There is value in collecting further evidence if the expected value of this additional evidence is higher than the expected costs of the research study. We evaluate the value of additional evidence as its potential to reduce the costs associated with making the wrong decision. More specifically, the costs of making the wrong decision, is called the forgone benefit, or potential lost value, and refers to the benefit that could have been gained if a more "optimal" decision had been made.[106]

Value of information

A decision-analytic method called value of information (VOI), can be used to quantify the monetary value of reducing decision uncertainty via additional evidence collection. VOI methods estimate the value of additional evidence collection in terms of its potential to reduce costs associated with current decision uncertainty. VOI takes into consideration both the probability of making the wrong decision based on the current evidence and the potential costs of making that wrong decision. If the value of obtaining additional evidence outweighs the costs associated with collecting that evidence, it is worthwhile to perform more research.

Deciding whether the decision should be made with current evidence or if more data collection is justified should be done in a sequential manner. There are a number of VOI measures that can be used to assess whether current evidence is sufficient to make the decision between implementation or no implementation of the evaluated interventions. Furthermore, it helps determine whether there is value in the collection of further evidence.[102,107–109] If further evidence collection is worthwhile, VOI can identify trial outcomes, that are used as model input parameters, that should be targeted in further research to decrease the current decision uncertainty. Finally, VOI analyses can guide how this evidence should be collected by evaluating different research designs.[86,87,110]

Although there has been increased awareness of VOI in the medical field, it still remains underutilized in research.[111] One of the reasons is that originally VOI calculations were computationally demanding.[112] However, there have been efforts to make these calculations more efficient[112–114] and to develop approximation methods.[115–119] Despite these improvements in methods, decision makers are still not familiar with these methods leading to the limited application of VOI in clinical trail design.[120–124]

Papers discussing VOI methods often provide theoretical explanations expressed in mathematical notations.[35,110,125,126] Clinically oriented readers may not be familiar with mathematical notation and might experience difficulties interpreting these complex equations. However, designing clinical trials using VOI analysis could be very valuable to improve allocation of resources. Therefore, in the second part of this thesis, I aim to communicate the potential value of VOI methods to a broad audience without mathematical notations. In addition, I will present it within a COVID-19 related decision problem.

R for resource prioritization

The use of R-based decision-analytic models is expected to benefit the implementation or research prioritization methods for two main reasons. First, the advantage of R is that both the decision-analytic model and VOI analysis can be integrated together. As discussed before, trial design via VOI is a sequential process. TreeAge can only estimate the first two VOI measures; the expected value of perfect information (EVPI) and the expected value of partial perfect information (EVPPI). These measures tell us what the value of eliminating all uncertainty would be and which parameters are driving the decision uncertainty, respectively. However, TreeAge cannot determine the optimal sample size of a new trial. Although, there are examples of how the optimal sample size of a new trial can be estimated (i.e., via expected value of sample information calculations) in Excel[124,126], these estimations are only feasible for the simplest decision-analytic models.[68] These limitations do not apply to R-based decision-analytic models as they allow for efficient implementation of the recently developed approximation methods.[115–119] Applications of these methods to more practical examples will contribute to a wider understanding of the methods and to further development of these methods.

Second, R-based decision-analytic models can be used to increase transparency in communicating the model and the used methods. Although, the fundamental concept of VOI methods could be relatively easily communicated to a clinical audience[124], their underlying calculations are complex and technical. R could be used to guide users through the levels of complexity. This can start with a

report that only communicates the main results, continues to provide interactive web-applications to test the effect of assumptions all the way down to giving access to the source code for further review and adaption. This stepwise reporting process can contribute to increase trust in the model results by guiding reviewers through the required VOI analysis assumptions like the willingness-to-pay threshold, the number of patients that can benefit from the collected information about the intervention and the period in which the extra evidence about the current intervention is used (e.g., how much time is there between now and the moment the intervention is replaced by a new technology). These assumptions are inherent in research prioritization methods, but could have a major impact on the final recommendation. Therefore, improved resource prioritization requires transparent communication and testing of these assumptions.

Objective of this thesis

The first aim of this thesis is to develop a structured programming method that is generic, uses an open-source programming language, and can be applied to a wide range of research topics. The structured programming method is provided in a set of papers that provide guidance to decision modelers when constructing R-based decision models with the ultimate aim to increase transparency in the field of health decision sciences.

Many decision models focus on evaluating comparative effectiveness or cost-effectiveness. These models aim to identify the optimal treatment, to get the best outcome for a patient or the most cost-effective strategy. However, prioritization in healthcare is more than optimizing patient outcomes or monetary healthcare resources. Both healthcare and research resources contribute to healthcare improvements. The allocation of all resources in healthcare can be more transparent, more evidence-based, and more consistent using decision modeling methods. Therefore, the second aim of this thesis is to demonstrate how open-source decision modeling can contribute to resource prioritization in healthcare.

Outline of this thesis

The thesis consists of two parts. Part I (**Chapters 2-7**) focuses on methods for open-source decision modeling in healthcare and is organized as follows:

Chapter 2 proposes a high-level framework for model-based decision and cost-effectiveness analysis in R. The adoption of this framework will facilitate the sharing and readability of decision models implemented in R. **Chapter 3** provides an overview of existing R functionalities that are applicable to the various stages of decision analysis, including model design, input parameter estimation, and analysis of model outputs. **Chapter 4** addresses how to implement time-*independent* cohort state-transition models in R. **Chapter 5** addresses how to implement time-*dependent* cohort state-transition models, where transition probabilities and rewards vary over time. **Chapter 6** proposes an alternative approach to compute and store cohort state-transition model outcomes. This approach produces a multidimensional array which captures both the state occupancy and the transition dynamics of the model. **Chapter 7** provides a general coding structure that can be used to build individual based state-transition models, also called microsimulation models, in R.

Part II (**Chapters 8-10**) provides guidance and practical examples of how open-source decision modeling methods can contribute to resource prioritization in healthcare.

Chapter 8 demonstrates how a decision-analytic model can be used to support prioritization of surgical care in times of scarce surgical capacity (e.g., during a pandemic) from a utilitarian perspective. This ethical perspective strives to achieve the greatest good for the greatest number. **Chapter 9** discusses two methods for the allocation of healthcare and research resources and reviews a framework for prioritization and design of clinical trials. **Chapter 10** demonstrates how a value-of-information approach can inform decision makers about both treatment decisions and research prioritization decisions regarding COVID-19 therapies in times of a pandemic.

Finally **Chapter 11** provides a general discussion about the main results of the studies presented in this thesis. It also provides recommendations to further optimize the use of open-source decision-analytic models for optimize resource allocation in healthcare.

References

1. Hannigan LJ, Phillippo DM, Hanlon P, Moss L, Butterly EW, Hawkins N, Dias S, Welton NJ, Mcallister DA. Improving the Estimation of Subgroup Effects for Clinical Trial Participants with Multimorbidity by Incorporating Drug Class-Level Information in Bayesian Hierarchical Models : A Simulation Study. *Med Decis Making*. 2021; epub:1–13.
2. Schlemm L, Schlemm E, Nolte CH. Pre-hospital Triage of Acute Ischemic Stroke Patients — Importance of Considering More Than Two Transport Options. *Front Neurol*. 2019; 10(437):1–8.
3. Venema E, Lingsma H, Chalos V, Mulder M, Lahr M, van der Lugt A, van Es A, Steyerberg E, Hunink M, Dippel D, Roozenbeek B. Personalized Prehospital Triage in Acute Ischemic Stroke. *Stroke*. 2019; 50(2):313–320.
4. Street AE, Street DJ, Flynn GM. Who Gets the Last Bed ? A Discrete-Choice Experiment Examining General Population Preferences for Intensive Care Bed Prioritization in a Pandemic. *Med Decis Making*. 2021; 41(4):408–418.
5. Gezondheidsraad. Griepvaccinatie: *herziening van de indicatiestelling* 2021. Den Haag: Gezondheidsraad, 2021; publicatienr. 2021/39.
6. Health Council of the Netherlands. *Phenytoin. Evaluation of the effects on reproduction, recommendation for classification.* The Hague: Health Councile of the Netherlands, 2018; publication no. 2018/15.
7. Health Council of the Netherlands. *New Anticoagulants: A well-dosed introduction*. The Hague: Health Council of the Netherlands, 2012; publicaion no. 2012/07E.
8. Opstal-van Winden AWJ, Kooyker AI, Toes-Zoutendijk E, Buskermolen M, Spaander MCW, Dekker E, van Vuuren HJ, van Kemenade FJ, Ramakers C, Nagtegaal I, van Veldhuizen H, Thomeer MGJ, van Velthuysen M-LF, van Ballegooijen M, Bonfrer HMG, de Koning HJ, Kuipers EJ, van Leerdam ME, Lansdorp-Vogelaar I. *Landelijke monitoring en evaluatie van het bevolkingsonderzoek naar darmkanker in Nederland 2014-2017*. 2019.
9. Zhang W, Mohammadi T, Sou J, Anis AH. Cost-effectiveness of prenatal screening and diagnostic strategies for Down syndrome : A microsimulation modeling analysis. *PLoS One*. 2019; 14(12):1–17.
10. Health Council of the Netherlands. Population Screening Act : NIPT for multiple pregnancies *Executive summary*. Health Council of the Netherlands, 2019; publicaion no. 2019/19.
11. Knudsen AB, Lansdorp-Vogelaar I, Rutter CM, Savarino JE, Van Ballegooijen M, Kuntz KM, Zauber AG. Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the medicare population. *J Natl Cancer Inst*. 2010; 102(16):1238–1252.
12. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. *Decision Making in Health and Medicine. Integrating Evidence and Values*. 2nd ed. Cambridge University Press, 2014.
13. Chhatwal J, Ferrante S, Brass C. Cost-effectiveness of boceprevir in patients previously treated for chronic hepatitis C genotype 1 infection in the United States. *Value Health*., http://www.sciencedirect.com/science/article/pii/S1098301513019049 (2013, accessed February 4, 2016).
14. Welton NJ, McAleenan A, Thom HHZ, Davies P, Hollingworth W, Higgins JPT, Okoli G, Sterne JAC, Feder G, Eaton D, Hingorani A, Fawsitt C, Lobban T, Bryden P, Richards A, Sofat R. Screening strategies for atrial fibrillation: A systematic review and cost-effectiveness analysis. *Health Technol Assess* (Rockv). 2017; 21(29):1–235.
15. Hoogendoorn M, Feenstra TL, Asukai Y, Briggs AH, Borg S, Dal Negro RW, Hansen RN, Jansson SA, Leidl R, Risebrough N, Samyshkin Y, Wacker ME, Rutten-van Mölken MPMH. Patient Heterogeneity in Health Economic Decision Models for Chronic Obstructive Pulmonary Disease: Are Current Models Suitable to Evaluate Personalized Medicine? *Value Health*. 2016; 19(6):800–810.
16. Jongeneel G, Greuter MJE, van Erning FN, Koopman M, Medema JP, Kandimalla R, Goel A, Bujanda L, Meijer GA, Fijneman RJA, van Oijen MGH, Ijzermans J, Punt CJA, Vink GR, Coupé VMH. Modeling Personalized Adjuvant TreaTment in EaRly stage coloN cancer (PATTERN). *Eur J Heal Econ*. 2020; 21(7):1059–1073.
17. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost-effectiveness subgroup analysis: Conceptual framework and application. *Med Decis Making*. 2014; 34(8):951–964.
18. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012; 32(5):678–689.
19. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials* 2017; 18(1):e122.
20. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun*. 2018; 11(August):156–164.
21. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, Sharples L. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. *Med Decis Making*. 2017; 37(4):377–390.
22. NIH U.S. National Library of Medicine. Clinicaltrials.gov. *Visited May 2020*, Clinicaltrials.org.
23. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the Generalizability of Randomized Trial Results to Target Populations. *Prev Sci*. 2015; 16(3):475–485.
24. Geller SE, Koch A, Pellettieri B, Carnes M. Inclusion, analysis, and reporting of sex and race/ethnicity in clinical trials: have we made progress? *J women's Heal*. 2011; 20(3):315–320.
25. Ramamoorthy A, Pacanowski MA, Bull J, Zhang L. Racial/ethnic differences in drug disposition and response: Review of recently approved drugs. *Clinical Pharmacology and Therapeutics* 2015; 97(3):263–273.

26. Heinrich J. GAO-01-286R Drugs Withdrawn From Market - Drug Safety: *Most Drugs withdrawn in recent years had greater health risks for women*, http://www.fda.gov/cder/drug/infopage/lotronex/lotronex-qa.htm (2001).
27. Caplan A, Friesen P. Health disparities and clinical trial recruitment: Is there a duty to tweet? *PLoS Biol*. 2017; 15(3):e2002040.
28. Sharpe N. Clinical trials and the real world: Selection bias and generalisability of trial results. *Cardiovasc Drugs Ther*. 2002; 16(1):75–77.
29. Latimer NR. Survival analysis for economic evaluations alongside clinical trials - Extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013; 33(6):743–754.
30. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making*. 2017; 37(4):340–352.
31. The SPRINT Research Group. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*. 2015; 373(22):2103–2116.
32. Bress AP, Bellows BK, King JB, Hess R, Beddhu S, Zhang Z, Berlowitz DR, Conroy MB, Fine L, Oparil S, Morisky DE, Kazis LE, Ruiz-Negrón N, Powell J, Tamariz L, Whittle J, Wright JT, Supiano MA, Cheung AK, et al. Cost-Effectiveness of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*. 2017; 377(8):745–755.
33. Jit M, Brisson M, Portnoy A, Hutubessy R. Cost-effectiveness of female human papillomavirus vaccination in 179 countries: A PRIME modelling study. *Lancet Glob Heal*. 2014; 2(7):e406–e414.
34. Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. *PLoS Med*. 2015; 12(9):1–5.
35. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996; 5(6):513–524.
36. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, Spackman E, Rothery C. A Comprehensive Algorithm for Approval of Health Technologies With, Without, or Only in Research: The Key Principles for Informing Coverage Decisions. *Value Health*. 2016; 19(6):885–891.
37. Garattini L, Padula A. HTA for pharmaceuticals in Europe : will the mountain deliver a mouse ? *Eur J Heal Econ*. 2020; 21(1):1–5.
38. Canadian Agency for Drugs and Technologies in Health. *CADTH Methods and Guidelines. Guidelines for the Economic Evaluation of Health Technologies: Canada,* https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf (2017).
39. Pershing S, Enns EA, Matesic B, Owens DK, Goldhaber-Fiebert JD. Cost-Effectiveness of Treatment of *Diabetic Macular Edema*, www.annals.org (2014).
40. Sathianathen NJ, Konety BR, Alarid-Escudero F, Lawrentschuk N, Bolton DM, Kuntz KM. Cost-effectiveness Analysis of Active Surveillance Strategies for Men with Low-risk Prostate Cancer(Figure presented.). *Eur Urol*. 2019; 75(6):910–917.
41. Lu S, Yu Y, Fu S, Ren H. Cost-effectiveness of ALK testing and first-line crizotinib therapy for non-small-cell lung cancer in China. *PLoS One.*; 13(10). Epub ahead of print October 1, 2018. DOI: 10.1371/journal.pone.0205827.
42. Smith-Spangler CM, Juusola JL, Enns EA, Owens DK, Garber AM. Population Strategies to Decrease Sodium Intake and the Burden of Cardiovascular Disease A Cost-Effectiveness Analysis. *Ann Intern Med*. 2010; 152(8):481–487.
43. Andrus JK, Toscano CM, Lewis M, Oliveiria L, Ropero AM, Dávila M, Fitzsimmons JW. A Model for Enhancing Evidence-Based Capacity to Make Informed Policy Decisions on the Introduction of New Vaccines in the Americas: Paho's Provac Initiative. 2007; 122(6):811–816.
44. Koleva D, de Compadri P, Padula A, Garattini L. Economic evaluation of human papilloma virus vaccination in the European Union: A critical review. *Internal and Emergency Medicine* 2011; 6(2):163–174.
45. Kim S-Y, Goldie SJ. Cost-Effectiveness Analyses of Vaccination Programmes A Focused Review of Modelling Approaches. *Pharmacoeconomics*. 2008; 26(3):191–215.
46. World Health Organization. Countries with National agency / unit / committee that produces HTA reports for the Ministry of Health, https://www.who.int/health-technology-assessment/NationalAgencieHTA.pdf (accessed December 16, 2021).
47. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on Medical Decision Analysis: Part 2-Building a Tree. *Med Decis Making*. 1997; 17(2):123–125.
48. 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.
49. 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. *Med Decis Making*. 2012; 32(5):690–700.
50. Vázquez-Serrano JI, Peimbert-García RE, Cárdenas-Barrón LE. Discrete-Event Simulation Modeling in Healthcare : A Comprehensive Review. *Int Joural Environ Res Public Heal*. 2021; 18(22):12262.
51. Glover MJ, Jones E, Masconi KL, Sweeting MJ, Thompson SG, Powell JT, Ulug P, Bown MJ. Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening. *Med Decis Making*. 2018; 38(4):439–451.
52. Degeling K, Franken MD, May AM, van Oijen MGH, Koopman M, Punt CJA, IJzerman MJ, Koffijberg H. Matching the model with the evidence: comparing discrete event simulation and state-transition modeling for time-to-event predictions in a cost-effectiveness analysis of treatment in metastatic colorectal cancer patients. *Cancer Epidemiol*. 2018; 57:60–67.

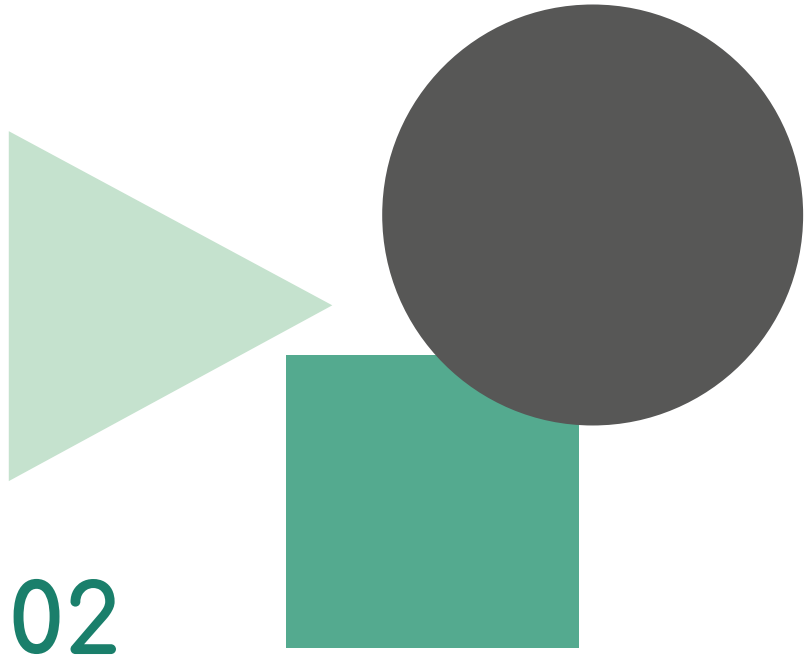
53. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Med Decis Making*. 2012; 32(5):701–711.
54. Martcheva M. Chapter 2: Introduction to Epidemic Modeling" in An Introduction to Mathematical Epidemiology. In: *An Introduction to Mathematical Epidemiology*. Springer, 2015, pp. 9–31.
55. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006; 15(12):1295–1310.
56. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on Medical Decision Analysis: Part 1 - Getting Started. *Med Decis Making*. 1997; 17(2):123–125.
57. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, ISPOR-SMDM Modeling Good Research Practices Task Force. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012; 32(5):722–732.
58. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. 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.
59. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012; 32(5):678–689.
60. Pitman R, Fisman D, Zaric G, Postma M, Kretzschmar M, Edmunds J, Brisson M, On Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. 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.
61. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. 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.
62. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, Koffijberg H, Strong M, Sanders Schmidler GD, Rothery C. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020; 23(2):139–150.
63. Rothery C, Strong M, Koffijberg H, Basu A, Ghabri S, Knies S, Murray JF, Sanders Schmidler GD, Steuten L, Fenwick E. Value of Information Analytical Methods: Report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020; 23(3):277–286.
64. Zorginstituut Nederland. *Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg*. 2016.
65. The National Institute for Health and Care. *Guide to the methods of technology appraisal*, http://doi.org/10.1183/13993003.01815-2018 (2013).
66. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*. 2013; 16(2):e1-5.
67. Strong M, Oakley JE, Chilcott J. Managing structural uncertainty in health economic decision models: a discrepancy approach. *Jourrnal R* Stat Soc. 2011; 61(1):25–45.
68. Incerti D, Thom H, Baio G, Jansen JP. R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. *Value Health*. 2019; 22(5):575–579.
69. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan T. Probabilistic sensitivity analysis for NICE technology assessment: Not an optional extra. *Health Economics* 2005; 14(4):339–347.
70. National Institute for Health and Care Excellence. *Guide to the processes of technology appraisal*, https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf (2018).
71. Elevier. What is peer review?, https://www.elsevier.com/reviewers/what-is-peer-review (accessed December 12, 2021).
72. Tappenden P, Caro JJ. Improving Transparency in Decision Models: Current Issues and Potential Solutions. *Pharmacoeconomics*. 2019; 37(11):1303–1304.
73. Jürgensen JS, Arns W, Ha B. Cost-effectiveness of immunosuppressive regimens in renal transplant recipients in Germany: A model approach. *Eur J Heal Econ*. 2010; 11(1):15–25.
74. Global Health CEA Registry, http://ghcearegistry.org/ghcearegistry/ (accessed October 18, 2021).
75. Open-Source Model Clearinghouse, http://ghcearegistry.com/orchard/about-the-clearinghouse (accessed December 12, 2021).
76. Github. GitHub, https://github.com/ (2020).
77. Hollman C, Paulden M, Pechlivanoglou P, McCabe C. A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. *Pharmacoeconomics*. 2017; 35(8):817–830.
78. TreeAge Software, Williamstown, MA; software available at http://www.treeage.com.
79. Software T. *TreeAge Pro Healthcare 2020 User ' s Manual*. 2020.
80. Microsoft Corporation. Microsoft Excel, https://office.microsoft.com/excel.

81. Baio G, Heath A. When Simple Becomes Complicated: Why Excel Should Lose its Place at the Top Table. *Glob Reg Heal Technol Assess Ital North Eur Spanish*. 2017; 4(1):e3–e6.
82. R Core Team. R: A language and Environment for Statistical Computing, https://www.r-project.org (2013).
83. Foundation PS. *Python Language Reference*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
84. Corbly JE. The free software alternative: Freeware, open-source software, and libraries. *Inf Technol Libr*. 2014; 33(3):65–75.
85. Vanni T, Karnon J, Madan J. Calibrating models in economic evaluation. *Pharmacoeconomics*. 2011; 29(1):35–49.
86. Ades AE, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Med Decis Making*. 2004; 24(2):207–227.
87. Griffin S, Welton NJ, Claxton K. Exploring the research decision space: the expected value of information for sequential research designs. *Med Decis Making*. 2010; 30(2):155–62.
88. The Comprehensive R Archive Network, https://cran.r-project.org.
89. R project. R blog, https://developer.r-project.org/Blog/public/ (accessed December 5, 2021).
90. Reddit, https://www.reddit.com (accessed December 5, 2021).
91. Stack Overflow, https://stackoverflow.com (2008, accessed December 5, 2021).
92. Team R. Rstudio, https://www.rstudio.com.
93. Hawkins N, Schulpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. *Med Decis Making*. 2005; 25(5):511–519.
94. Ministerie van Financien. Rijksbegroting: Volksgezondheid, Welzijn en Sport, https://www.rijksfinancien.nl/visuals/2020/begroting/uitgaven/XVI (accessed December 6, 2021).
95. National Health Service England. *NHS England and NHS Improvement funding and resource 2019/20: supporting 'The NHS Long Term Plan,'* https://www.england.nhs.uk/publication/funding-and-resource-2019-20/ (2019).
96. Zorgkaart Nederland Patientenfederatie Nederland. Wachttijden, https://www.zorgkaartnederland.nl/wachttijden (accessed December 5, 2021).
97. Nederlandse Zorgautoriteit (NZa). Regeling Aanleveren wachttijden medisch-specialistische zorg - NR/REG-2127, https://puc.overheid.nl/nza/doc/PUC_642745_22/1/ (accessed December 5, 2021).
98. Contsys.org. healthcare resource, https://contsys.org/concept/healthcare_resource.
99. Neumann PJ, Weinstein MC. Legislating against use of cost-effectiveness information. *N Engl J Med*. 2010; 363(16):1495–1497.
100. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. J *Health Econ*. 1999; 18(3):341–364.
101. Fenwick E, Claxton K, Sculpher M, Briggs A. *Improving the efficiency and relevance of health technology assessment: the role of iterative decision analytic modelling*. York, http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE Discussion Paper 179.pdf (2000).
102. Griffin SC, Claxton KP, Palmer SJ, Schulpher MJ. Dangerous Omissions: The consequences of ignoring decision uncertainty. *Health Econ*. 2011; 20(2):212–224.
103. Weinstein MC, Toy EL, Sandberg EA, Neumann PJ, Evans JS, Kuntz KM, Graham JD, Hammitt JK, E F, K C, M S, A. B, Weinstein MC, Toy EL, Sandberg EA, Neumann PJ, Evans JS, Kuntz KM, Graham JD, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health*. 2001; 4(5):348–361.
104. Health NI of. NIH Data Book - Success Rates: R01-Equivalent and Research Project Grants, https://report.nih.gov/nihdatabook/category/10 (accessed November 12, 2021).
105. Kasenda B, Von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, Meerpohl JJ, Stegert M, Tikkinen KAO, Neumann I, Carrasco-Labra A, Faulhaber M, Mulla SM, Mertz D, Akl EA, Bassler D, Busse JW, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA - J Am *Med Assoc*. 2014; 311(10):1045–1051.
106. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. "Time Traveling Is Just Too Dangerous" But Some Methods Are Worth Revisiting: The Advantage of Expected Loss Curves Over Cost-Effectiveness Acceptability Curves and Frontier. *Value Healt*h. 2018; 22(5):611–618.
107. Walker S, Sculpher M, Claxton K, Palmer S. Coverage with Evidence Development, Only in research, Risk sharing, or Patient Access Scheme? A Framework for Coverage Decisions. *Value Health*. 2012; 15(3):570–579.
108. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: Combined and uneven development. *Med Decis Making*. 2008; 28(1):21–32.
109. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Makin*g. 1998; 18(1):95–109.
110. Conti S, Claxton K. Dimensions of design space: a decision-theoretic approach to optimal research design. *Med Decis Making*. 2009; 29(6):643–660.
111. Keisler JM, Collier ZA, Chu E, Sinatra N, Linkov I. Value of information analysis: The state of application. *Environ Syst Decis*. 2014; 34(1):3–23.
112. Sadatsafavi M, Bansback N, Zafari Z, Najafzadeh M, Marra C. Need for speed: an efficient algorithm for calculation of single-parameter expected value of partial perfect information. *Value Health*. 2013; 16(2):438–448.

113. Brennan A, Kharroubi S, O'hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med Decis Makin*g. 2007; 27(4):448–470.
114. Madan J, Ades AE, Price M, Maitland K, Jemutai J, Revill P, Welton NJ. Strategies for efficient computation of the expected value of partial perfect information. *Med Decis Making*. 2014; 34(3):327–342.
115. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast, Nonparametric Regression-Based Method. *Med Decis Making*. 2015; 35(5):570–583.
116. Menzies NA. An Efficient Estimator for the Expected Value of Sample Information. *Med Decis Making*. 2016; 36(3):308–320.
117. Jalal H, Alarid-Escudero F. A Gaussian Approximation Approach for Value of Information Analysis. *Med Decis Making*. 2018; 38(2):174–188.
118. Jalal H, Goldhaber-Fiebert JD, Kuntz KM. Computing expected value of partial sample information from probabilistic sensitivity analysis using linear regression metamodeling. *Med Decis Making*. 2015; 35(5):584–595.
119. Heath A, Manolopoulou I, Baio G. Efficient Monte Carlo Estimation of the Expected Value of Sample Information Using Moment Matching. *Med Decis Making*. 2018; 38(2):163–173.
120. Hassan C, Hunink MGM, Laghi A, Pickhardt PJ, Zullo A, Kim DH, Iafrate F, Di Giulio E. Value-of-information analysis to guide future research in colorectal cancer screening. *Radiology*. 2009; 253(3):745–752.
121. Grutters JPC, Pijls-Johannesma M, Ruysscher D De, Peeters A, Reimoser S, Severens JL, Lambin P, Joore MA. The cost-effectiveness of particle therapy in non-small cell lung cancer: Exploring decision uncertainty and areas for future research. *Cancer Treat Rev*. 2010; 36(6):468–476.
122. Miquel-Cases A, Retèl VP, van Harten WH, Steuten LMG. Decisions on Further Research for Predictive Biomarkers of High-Dose Alkylating Chemotherapy in Triple-Negative Breast Cancer: A Value of Information Analysis. *Value Health*. 2016; 19(4):419–430.
123. Wong WB, Ramsey SD, Barlow WE, Garrison LP, Veenstra DL. The value of comparative effectiveness research: Projected return on investment of the RxPONDER trial (SWOG S1007). *Contemp Clin Trials*. 2012; 33(6):1117–1123.
124. Groot Koerkamp B, Spronk S, Stijnen T, Hunink MGM. Value of information analyses of economic randomized controlled trials: the treatment of intermittent claudication. *Value Health*. 2010; 13(2):242–250.
125. McKenna C, Claxton K. Addressing Adoption and Research Design Decisions Simultaneously The Role of Value of Sample Information Analysis. *Med Decis Making*. 2011; 31(6):853–865.
126. Wilson ECF. A practical guide to value of information analysis. *Pharmacoeconomics*. 2015; 33(2):105–121.

PART I

OPEN-SOURCE DECISION MODELING IN HEALTHCARE



02

A NEED FOR CHANGE! A CODING FRAMEWORK FOR IMPROVING TRANSPARENCY IN DECISION MODELING

Fernando Alarid-Escudero
Eline Krijkamp
Petros Pechlivanoglou
Hawre Jalal
Szu-Yu Zoe Kao
Alan Yang
Eva A. Enns

Pharmacoeconomics 2019

“What will not be simple, simply will not be”

Professor Joseph Pliskin during NIHES course Operations Management

A NEED FOR CHANGE! A CODING FRAMEWORK FOR IMPROVING TRANSPARENCY IN DECISION MODELING

Abstract

The use of open-source programming languages, such as R, in health decision sciences is growing and has the potential to facilitate model transparency, reproducibility, and shareability. However, realizing this potential can be challenging. Models are complex and primarily built to answer a research question, with model-sharing and transparency relegated to being secondary goals. Consequently, code is often neither well-documented nor systematically organized in a comprehensible and shareable approach. Moreover, many decision modelers are not formally trained in computer programming and may lack good coding practices, further compounding the problem of model transparency. To address these challenges, we propose a high-level framework for model-based decision and cost-effectiveness analyses (CEA) in R. The proposed framework consists of a conceptual, modular structure and coding recommendations for the implementation of model-based decision analyses in R. This framework defines a set of common decision model elements divided into five components: (1) model inputs, (2)

decision model implementation, (3) model calibration, (4) model validation, and (5) analysis. The first four components form the model development phase. The analysis component is the application of the fully developed decision model to answer the policy or the research question of interest, assess decision uncertainty, and/or to determine the value of future research through value of information (VOI) analysis. In this framework, we also make recommendations for good coding practices specific to decision modeling, such as file organization and variable naming conventions. We showcase the framework through a fully functional, testbed decision model, which is hosted on GitHub for free download and easy adaptation to other applications. The use of this framework in decision modeling will improve code readability and model sharing, paving the way to an ideal, open-source world.

Introduction

Many journals now strongly encourage that the data and the code underlying an analysis be archived and made publicly available alongside the publication.[1,2] There are similar calls for making mathematical models that are the basis for health technology assessments (HTA) and cost-effectiveness analyses (CEAs) available to promote transparency, support reproducibility, and facilitate adaptation of existing models to new applications.[3,4] In formal HTA submissions, it is already expected that the model itself will be provided to clients and stakeholders for them to scrutinize and manipulate, necessitating a certain degree of model transparency and usability.[5,6] More broadly, the Open-Source Model Clearinghouse was recently launched as a database of open source models with a mandate to, in part, "facilitate adherence to standards calling for open disclosure of scientific software".[7] Though it has been common wisdom that a detailed methods section and a lengthy appendix of equations should be sufficient to reproduce a mathematical model, this is not generally the case. Thus, to support the transparency of mathematical modeling, more and more emphasis is being placed on sharing the underlying model construction, be it implemented in a specific software platform or coded in a programming language.[8–10]

For anyone who has ever looked under the hood of software source code, the naivety of transparency being achieved by sharing such code is obvious. Even for a well-trained and sophisticated programmer, coding entails a certain amount of personal style and preferences which may or may not be intuitive to the reader. Imagine, then, the even more extreme, yet still common, situation of releasing code that was never intended for public use to the public. If this is done as an after-thought, documentation may be lacking, and the code structure will likely be a byproduct of the complex decision history that it took to arrive at the final model structure rather than a pre-planned organizational structure. All these issues may be further obscured when proprietary software is required to view/operate the model, which may limit access to those with active licenses and installations. The appeal of proprietary software is often in the facilitation of model construction through a graphical interface that allows a user to point-and-click their way through an analysis. However, despite the initial user-friendliness of these software platforms, sharing the model alone still does not necessarily achieve transparency or reproducibility, as any manual point-and-click steps are not captured or recorded.[11]

Health decision science and HTA are fields situated at the intersection of operations research, economics, statistics, medicine, and public health. Computer science and software development are generally not the major foci of decision-analytic training, as models are used to answer specific research questions, not necessarily as general tools for a client user-base. Thus, in order for the benefits of transparent and open model-sharing to be fully realized, guidance is needed on coding best practices as it relates to decision modeling so that these models can be read, scrutinized, and understood by their consumers.

The aim of this paper is to provide a high-level framework that sets a common structure for decision-model building for both model developers and model consumers. The development of this framework is the culmination of the research and pedagogical experiences of The Decision Analysis in R for Technologies in Health (DARTH)

workgroup.[12] The DARTH framework modularizes decision models into a set of core components that are common across CEAs and HTA submissions, regardless of the type of mathematical model used. In this paper, we also provide a number of recommendations specific to decision-modeling applications relating to file organization, variable naming conventions, use of functions and data structures, and unit testing. However, these more detailed recommendations are suggestions only; the primary purpose of the DARTH framework is to outline a high-level organizational structure to code underlying a decision-modeling analysis. Given the diversity of applications and methodological needs of different analyses, we hope that the DARTH framework provides a scaffolding to facilitate readability, usability, and reproducibility of the analysis to others, without overly restricting the kinds of models and analyses that can be implemented in this framework. We showcase the DARTH framework through a fully functional, testbed decision model developed in R[11,13], implemented as an R package (`darthpack`) that is freely available for download via GitHub (https://github.com/DARTH-git/darthpack). The testbed model was designed to serve as a template for organizing and sharing model and analysis source code[14] that can be easily adapted to other applications and enhanced by other decision modelers. The adoption and promotion of this framework will create more readable, and thus more shareable models, paving the way to an open-source culture in health decision sciences.

Methods

Components of a decision mode

The DARTH framework is based on the premise that a comprehensive model-based decision and/or CEA will involve the same high-level model-development analysis components, regardless of the specific structure of the decision model being applied, be it a decision tree, Markov model, stochastic simulation model, and so on. In developing this framework, we strived to create a flexible framework that can successfully organize code relating to a diversity of model types and applications.

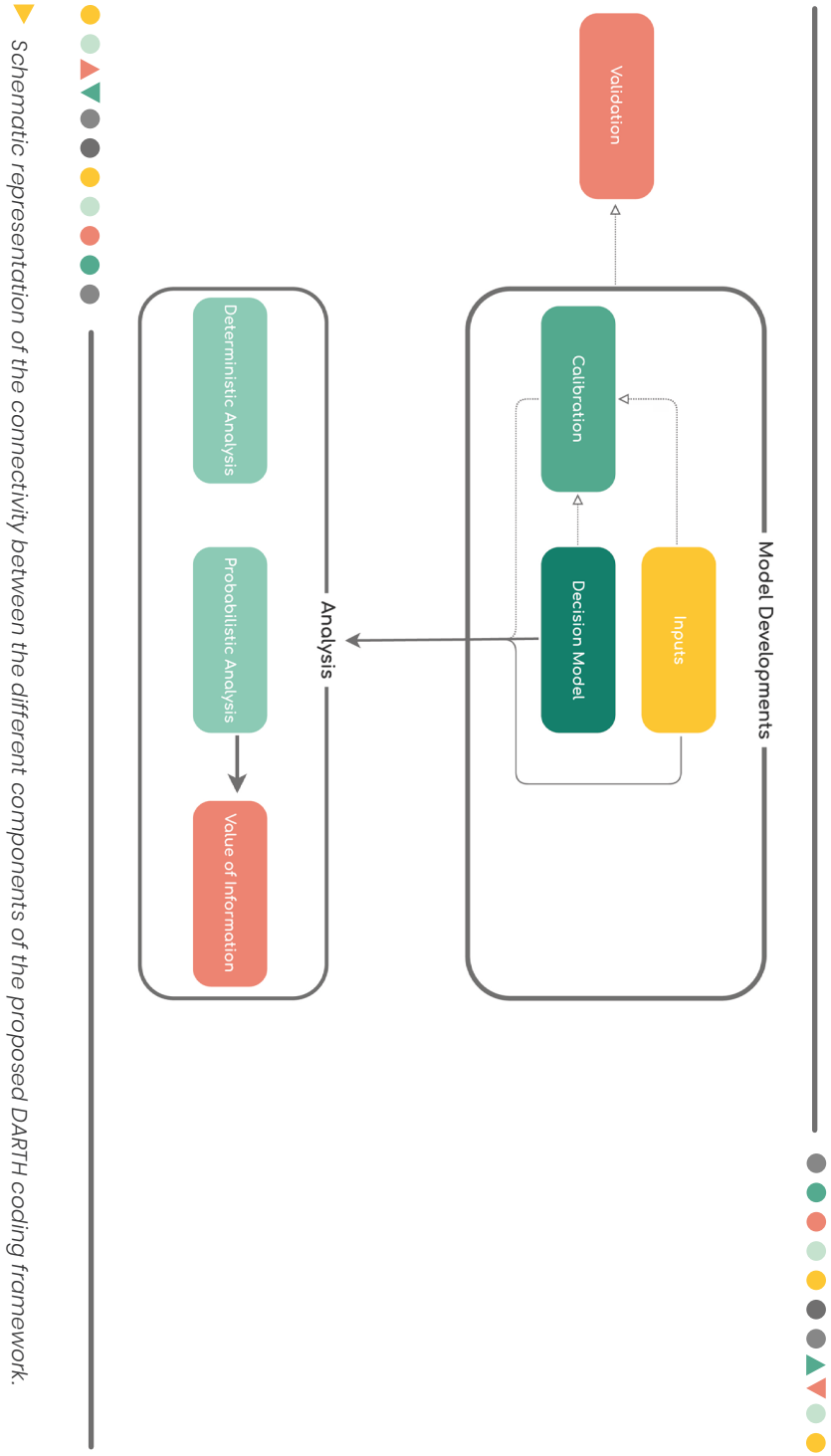
The framework we present here focuses on the organization of R code for the conduct of a decision analysis, but not on the specific content of the code within each component. We also assume that an analyst has already fully documented their biological, behavioral, and mathematical assumptions and decisions that went into their model and analysis in some kind of technical appendix. In our case study, we provide an example of how such documentation might look, but it is not the primary focus of this work. Thus, commentary in the code will primarily explain the functionality of that code, with the assumption that broader descriptions of the disease processes, interventions, and policy questions are provided alongside the code in a separate document.

The DARTH framework divides a decision analysis into five components: (1) model inputs, (2) decision model implementation, (3) model calibration, (4) model validation, and (5) analysis. The first four components form the *model development* phase, whereas the *analysis* component is the application of the final model to answer the policy or research question of interest, assess decision uncertainty, and/or to determine the value of future research through value of information (VOI) analysis. The same model from the development phase could be used to answer multiple research questions, which is why we make this distinction. The relationship between the five components is illustrated in Figure 1 and described in detail in the sections that follow.

Component 1: Define model inputs

In this component, all model input variables are declared and values are set. We broadly categorize input variables into three categories depending on how their values are informed: external, estimated, and calibrated. Parameters informed by external sources are set to a value either directly into an R script or read in from an external source, such as a .csv file or a data repository. These parameter values (and uncertainty ranges and distributions for probabilistic analyses) are derived from published literature or external data analyses not embedded into the analysis itself. Estimated parameters are

Figure 1



Schematic representation of the connectivity between the different components of the proposed DARTH coding framework.

those whose values are estimated through a primary data analysis conducted within the decision analysis. R has the advantage of being both a statistical and programming environment. This allows any necessary statistical analyses to be embedded directly within the decision analysis, further improving analysis transparency and reproducibility. The third type of model parameters are those that will be estimated via model calibration. In this first stage of the DARTH framework, we simply set these parameters to some valid but arbitrary 'dummy' values that are compatible with the next phase of the analysis - model implementation - but are ultimately just placeholder values until we conduct the calibration phase. Not all models will utilize all three types of input variables or different models may rely more heavily on one input type than another. While we selected the three input parameter categories based on how models are typically parameterized, for any given application, it may make sense to organize input parameters according to a different set of categories. The point of this component is to group input variables together and organize them in a logical fashion that can be easily communicated to a user rather than rigidly prescribe a universal input parameter organizational structure.

Component 2: Decision model implementation

This implementation of the decision model component is the heart of the decision analysis. In this section of the DARTH framework, a function that maps model inputs to outputs is created, via the dynamic and/or stochastic processes that the decision model represents. The model itself could be a decision tree, Markov model, stochastic simulation, and so on. The output stored from the model at this stage should be sufficiently general and comprehensive to accommodate calibration, validation, and the main policy analysis. Constructing the model as a function at this stage facilitates subsequent components of model development and analysis, as these processes will all call the same model function but pass different parameter values and/or calculate different final outcomes from the model outputs. The model function also facilitates the use of parallel computing efforts for computationally intensive tasks, such as calibration and probabilistic sensitivity analysis (PSA).

We should note explicitly that the model function created here should have the capacity to capture the effect of any interventions or policy scenarios of interest on the outcomes of interest. The specific ways that intervention effects are incorporated into the model is a choice for the analyst. Interventions that reflect changing intensities of existing processes (e.g., increasing the frequency of screening) may be implemented by changing the values of relevant model input parameters. However, it is often the case that different interventions enable completely different pathways and processes in the model (e.g., medical management vs. surgery) and would be better captured by passing a categorical parameter value that indicates the intervention to be simulated. We do not recommend mixing these two cases and generally recommend using an explicit categorical intervention variable for generalizability. Ultimately, the analyst should decide how best to implement the functionality required for their application.

Component 3: Model calibration

In the model calibration component, unknown or highly uncertain model parameters are estimated by calibrating model outputs to match specified calibration targets.[15–17] This component involves both the setup of the calibration (specification of plausible ranges or prior distributions for input parameters to be calibrated, specification of calibration targets, calculation of corresponding values from model outputs and assessment of fit to targets) as well as the carrying out the calibration itself with a chosen algorithm. Once appropriate values, ranges, and/or distributions have been identified for calibrated parameters, these values will replace the placeholder values established in the model inputs component for the subsequent validation and analysis components. Though rare, not all models will have parameters that need to be calibrated. In such cases, the model calibration component can simply be omitted.

Component 4: Validation

Model validation should at the very least demonstrate the internal validity of the model. This means that the model reproduces outputs that correspond to its inputs.[3,18] For example, if an input parameter to the model was set to reflect a screening frequency of every 2 years, then the number of screenings conducted in the population over a given period of time should correspond to an average per-person screening frequency of every 2 years. Internal validity may also be demonstrated by plotting model-predicted outputs against calibration targets. Additionally, comparison of model outputs to other data sources not used in the model development (external validation) or to other models (comparative validity) may also be conducted here.[19–23]

Component 5: Analysis

The analysis component is where the model developed in components 1-4 is applied to answer the question(s) of interest given current information. An analysis will generally be broken down into several subcomponents. As an example, we describe an analysis with three subcomponents: a probabilistic analysis (which includes the base case analysis), a deterministic scenario and sensitivity analysis (such as one- and two-way sensitivity analyses), and a VOI analysis. However, in any given analysis, the analyst should create subcomponents as is relevant and appropriate for their application. Though the purposes and the structure of our example subcomponents vary, the general setup is similar. First, the analyst must specify the input parameter values that should be passed to the decision model function. Second, the analysis must setup calculation of the desired output values from the decision model outputs, which again are more comprehensive and detailed than may be necessary. For example, a Markov cohort model might output the cohort trace (distribution of the cohort across health states over the time horizon) or the transition dynamics array (proportion of the cohort that transitioned between any two health states in each cycle over the time horizon)[24], but in a given analysis, perhaps only the cohort's survival over time is of interest. Within CEA in particular, there are many standard calculations, comparisons, and

visualizations that are conducted based only on the total costs and quality-adjusted life-years (QALYs) calculated from model outputs for a set of strategies.

Subcomponent 5a: Probabilistic analysis

The probabilistic analysis subcomponent is the primary analysis component in the DARTH framework, which is typical for CEA following the recent guidance from the Second Panel on Cost-Effectiveness in Health and Medicine and satisfying the requirements of many health technology assessment agencies.[6,25] In a probabilistic analysis, also called a probabilistic sensitivity analysis (PSA), sets of input parameter values are randomly sampled from specified distributions. The model is then run for each set of parameter values, producing corresponding model outputs. Using analyst-specified functions that calculate outcomes of interest based on the model function output, means and standard deviations of these outcomes can be calculated from the PSA samples. For CEAs, primary outcomes of interest are generally total discounted costs and QALYs accrued over the analysis time horizon, though other intermediate outcomes may also be of interest. Interventions are then compared by calculating incremental cost-effectiveness ratios (ICERs) based on the expected cost and QALY outcomes from the PSA. We note that in the past, a primary analysis was often conducted using a single, deterministic set of base case parameter values but this practice is no longer recommended.[6] The distributions of outcomes produced from the PSA are also used to produce additional results regarding decision uncertainty, including cost-effectiveness acceptability curves (CEACs) and frontier (CEAF), expected loss curves (ELCs), and others.[26] For these common procedures, we rely on the decision-analytic modeling in R package, `dampack`, which is available for download here: https://github.com/DARTH-git/dampack. Instructions for installing `dampack` are described in the `dampack` GitHub repository.

Subcomponent 5b: Scenario and deterministic sensitivity analysis

The scenario and deterministic sensitivity analysis subcomponent is where the impact of individual or pairs of parameters on model outcomes can be assessed systematically through one- and two-way sensitivity analyses. An analyst may also wish to compare different scenarios (e.g., a high- vs. low cost scenarios), either in a probabilistic or deterministic framing. Generally, these scenario analyses and sensitivity analyses would be secondary to the primary results presented in subcomponent 5a.

Subcomponent 5c: Value of information (VOI) analysis

In the VOI component, we determine whether further potential research is needed using the results from the PSA generated in the probabilistic analysis subcomponent. The most common VOI measures are the expected value of perfect information (EVPI), the expected value of partial perfect information (EVPPI), the expected value of sample information (EVSI)[27,28], and, more recently, the curve of optimal sample size (COSS).[29]

File structure and organization

A model implemented in the R programming environment will involve a series of scripts with the file extension '.R'. The analysis will also generally use and/or generate a number of data and output files, which may be either stored as internal R data files (using '.RData', '.rda', or '.rds' extensions) or as external data files, such as comma-separated-values ('.csv') files. In the suggested organizational file structure of the DARTH framework, we use folders to delineate the different purposes that these files serve in the analysis. Within a folder, we append the relevant component number to the beginning of each file name to indicate where the file will be used or was created (in the case of outputs). Our suggested folder structure is summarized in Table 1. This structure is inspired by the organizational recommendations for an R package[30] and a simple reproducible workflow developed in the field of ecology and evolution.[31]

Table 1

▶ *File folder structure for organizing model and analysis files used in the proposed DARTH coding framework.*

Folder name	Folder function
data-raw	This is where raw data is stored alongside '.R' scripts that read in raw data, process these data, and calls `use_this::use_data(<processed data>)` to save '.rda' formatted data files in the 'data' folder. These data could include a '.csv' file with input parameters derived from the published literature, as well as internal R data files (with '.RData', '.rds', or '.rda' extensions) containing primary data from which model input values will be estimated through statistical models embedded into the analysis.
data	This is where input data is stored to be used in the different components of the CEA. These data could be generated from raw data stored in the 'data-raw' folder. Essentially, this folder stores the cleaned or processed versions of raw data that has been gathered from elsewhere.
R	This is where '.R' files that define functions to be used as part of the analysis are stored. These are functions that are specific to the analysis. The model will be one such function; however, other functions will likely be used, such as computing the fit of the model output to the specific calibration targets of the analysis. This folder also stores '.R' scripts that document the datasets in the 'data' folder.
analysis	This is where interactive scripts of the analysis would be stored. These scripts control the overall flow of the analysis. This is also where many operations that ultimately become functions will be developed and debugged.
output	This is where output files of the analysis should be stored. These files may be internal R data files ('.RData', '.rds', '.rda') or external data files (such as '.csv'). Examples of files stored here would be the output of the model calibration component or the PSA dataset generated in the uncertainty analysis component. These data files can then be loaded by other components without having to first rerun previous components (e.g., the calibrated model values can be loaded for a base case analysis without re-running the calibration).
figs	For analyses that will include figures, we generally create a separate figures folder. Though these could be stored in the output folder, it can be helpful to have a separate folder so that the images of the figure files can be easily previewed. This is particularly important for analyses that generate a large number of figures.
tables	This folder includes tables to be included in a publication or report, such as the table of intervention costs and effects and ICERs.
report	A report folder could be used to store R Markdown files to describe in detail the model-based CEA by using all the functions and data of the framework, run analyses and display figures. The R Markdown files can be compiled into '.html', '.doc' or '.pdf' files to generate a report of the CEA. This report could be the document submitted to HTA agencies accompanying the R code of the model-based CEA.
vignettes	A vignettes folder could be used to describe the usage of the functions and data of each of some or all components of the framework through accompanying R Markdown files as documentation. The R Markdown file can use all the functions, outputs, and figures to integrate the R code into the Markdown text.
tests	A tests folder includes '.R' scripts that runs all the unit tests of the functions in the framework. A good practice is to have one file of tests for each complicated function or for each of the components of the framework.

CEA cost-effectiveness analysis; HTA health technology assessment; ICERs incremental cost-effectiveness ratios; PSA probabilistic sensitivity analysis.

As an example, consider the 'data-raw' folder. The purpose of the data folder is to store the raw data files that will be cleaned, processed and/or analyzed to be used as inputs in the different components. The processed data would then be placed in the 'data' folder, perhaps stored as the file '01_primary_data.RData' to indicate that it will inform model parameter values (the first framework component). Within this folder, we would likely also have a file named '01_inputs.csv', which would contain model parameter values derived externally from published literature. Finally, in addition to input data for the input generation component, an analyst might also have a '03_calibration_targets.RData' which stores the calibration target data that will be used to estimate unknown model parameters through calibration.

Our suggested file folder structure is fairly self-explanatory and certainly customizable. However, two folders that warrant further clarification are the 'R' and 'analysis' folders. The 'R' folder is the traditional directory for storing functions for an R package. Here,

we store a separate '.R' script with all the functions for each framework component as well as some auxiliary '.R' scripts, such as the description of the different data included in the R package. For example, the 'data_init_params.R' script includes the description of the initial set of base-case parameters. To document the functions and processed data to be used as package data, we used `roxygen2`. `roxygen2` is the recommended format to produce documentation for R packages. For a more detailed description on the different components and steps for building an R package, we refer the reader to the R package book by Hadley Wickham.[30] Formalizing operations into functions is especially advantageous for operations that will be repeated (e.g., calculation of total costs and QALYs from model output). A single function can replace multiple lines of code and modularizes operations, and any updates to these repeated operations will be propagated across all function calls. Using functions is considered a good programming practice.[32] Functions that are customized for the particular application, model, and/or analysis should be defined in the '.R' file corresponded to the component where they are first needed. For example, the decision model is implemented as a function, which is important since it will be called by so many other processes (calibration, validation, as well as the analysis components). Model calibration would also involve several custom functions, such as functions to derive outputs corresponding to the calibration targets from the model's more generic, full output and functions to compare those model outputs to the calibration target values in terms of some kind of measure of 'fit'. The analysis components will have many functions for calculating outputs of interests (e.g., aggregating costs and QALYs over a time horizon of interest), and running the model over different sets of input parameter values in deterministic and probabilistic sensitivity analyses.

The 'analysis' folder is the traditional directory for storing the scripts with the code of R-based analyses. In this folder, we store a '.R' script for each framework component. These scripts are the overall control for these processes.

Naming conventions

Within the outlined file structures, we recommend that analysts use a consistent naming convention for variables and files throughout their code that balance readability and brevity. Different well thought out naming conventions have been proposed, including coding styles recommended by the `tidyverse` collection of R packages[33] and the Google R Style Guide.[34] We summarize our own naming convention, tailored to the specific types of parameters and files used in decision analytic modeling, in Table 2. In our naming convention, file names begin with the component number followed by some content descriptor, separated by underscores. User-defined functions are named starting with an action, followed by a descriptor, separated by underscores.

Our variable naming conventions involve encoding certain features of the variable in the name. The suggested naming structure would be `<x>_<y>_<var_name>`, where `x` indicates the data type (e.g., scalar, vector, matrix, data frame, etc.), `y` is the variable type (e.g., probability, rate, relative risk, cost, utility, etc.), and `var_name` is some description of the variable presented separated by underscores. Suggested prefixes are summarized in Table 3.

Unit testing

A full decision-analytic model, complete with all the modules outlined in the DARTH framework, will have complex interdependence between the various functions and processes. It is important to ensure that these functions behave as expected to maintain the integrity of the project. Thus, systematic testing is recommended alongside the development of the decision analysis source code. Testing increases confidence in the results and conclusions of the model and associated analyses and also allows the analyst to quickly identify whether modifications or additions to the analysis code impacts the behavior of the previously developed functions and processes.[35] A widely used testing method is unit testing, which tests a unit of code (often a function or a small process) to verify whether the code executes and generates outputs as intended.

Table 2

File and variable naming conventions in the proposed DARTH coding framework.

Object type	Naming recommendation	Examples
Files	dir/<component number>_<description>.<ext>	• `analysis/01_model_inputs.R` • `R/02_simulation_model_functions.R`
Functions	<action!>_<description>	• `generate_init_params()` • `generate_psa_params()`
Variables	<x>_<y>_<var_name> where x = data type prefix y = variable type prefix var_name = brief descriptor	• `n_samp` • `hr_S1D` • `v_r_mort_by_age` • `a_M` • `l_params_all` • `df_out_ce`

For a comprehensive decision-analytic project, we suggest writing tests alongside the development of any new function or process or whenever a bug is found.[30,35] This practice results in a high level of test coverage of the analysis code, reducing the likelihood that unintended interactions or incompatibilities between functions and/or processes will go undetected. In practice, we suggest that each R script in the 'R' folder have a corresponding testing '.R'script in a separate 'tests' folder. The naming of a test file could begin with 'test_', followed by the file name of the source code that is the target of the testing. This file structure is also compatible with the R package structure.

In each test file, tests are organized by the functions or processes to be tested. A single function or process will likely be associated with multiple tests. For instance, unit testing of a function will involve testing that the function runs when inputs are of the right data type, that the function outputs of the right data type, and that the function outputs are correct in dimension and value for specific sets of input values.

Table 3

Recommended prefixes in variable names that encode data and variable type.

Prefix	Data type	Prefix	Variable type
<> (no prefix)	scalar	n	number
v	vector	p	probability
m	matrix	r	rate
a	array	u	utility
df	data frame	c	cost
dtb	date table	hr	hazard ratio
l	list	rr	relative risk
		ly	life years
		q	QALYs
		se	standard error

QALYs quality-adjusted life years

It is also important to test the error checking within a function, such that the function returns an error when invalid inputs are provided or unexpected results are found .

Comprehensive testing facilitates model sharing, as any downstream user wishing to modify the code can easily verify whether their changes to the original source code requires adjustments to be made to other parts of the code. To illustrate the use of unit testing, we provided examples of unit testing on two selected source code files in our case study using the R package `testthat`.[30,36] We only include a small number of tests so as to not overwhelm those new to testing; however, in practice, a comprehensive set of unit tests should be included.

Additional tools to support model transparency

A number of tools exist that can facilitate the decision modelers interaction with the R language. A useful and commonly used tool is RStudio, an open-source, integrated development environment (IDE) for R. RStudio offers functionality that facilitates R coding (e.g., syntax-highlighting). With RStudio it is possible to create projects, which are files with the '.Rproj' extension. An RStudio project creates a specific R session for the DARTH framework with its own working directory, workspace, history and source documents.[37] In other words, the RStudio project makes a standalone working environment without the trouble of having to specify where files are located when used in different computers. Additional functionality is embedded within the RStudio platform that allows the modeler to present the output of the analysis in a visually attractive and dynamic form. In particular, through the `Shiny` package, an interactive web app[38] can be developed to facilitate the usage of the decision model.[39] The Shiny app allows the user to modify the input parameters, rerun the model through the app's interface and navigate through the updated results. Although Shiny has been developed to support web access to R models, it can also be downloaded and run locally. We have added the 'Shiny_framework.R' file that generates the Shiny app in the GitHub repository, which can be executed locally once the `darthpack` repository is either downloaded or installed. An additional advantage of having the DARTH framework as an R package via `darthpack` is its integration with other packages to develop web applications with JavaScript, such as OpenCPU.

Once the analysis is completed, the user might be interested in generating a report of the findings. R Markdown is a functionality within RStudio that provides a dynamic solution to developing reports within an R environment. Once written, an R Markdown file can be 'knitted' (transcribed) to a variety of different formats (.docx, .pdf, .html). There are a number of advantages associated with the use of R Markdown. The primary one is the integration of the report writing process with the data analysis or the simulation modeling. This allows for a better documented model-based CEA and a dynamic element to the report. For example, a report could be built in R Markdown while allowing for narrative that can be automatically updated conditional on the findings of the analysis. Another advantage of R Markdown is the ease of making a report publicly accessible because the ability for documents to be knitted in different web formats, allows them to be easily published on the web. With R Markdown, the description and reporting of the workflow of a CEA can be made more efficient with limited entry costs for those not already familiar with this functionality. Recently developed packages further enhance the functionality of R Markdown. For example, the `bookdown` facilitates the development of long reports.[40] We provide an example of how a report could be written in R Markdown with `bookdown` by describing the use of the functions of all the components of the DARTH framework using the case-study described below.

Case study: Sick-Sicker model

To showcase the DARTH framework, we performed a CEA of a hypothetical treatment using a state-transition cohort model on a hypothetical disease. For this CEA, we used the previously published Sick-Sicker model first described by Enns et al.[41] Briefly, the Sick-Sicker model simulates a hypothetical cohort of 25-year-old healthy individuals with an age-specific background mortality that are at risk of developing a disease with two different stages of illness, 'Sick' (S1) and 'Sicker' (S2). Individuals in both the S1 and S2 states face an increased mortality and reduced quality of life (QoL) compared to healthy individuals. The hypothetical treatment improves QoL for individuals in the S1 state but has no effect on the QoL of those in the S2 state. While individuals who are afflicted with the illness can be identified through obvious symptoms, those in S1 cannot be easily distinguished from those in the S2 state. Thus, under the treatment strategy, all afflicted individuals are treated and accrue the costs of treatment, even though only those in S1 experience any benefit.

We assume that most parameters of the Sick-Sicker model and their uncertainty are known to the analyst and do not require any statistical estimation. However, because we cannot distinguish

between S1 and S2, neither state-specific mortality hazard ratios nor the probability of progressing from S1 to S2 can be directly estimated. Therefore, we estimated these parameters by calibrating the model to epidemiological data. We internally validated the calibrated model by comparing the predicted outputs from the model, evaluated at the calibrated parameters, against the calibration targets.

As part of the CEA, we conducted different deterministic SA, including one-way and two-way SAs. To quantify the effect of parameter uncertainty on decision uncertainty, we conducted a PSA and reported our uncertainty analysis results with incremental costs and QALYs, ICERs, CEACs, CEAF and ELCs.[26] We also conducted a VOI analysis to determine whether potential future research is needed to reduce parameter uncertainty.

The CEA of the Sick-Sicker model implemented in the DARTH framework may be downloaded from GitHub (https://github.com/DARTH-git/darthpack). We recommend either using the repository of this framework as a GitHub template or installing it as an R package. Using `darthpack` as a template allows users to easily modify any of the included files and is most appropriate for users wishing to adapt the DARTH framework to their own application model and analyses. To use `darthpack` as a template, users should first either clone the repository to their GitHub account or download it locally as a .zip file containing all files and folders. For users simply wishing to reproduce the existing analyses of the Sick-Sicker model in `darthpack` or conduct simple explorations using the included model and/or analysis functions, installing `darthpack` as a package is most appropriate. To install `darthpack` as a package, users should make use of the `devtools` package by typing **`devtools::install_github("DARTH-git/darthpack")`**. Detailed instructions on how to use and install the repository can be found in `darthpack` website (https://darth-git.github.io/darthpack/).

The DARTH framework is divided into different folders, described in Table 1, that could be accessed from the RStudio project 'darthpack.Rproj'. A detailed description on how to install and use the DARTH framework on the Sick-Sicker mode can be found in the `darthpack` GitHub repository (https://github.com/DARTH-git/darthpack) and website (https://darth-git.github.io/darthpack). The framework of the case study is considered a finalized CEA so each of the components in the 'analysis' folder should be able to run independently of the rest of them. For example, if there is interest in reproducing the calibration component, the analyst or reviewer of the CEA can start by running the file '03_calibration.R' in the 'analysis' folder, and so on. To reproduce the entire CEA, including all model development components and all analyses, the analyst should run the '_master.R' file in the 'analysis' folder, which will execute the R scripts of each of the components. For a more detailed description of how the elements (functions, data and procedures) are interconnected within and between components for the Sick-Sicker model CEA case study, we recommend reading the vignettes of `darthpack` stored in the 'vignettes' folder of the repository. In addition, a detailed description of the CEA of Sick-Sicker model can be found in the file 'report.pdf' stored in the 'report/_book' folder and attached as supplementary material to this manuscript. This report could be used as a template for CEA that are submitted to HTA agencies for their approval. These documents describe the code in detail and will guide the reader on how to run code of the Sick-Sicker model implemented in the DARTH framework.

Discussion

We developed the DARTH framework as a way to support transparency, reproducibility, and model sharing in R-based decision analytic models and CEA. Adoption of this general framework will facilitate the sharing and readability of decision analytic models implemented in R as analysts adopting the framework will be familiar with the component structure and the specific choices and assumptions of each component can be easily scrutinized. The standardization of R code presented here may also support the broader use of R in formal HTA submissions, allowing for more complex modeling methods to be more transparently incorporated into decision-making regarding coverage of new health technologies.

As we illustrated in this paper, a traditional model-based decision analysis follows a well-defined conceptual structure. Despite this, our field lacks practical guidance for the implementation of decision modeling in programming languages. The DARTH framework addresses this gap and will facilitate overall improvement in the quality, transparency, and reproducibility of decision models and analyses conducted in R. Frameworks like the one we propose have been adopted in other fields such as engineering, mathematics, and computer science to routinize frequently conducted analyses, leading to improvements in quality and efficiency in these methods.[42] There are additional benefits of using R as the platform to develop model-based CEA. One such benefit is that R has established packages that allow the evaluation of functions in parallel using different cores of computing systems. If components have processes that require the evaluation of the model multiple times (e.g., calibration, validation or PSA), the model evaluations can be carried out more efficiently by parallelizing these processes.

While a standardized framework can facilitate model sharing and readability, it must still be flexible enough to accommodate a wide variety of needs and applications. The framework we describe here is meant merely as the scaffolding for any given analysis; ultimately, the analyst should make design decisions that work for their particular use and that facilitate transparency to their audience, be it clients, stakeholders, a government agency, other academics, or the general public. Alongside the details of the DARTH framework, we have also attempted to provide the rationale behind our recommendations so that analysts may adapt the specific structures and recommendations to their needs while following the spirit of the framework.

The DARTH framework is focused on the structure and organization of the source code underlying a decision model and analysis to support transparency and sharing. The DARTH framework facilitates dissemination by organizing all the code necessary to conduct a given set of analyses into a single directory that can be easily shared via a repository hosting service, such as GitHub, as we have done in our example model, or through open-source initiates, such as the Open-Source Model Clearinghouse.[7] The DARTH framework is built as an R package, which allows the model and analysis source code to be loaded directly into R. For a description of the steps involved in package development, see the R package book by Hadley Wickham[29]. A package has the advantage of generating a self-contained collection of code with explicit dependencies on other packages and versions with a standard downloading and installation process for users. A package is also advantageous if computationally-intensive functions have been compiled from C/C++ source code, as these functions will be available to the user as R functions. The C/C++ source code can be stored in a folder named 'src' as part of an R package.

An R package makes it easier for others to use built-in functions, say for running a model with different input values or exactly reproducing the results of a set of pre-defined analyses. To modify the model structure or adapt it to a new application, the corresponding functions need to be modified and the package must be recompiled. This may be cumbersome in a model-development phase, when debugging and internal validation studies are being conducted. However, if RStudio is used for the package development or adaptation, compiling the

package is an effortless task as long as all the R code is sound and well-implemented. If an analyst truly wants their model to be broadly used by practitioners, tools such as R Shiny can make interacting with models more user-friendly. Documenting the model structure and different components in the CEA using R documentation and R Markdown also enhances the transparency of the decision models and associated analyses. The use of the DARTH framework alongside these complementary dissemination tools are the foundations for open, transparent and reproducible decision modeling, paving the way to an ideal, open-source world.

Supplemental material

Data and statistical code are provided in the GitHub repository (https://github.com/DARTH-git/darthpack) and the `darthpack` website (https://darth-git.github.io/darthpack).

The version of `darthpack` released in this article is available at https://doi.org/10.5281/zenodo.3445451.

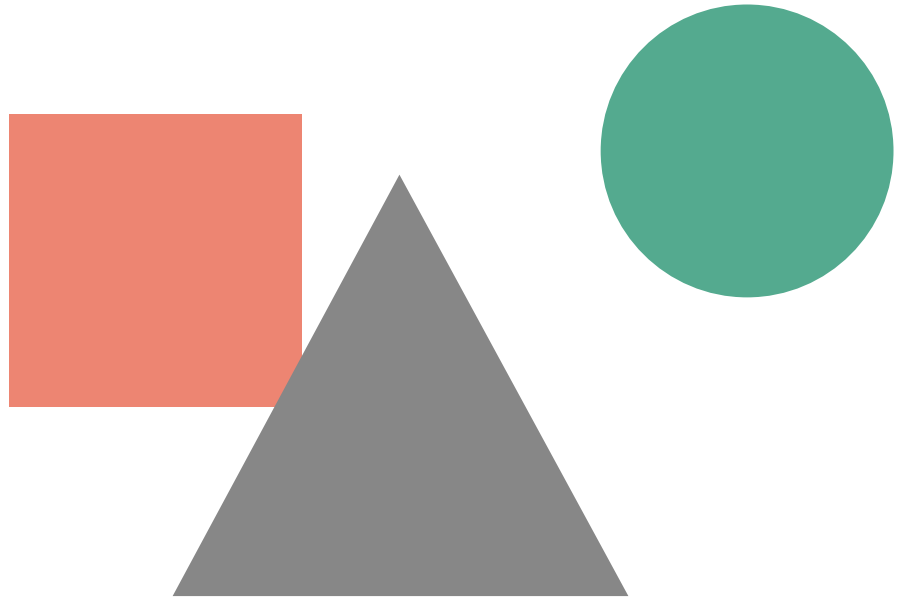
Acknowledgments

We thank Mr. Caleb Easterly for his helpful comments and suggestions on the code developed for this framework and Dr. Myriam Hunink for her overall contribution in the DARTH workgroup.

References

1. Taichman DB. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *Ann Intern Med*. 2017; 14(6):e1002315–e1002315.
2. Stanford. Data Availability Policies at Top Journals, https://web.stanford.edu/~cy10/public/data/Data_Availability_Policies.pdf. (2019, accessed August 2, 2019).
3. 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.
4. Cohen J, Neumann P, Wong J. A Call for Open-Source Cost-Effectiveness Analysis. *Ann Intern Med*. 2017; 167(6):432–433.
5. Baio G, Heath A. When Simple Becomes Complicated: Why Excel Should Lose its Place at the Top Table. *Glob Reg Heal Technol Assess Ital North Eur Spanish*. 2017; 4(1):e3–e6.
6. Canadian Agency for Drugs and Technologies in Health (CADTH). Procedure and Submission Guidelines for the CADTH Common Drug Review. 2018; (August):1–113.
7. Center for the Evaluation of Value and Risk in Health. Open-Source Model Clearinghouse. *Tufts University Medical Center*, http://ghcearegistry.org/ghcearegistry/ (2019, accessed February 1, 2019).
8. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. *Pharmacoeconomics*. 2017; 35(1):125–128.
9. Sampson CJ, Wrightson T. Model Registration : A Call to Action. *PharmacoEconomics - Open*. 2017; 1(2):73–77.
10. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, Hawkins N, Langham S, Marshall D, Sadatsafavi M, Sullivan W, Wilson ECF, Wrightson T. Transparency in Decision Modelling: What, Why, Who and How? *Pharmacoeconomics*. 2019; 37(11):1355–1369.
11. Jalal H, Pechlivanoglou P, Krijkamp EM, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making*. 2017; 37(7):735–746.
12. Decision Analysis in R for Technologies in Health (DARTH) workgroup. Decision Analysis in R for Technologies in Health, http://darthworkgroup.com (2019).
13. R Core Team. R: A language and Environment for Statistical Computing, https://www.r-project.org (2013).
14. Marwick B, Boettiger C, Mullen L. Packaging Data Analytical Work Reproducibly Using R (and Friends). *Am Stat*. 2018; 72(1):80–88.
15. Stout NK, Knudsen AB, Kong CY (Joey), Mcmahon PM, Gazelle GS. Calibration Methods Used in Cancer Simulation Models and Suggested Reporting Guidelines. *Pharmacoeconomics*. 2009; 27(7):533–545.

16. Briggs AH, Weinstein MC, Fenwick E a L, Karnon J, Sculpher MJ, Paltiel a D. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012; 32(5):722–732.
17. Alarid-Escudero F, MacLehose RF, Peralta Y, Kuntz KM, Enns EA. Nonidentifiability in Model Calibration and Implications for Medical Decision Making. *Med Decis Making*. 2018; 38(7):810–821.
18. Sargent RG. Verification and validation of simulation models. *J Simul*. 2013; 7(1):12–24.
19. Goldhaber-Fiebert JD, Stout NK, Goldie SJ. Empirically evaluating decision-analytic models. *Value Health*. 2010; 13(5):667–674.
20. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: Validation and application. *Cancer Epidemiol Biomarkers Prev*. 2010; 19(8):1992–2002.
21. Rutter CM, Knudsen AB, Marsh TL, Doria-Rose VP, Johnson E, Pabiniak C, Kuntz KM, Van Ballegooijen M, Zauber AG, Lansdorp-Vogelaar I. Validation of Models Used to Inform Colorectal Cancer Screening Guidelines. *Med Decis Making*. 2016; 36(5):604–614.
22. Kopec JA, Finès P, Manuel DG, Buckeridge DL, Flanagan WM, Oderkirk J, Abrahamowicz M, Harper S, Sharif B, Okhmatovskaia A, Sayre EC, Rahman MM, Wolfson MC. Validation of population-based disease simulation models: a review of concepts and methods. *BMC Public Health*. 2010; 10(1):710.
23. Cancer Intervention and Surveillance Modelling Network (CISNET). About CISNET, https://cisnet.cancer.gov/about/index.html (2019, accessed July 16, 2019).
24. Krijkamp EM, Alarid-Escudero F, Enns EA, Pechlivanoglou P, Hunink MGM, Yang A, Jalal HJ. A Multidimensional Array Representation of State-Transition Model Dynamics. *Med Decis Making*. 2020; 40(2):242–248.
25. Sculpher MJ, Basu A, Kuntz KM, Meltzer DO. Reflecting Uncertainty in Cost-Effectiveness Analysis. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG (eds) *Cost-Effectiveness in Health and Medicine*. New York, NY: Oxford University Press, 2017, pp. 289–318.
26. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. "Time Traveling Is Just Too Dangerous" But Some Methods Are Worth Revisiting: The Advantage of Expected Loss Curves Over Cost-Effectiveness Acceptability Curves and Frontier. *Value Health*. 2018; 22(5):611–618.
27. Raiffa H, Schlaifer RO. *Applied Statistical Decision Theory*. Cambridge, MA: Harvard Business School, 1961. Epub ahead of print 1961. DOI: 10.1017/CBO9781107415324.004.
28. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996; 5(6):513–524.
29. Jutkowitz E, Alarid-Escudero F, Kuntz KM, Jalal H. The Curve of Optimal Sample Size (COSS): A Graphical Representation of the Optimal Sample Size from a Value of Information Analysis. *Pharmacoeconomics*. 2019; 37(7):871–877.
30. Wickham H. R packages: *Organize, test, document, and share your code*. Sebastopol, CA: O'Reilly Media, 2015. Epub ahead of print 2015. DOI: 10.1017/CBO9781107415324.004.
31. Cooper N, Hsing P-Y, editors. *A Guide to Reproducible Code in Ecology and Evolution*. London, UK: British Ecology Society, 2017.
32. Kleijnen JPC. Verification and validation of simulation models. E*ur J Oper Res*. 1995; 82(1):145–162.
33. Wickham H. The tidyverse style guide, https://style.tidyverse.org (2019, accessed July 19, 2019).
34. Google. Google's R Style Guide, https://google.github.io/styleguide/Rguide.html (2019, accessed July 24, 2021).
35. Martin RC. Clean code: *a handbook of agile software craftsmanship*. Boston, MA: Pearson Education, 2009.
36. Wickham H. testthat: Get Started with Testing. *R J*. 2011; 3(1):5.
37. RStudio. Using projects, https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects (2019, accessed February 1, 2019).
38. Beeley C. *Web Application Development with R using Shiny*. Birmingham, UK: Packt Publishing Ltd, 2013. Epub ahead of print 2013. DOI: 10.1017/CBO9781107415324.004.
39. Incerti D, Curtis JR, Shafrin J, Lakdawalla DN, Jansen JP. A Flexible Open-Source Decision Model for Value Assessment of Biologic Treatment for Rheumatoid Arthritis. *Pharmacoeconomics*. 2019; 37(6):829–843.
40. Xie Y. Bookdown: *Authoring Books with R Markdown*. Boca Raton, FL: CRC Press, 2016.
41. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying best-fitting inputs in health-economic model calibration: A pareto frontier approach. *Med Decis Making*. 2015; 35(2):170–182.
42. David O, Ascough JC, Lloyd W, Green TR, Rojas KW, Leavesley GH, Ahuja LR. A software engineering perspective on environmental modeling framework design: The Object Modeling System. *Environ Model Softw*. 2013; 39:201–213.



03

AN OVERVIEW OF R IN HEALTH DECISION SCIENCES

Hawre Jalal
Petros Pechlivanoglou
Eline Krijkamp
Fernando Alarid-Escudero
Eva A. Enns
M.G. Myriam Hunink

Medical Decision Making 2017

“You don’t have to be first to be original. You have to be different and better”

Adam Grant in Originals

AN OVERVIEW OF R IN HEALTH DECISION SCIENCES

Abstract

As the complexity of health decision science applications increases, high-level programming languages are increasingly adopted for statistical analyses and numerical computations. These programming languages facilitate sophisticated modeling, model documentation, and analysis reproducibility. Among the high-level programming languages, the statistical programming framework R is gaining increased recognition. R is freely available, cross-platform compatible, and open source. A large community of users who have generated an extensive collection of well-documented packages and functions supports it. These functions facilitate applications of health decision science methodology as well as the visualization and communication of results. Although R’s popularity is increasing among health decision scientists, methodological extensions of R in the field of decision analysis remain isolated. The purpose of this article is to provide an overview of existing R functionality that is applicable to the various stages of decision analysis, including model design, input parameter estimation, and analysis of model outputs.

Introduction

Decision analyses often involve building mathematical simulation models to simplify real-life complexity. Applications of decision analysis in health have predominantly been conducted with software that mostly use a "point-and-click" approach (e.g., TreeAge, TreeAge Software, Inc., Williamstown, MA) or "hybrid" software that have interfaces and to some extent support programming languages such as Microsoft Excel (Microsoft Corportaton, Redmond, WA), Corporation, Redmond, WA), which allows programming using Visual Basic for Application code. To a lesser extent, high-level computer-programming languages (e.g., C++) have been used for computationally expensive models.[1] However, as the field of health decision sciences evolves, highly sophisticated models are increasingly used to better represent real-life complexity. In addition, new methodological techniques are being developed rapidly that rely heavily on computationally intensive statistical and mathematical methods (e.g., expected value of information [VOI] analysis). Thus, health decision scientists are beginning to adopt high-level computer programming languages for model development and analysis that also support statistical and mathematical functionalities, such as R (R Core Development Team). Developing models and conducting statistical analyses in a programming language offers several advantages over the currently available special-purpose software. First, programming languages offer more extensive integration with complex statistical approaches and are more flexible compared with special purpose software. In addition, well-conducted analyses that rely on a programming language can be more easily reproduced compared with software that does not reveal the underlying programming code. Furthermore, programming languages allow for clear documentation and model transparency using inline text that explains the purpose of each line, custom function, or section of the code. Thus, modelers generally have more control for model debugging in programming languages.

Several different high-level programming languages, such as C++, Java (Oracle Corporation, Redwood Shores, CA), Python, MATLAB (The MathWorks, Inc., Natick, MA), and R (R Core Development Team) are currently used in health decision sciences. Among these languages, R is uniquely positioned because it is freely available, supported by a large community of professionals and academicians, and capable of integrating many features needed in the process of decision analysis. In addition, there are numerous resources on tutorials, guidelines, and good coding practices and programming styles that help make the code easier to read and share.

R is a programming language and an environment for conducting statistical analyses. R is freely available as part of the GNU project and is largely based on the S language. Version 4 of S program (S4) provides advanced object-oriented programming, which is particularly useful for building microsimulation models. In addition, advanced users can integrate their programs in C++ for increased computational efficiency. These features make R particularly suitable for designing and conducting simulation analysis.

R has a basic set of packages (R core packages) that provide a substantial collection of useful functions for health decision sciences. This core functionality has been readily extended by its active community of users who have generated many well documented, user-written packages that implement commonly used statistical and computational tasks. R is maintained by the R Core Development Team, and currently there are more than 2000 packages that have been contributed to the Comprehensive R Archive Network (CRAN), which is a network of file transfer protocol and Web servers around the world that store identical and up-to-date versions of code and documentation for R to avoid conflicting versions.

Despite the many advantages of R and its increasing popularity among health decision analysts,[2,3] applications of R remain isolated, and a general familiarity with the capabilities of R in the health decision sciences is lacking. The purpose of this article is to provide an overview of R for various applications in health decision analysis, including 1) model design and implementation, 2) input parameter estimation and calibration, and 3) summariza tion of results and sensitivity analyses.

In this article, we first search the literature for applications in health decision sciences that relied on R compared with a sample of other commonly used software. Next, we discuss specific R packages and functions that are of relevance to health decision sciences, grouped by stages of model development and analysis (Appendix A). This article is the first in a series of tutorial papers on the use of R in health decision sciences.

Software used in health decision sciences

We conducted a literature search to measure how frequently R is used in health decision sciences compared with other commonly used software and to identify potential trends in its use over time. We considered the use of this software at any stage of modeling development (e.g., parameter estimation, data analysis, mathematical model building, VOI, etc.). We searched full-text articles published in *Medical Decision Making, Pharmacoeconomics, Health Economics, Journal of Health Economics, European Journal of Health Economics, Value in Health, Health Economics Review*, and *Statistics in Medicine*. We used Google Scholar, which allows for full-text search within specific journals. Because Google Scholar limits the number of characters in the search string, we limited our search query to ([software name]) AND ("*cost-effectiveness*" OR "*cost-utility*" OR "*QALY*" OR "*Decision model*" OR "*microsimulation*" OR "*value of information*"). We modified the "software name" keyword to capture applications in health decision sciences that used a list of relevant software, including R, Microsoft Excel (Microsoft Corporation), TreeAge (TreeAge Software, Inc.), SAS (SAS Software, Cary, NC), and STATA (Stata Statistical Software, College Station, TX). Since R is a single letter, we used these keywords to search for R-relevant mentioning in Google Scholar: ("*Used R*" OR "*R Core*" OR "*R software*" OR "*R development*" OR "*R statistical*" OR "*R package*" OR "*R Project*") AND ("*cost-effectiveness*" OR "*cost-utility*" OR "*QALY*" OR "*Decision model*" OR "*microsimulation*" OR "*value of information*").

We manually reviewed the entries and excluded all conference abstracts and posters and limited the search to the period between 1 January 1994 and 31 December 2015. We decided to start our trend analysis in 1994 as R was created in 1993. We extracted the PubMed identifiers from the relevant abstracts. All analyses, including the literature search, were conducted in R, and we used the R package `RISmed` to analyze the findings of the literature search (Figure 1).

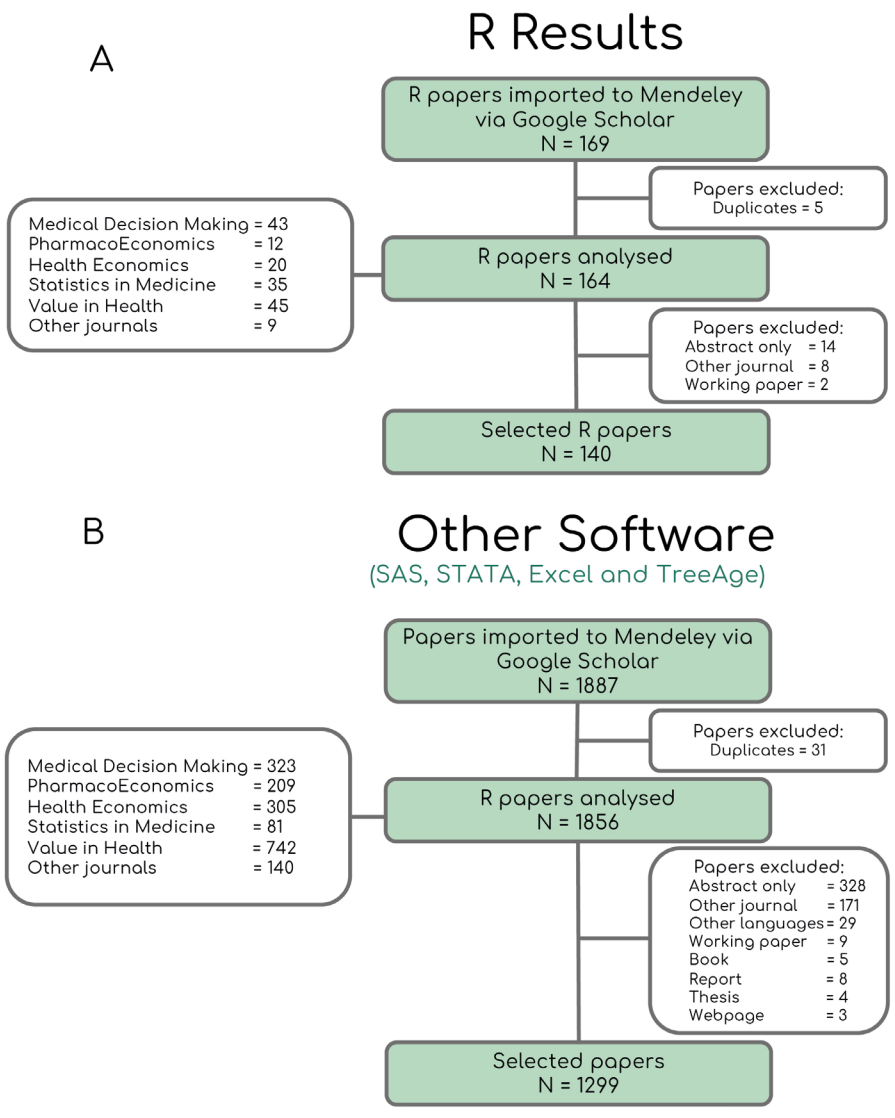
Figure 2 illustrates the trends of various software use in health decision sciences over time. R's diffusion into the health decision sciences literature was measured through the proportion of studies per year that were conducted using R versus any other software. In the journals searched, R appears to increase in popularity, with the proportion of studies using R increasing by nearly 50% over the past 5 years. In fact, in these journals, R is the fastest growing software compared with the other software. These results seem consistent with other disciplines. (http://r4stats.com/articles/popularity/).

Existing R packages and functions useful in health decision sciences

In this section, we provide a list of R packages and user-defined functions that we found relevant to health decision science. We divided these packages into 3 categories based on 3 core stages of decision analysis: 1) model design, 2) input parameter estimation and calibration, and 3) analysis of model outputs, presentation of results, and sensitivity analysis. The packages and their sources are summarized in Appendix A.

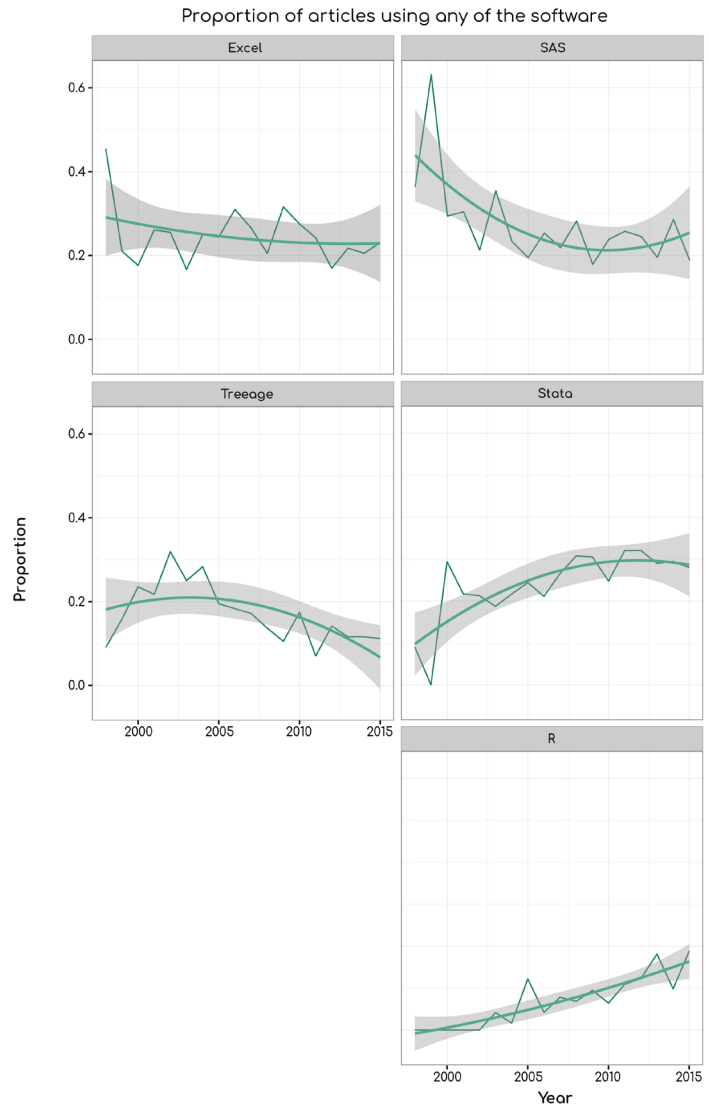
In addition, CRAN has created the R Task Views (https://cran.r-project.org/web/views/) that group sets of R packages and functions by type of analysis, fields, or methodologies. For example, some task views included in this repository are the optimization and mathematical programming and Bayesian inference that are especially useful for calibration and model parameter estimation.

Figure 1



Flow diagram for the literature search. (a) Flow chart of the literature search for articles that used R. (b) Chart of the literature search for articles that used other software, such as SAS, STATA, Microsoft Office Excel, or TreeAge.

Figure 2



Proportion of articles in health decision sciences using the identified software. For comparing R to the other software, we chose to start the x-axis from 1998 because there were only 4 articles that used R in the period between 1994 and 1998.

Model design

Models in health decision sciences range from simple decision trees to complex agent-based micro-simulation and infectious disease models. Below, we provide examples of specialized functions that facilitate the implementation of the commonly used models in health decision sciences as described by the SMDM-ISPOR Joint Task Forces[4–10]:

Decision trees are the simplest form of decision models in which all possible alternatives and the pathways associated with each alternative are represented in a treelike graph structure. The analytical computations underlying a decision tree are mostly simple enough to be conducted in R using the R core packages functionality. R is capable of performing element-wise and matrix operations that drastically reduce the computational time needed for decision trees, even for decision trees with multiple nested branches.

State-transition models

Markov model cohort simulations

Markov models are used to describe the transitions of a simulated cohort of patients among several mutually exclusive and exhaustive health states during a series of short intervals or cycles. Markov models are useful for simulating disease progression and the effects of interventions over time and have been a fundamental tool in health decision sciences for more than 30 years.[3,11] The typical way of simulating a Markov model is through matrix operations. The R core packages efficiently handle matrix operations as single objects, which allow calculating discrete-time Markov models without the need for additional packages. However, user contributed packages such as `matrixStats` provide added matrix- and vector-based methods to compute a wide range of summary statistics on one or more dimensions of multidimensional matrices, which are optimized for speed and memory usage. R also has packages with already built-in functions to both simulate and determine properties of discrete-time Markov models, such as the first passage time into each state and the steady-state probabilities. For example, the packages `DTMCPack` and `markovchain` provide methods to develop and run discrete-time Markov models. In addition, `markovchain` and `heemod` packages provide functions to perform statistical and probabilistic analysis, estimate transition probabilities from data, and plot state-transition diagrams. Furthermore, continuous-time Markov models can be stochastically simulated using `GillespieSSA`, as illustrated in reference 12.

Markov decision processes (MDPs) are specialized Markov models that allow decisions to be made during each Markov cycle and produce a sequence of decisions that optimize the overall expected utility. These tools are useful when the decision problem involves estimating the optimal timing of an intervention.[13] MDPs generally use a backward induction algorithm that computes the expected value of each decision starting from the last cycle backward until it reaches the first cycle. The decisions that produce the highest utility in each cycle are then recorded as defining the optimal policy. Standard decision analysis software generally does not support this algorithm because this software generally uses a forward algorithm that would typically be computationally very expensive in determining the optimal timing. The package `MDPtoolbox` is specifically written to allow R users to solve MDPs.

Individual-level microsimulation

Microsimulation models generally involve sampling many hypothetical individuals and tracking their progress over time. One important advantage of this approach over standard Markov models is that it allows for history of the simulated individuals to be incorporated more efficiently. R's core packages contain many of the necessary functionalities for the development and simulation of standard microsimulation models. The `MILC` package is a microsimulation model of the natural history of lung cancer progression and is developed entirely in R.

In addition, the S4 language, which can be used in R, natively supports object-oriented programming in which individuals can

be assigned attributes and inherit attributes from parent classes. Additional packages such as `RNetLogo` allow the integration of R with NetLogo,[14] a software dedicated to agent-based simulation modeling. Furthermore, the package `spaDES` is specifically designed for building discrete-event simulations.

Dynamic transmission models (compartmental models)

Dynamic transmission models are most often used to model the spread of infectious diseases. Like Markov models, compartmental models describe transitions of a population through a set of states; however, unlike most Markov models, certain transition probabilities may depend on the current composition of the population. For example, the probability of new infections depends on the number of infectious individuals in the population.

One of the simplest infectious disease compartmental models is the 3-compartment susceptible infected-recovered (SIR) model. In an SIR model, the probability of transitioning from the susceptible to infected compartment is proportional to the number of currently infected individuals in the population. Because of the state dependence of the transition probabilities, these models are often represented as either ordinary differential equations (ODEs) or partial differential equations, which can be solved numerically in R. Standard ODE solvers are implemented in R through the `deSolve` package. R also supports the stochastic simulation of system dynamics represented by ODEs through the `GillespieSSA` and `adaptivetau` packages. These packages also include functions that implement basic infectious disease model structures, such as SIR, that can be customized and extended to more complex models.

The `EpiModel` package was developed explicitly to facilitate the development and simulation of infectious disease models in R. `EpiModel` is a framework that includes built-in functionality for the most commonly used infectious disease model types, such as SIR, as well as a module-based structure that is designed to facilitate user customization for more complex disease models. While `EpiModel` supports deterministic compartmental models, it can also simulate stochastic infectious disease spread through either a random contact process or a dynamic contact network, which is implemented using the network analysis and simulation package `statnet`.

Parameter estimation

R provides various powerful statistical and numerical tools for parameter estimation through statistical analyses of patient-level or aggregated data as well as tools for estimating unobservable or unknown parameter values through model calibration. In the following sections, we summarize some of these functionalities in R.

Analysis of patient-level or aggregate preexisting data

Decision analysts often rely on the analysis of existing data to populate their model parameters. R has numerous packages for conducting various forms of basic and advanced statistical analyses. In addition, special packages can be used to directly access commonly used data sources (e.g., the Surveillance, Epidemiology, and End Results [SEER] data set).

Statistical analysis of primary data

Input parameter estimates in decision models are often obtained from statistical analysis of primary data. For example, partitioned survival models are frequently used in cancer models and rely on estimates from parametric survival regression models.[15] Several R packages (e.g., the core package `survival` and the contributed package `flexsurv`) allow the analysis of survival data in order to be integrated in a decision model. Multistate models (MSM) are the generalization of survival analysis including more than two competing states. R has various packages that estimate MSM. For example, `msm`, `mstate`, and `SemiMarkov` are used to fit MSM to panel data in survival analysis applications, `TPmsm` for estimating transition probabilities for 3-state progressive disease models, and HMM and `depmixS4` for fitting hidden Markov models from individual-level data.

In addition, R provides tools for performing numerous types of regression-based analyses, including linear and generalized linear regression models (functions `lm` and `glm`), generalized linear mixed models (e.g., packages `nlme` and `lme4` being the most popular), generalized estimating equations (e.g., package `gee`), instrumental variables (e.g., package `AER`), and marginal structural models (e.g., package `ipw`). These are only a few examples of the wide variety of packages that are offered by R for regression-based analyses. For a more comprehensive list of packages that can be useful in parameter estimations, the reader is advised to use the CRAN Task View: Econometrics (https://cran.r-project.org/web/views/Econometrics.html) and the extensive online support available for R.

Meta-analysis and evidence synthesis

It is suggested that economic evaluations should ideally compare all treatment strategies and should rely on best available evidence.[16] For this reason, the process of systematic review and meta-analysis precedes most economic evaluations, either as part of that analysis or as input for the economic model.

R serves as a rich environment for conducting evidence synthesis. R includes a convenient set of functions that can facilitate conducting systematic reviews, manipulating the study-extracted data, and performing meta-analyses, including network meta-analyses. One of the most widely used packages for conducting systematic reviews through R is the `RISmed` package. Many examples rely on `RISmed` to automate study identification such as extracting information for respiratory health studies from PubMED.[17] This package allows searching within any of the Entrez databases, including PubMed and the National Library of Medicine collection, and extraction of extensive information from the identified studies (e.g., abstract title, authors, year of publication, journal published, and more). The structured approach in this metadata extraction in combination with the graphic tools of R allows for intuitive summaries of this metadata information.

There are several packages that allow the R user to conduct meta-analysis (e.g., `meta`, `metafor`, `rmeta`). These packages provide specific meta-analytic functionality, such as forest and funnel plots, as well as standard statistical analyses. For example, fixed- and random-effects meta-analyses can be conducted using either the `rmeta` or `metaphor` packages but can also be executed using the generalized linear mixed model functions from the `nlme` of `lme4` packages. Extensions of the standard univariate, inverse variance methods can also be conducted using R. The package `mvmeta` allows for multivariate fixed- or random-effects models, which can incorporate correlation structures between outcomes within the same sample of studies.

R has been used widely for network meta-analysis. A series of recently published articles on evidence synthesis for decision making describes a set of tutorials for a framework of evidence synthesis using fixed- or random-effect, indirect comparisons and network meta-analysis.[18–24] Using R is advantageous because it provides a wide array of tools to conduct these analyses. For example, the packages `netmeta` and `gemtc` allow for conducting network meta-analyses within a frequentist or a Bayesian framework, respectively. R has a highly evolved integration with Bayesian inference using a set of probabilistic programming languages, referred to as Bayesian inference Using Gibbs Sampling (BUGS) languages (e.g., WinBUGS, OpenBUGS, and JAGS). These applications can be programmatically accessed through R using the packages `R2WinBUGS`, `R2OpenBUGS`, and `R2jags`. These packages allow Bayesian methods to be seamlessly integrated with decision-analytic models. In addition, these packages are often used in network meta-analyses for evidence syntheses. The results of the meta-analyses and network meta-analyses can be readily incorporated in the economic evaluation, allowing for uncertainty to be appropriately propagated in the economic model.[16]

Direct analyses of public data sets

R also facilitates access to publicly available databases of primary and aggregated data, such as life tables or cancer surveillance data. This input can be used by decision modelers for the direct estimation of input parameters (e.g., all-cause mortality) but also for the calibration and validation of the simulation models. Commonly accessed databases include the Human Fertility Database[25] through the `HMDHFDplus` package, the Human Mortality Database[26] through the `demography` package, and the SEER Database[27] through the `SEER2R` package. Some of these packages also include useful functions for the analysis of such epidemiological data. For example, the demography package allows for the estimation of Lee-Carter type of models, which use age-specific mortality rates to forecast future mortality rates, accounting for changes in life expectancy over time.[28]

Calibration

Calibration involves identifying appropriate parameter values based on the model fit to some observed target data. Often, there is a subset of model parameters that cannot be directly observed or estimated using traditional statistical models, such as the probability of a preclinical cancer becoming symptomatic. Modelers often adopt calibration as a tool to infer the value for these unobservable parameters. R provides a wide selection of packages for optimization that could potentially be used for model calibration. The `optim` function included in the `stats` package is a general-purpose optimization tool that includes different algorithms, such as directed search optimization (i.e., Nelder–Mead), quasi-Newton methods, and the conjugate-gradient algorithm. It can also conduct box-constrained optimization and simulated annealing. The `optimx` package provides more features including additional optimization algorithms and constraint handlings. Nelder–Mead is a commonly used algorithm in model calibration in health decision sciences and has its own standalone implementation in R through the package `neldermead`, which has been used in decision models in the past (examples include references 29–34). R also includes packages that implement global optimization algorithms. For example, the package `genalg` can perform a genetic algorithm search of the parameter space for appropriate estimates, and the package `DEoptim` implements the differential evolution algorithm for global optimization of a function of a real-valued parameter vector. A generalized version of the simulated annealing algorithm has its own standalone package named `GenSA`. A comprehensive list of R packages that provides tools for solving optimization problems is maintained by the CRAN Task View: Optimization and Mathematical Programming (https://cran.r-project.org/web/views/Optimization.html). Furthermore, Latin hypercube sampling (LHS), an efficient sampling method, can be used for model calibration where the parameter space is split recursively until the parameter subspace that contains the best-fitting parameter space is found. `LHS` in R can be implemented using the lhs package.

Bayesian calibration can be implemented with either Markov chain Monte Carlo (MCMC) or integrated nested Laplace approximation algorithms. For example, `MHadaptive` implements an adaptive Metropolis-Hastings (MH) sampling algorithm of a user-defined function. The package also provides some functions for Bayesian inference including Bayesian credible intervals and deviance information criterion calculation. The package `mcmc` simulates continuous distributions of random vectors using MCMC. This package provides an implementation of the MH sampling algorithm that uses a multivariate normal proposal distribution with variance proportional to the identity matrix. In addition to `mcmc`, several other packages can perform Bayesian inference, such as `MCMCpack`. As an alternative to MCMC, the package `INLA` could be used to approximate Bayesian inference for latent Gaussian Markov random field models. The CRAN Task View also maintains a comprehensive list of R packages that provides tools for Bayesian inference (https://cran.r-project.org/web/views/Bayesian.html).

Model analysis and output

After the model is built and input parameters are estimated, modelers can use R to conduct various analyses including PSA and VOI analysis. In addition, many analyses in R can be run more efficiently by taking advantage of the available computing resources and R's native capability of handling matrix and vector operations efficiently. Furthermore, R provides superior graphing capabilities through a set of packages that produce publication-quality figures, which allows a high degree of customization of these figures.

Probabilistic analyses

As decision models usually rely on input parameters that are estimated and hence not known with certainty, the results of these models are also associated with uncertainty. To incorporate this uncertainty, decision modelers rely on probabilistic sensitivity analysis (PSA) in which each input parameter is associated with a distribution rather than a single value. Monte Carlo methods are applied to propagate parameter uncertainty to the model outputs. PSA is computationally demanding in many models and requires random number-generating or resampling capabilities. The core R packages include basic functionalities necessary for probabilistic sampling, including random-number generation functionality for several univariate distributions (e.g., normal, beta, and gamma) and calculating the density, cumulative density, and quantiles for many standard probability distributions that are frequently used in decision analyses. Sampling from more complex distributions, such as multivariate normal and the Dirichlet distributions, can also be achieved using the `MASS`, `LCA`, and `mvtnorm` packages. In addition, users can sample from several independent parameters and later induce correlations using published user-written R functions.[35]

Much of R's functionality used for primary data analysis can also be used to summarize the results from decision models. In addition, there is an increasing number of efforts to streamline and simplify the production of figures and tables specific to cost-effectiveness analyses. For example, the package `BCEA` performs several analyses on probabilistic decision models that are relevant to health economic evaluations, including VOI analysis (see below). This package uses a probabilistic data set as an input and generates various plots, such as cost-effectiveness planes, a cost-effectiveness acceptability curve (CEAC), the contours and scatterplots for the distribution of incremental costs and effectiveness, and expected value of perfect information figure.

VOI analysis

VOI measures the value of reducing uncertainty in a cost-effectiveness or decision analysis. VOI analysis is gaining increasing recognition. The packages `BCEA` and `SAVI` (available on GitHub) can calculate expected value of perfect information from a PSA data set. In addition, many authors have published R code to conduct VOI. Examples of such efforts include using generalized additive regression models and Bayesian updating,[36] using the unit normal loss integral and regression meta-modeling,[37] and Bayesian Laplace approximation.[38]

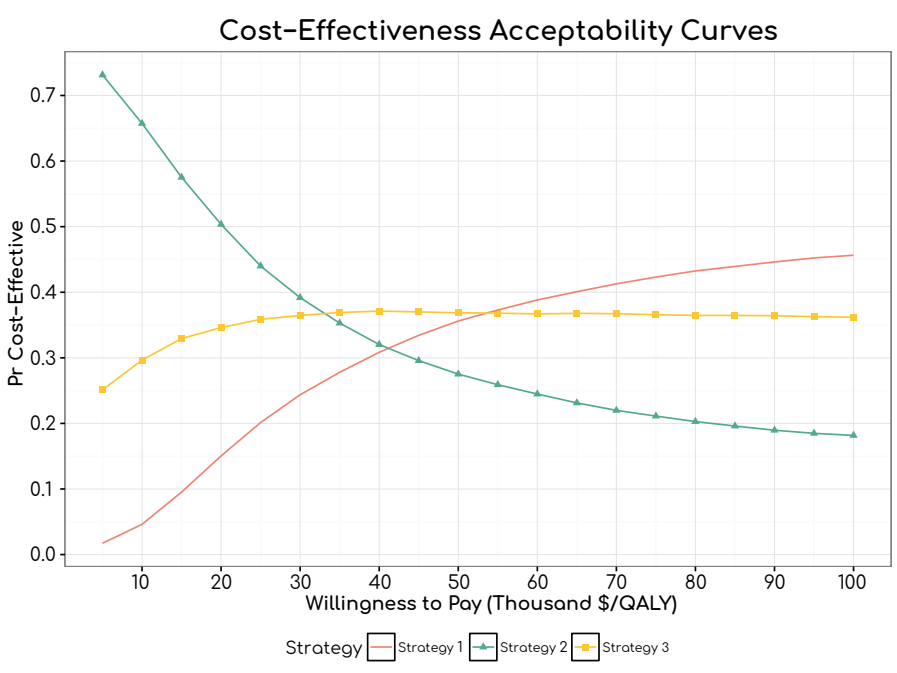
Increased computational efficiency

By default, R uses only one computational thread of the computer's central processing unit. However, most modern computers can run multiple threads in parallel. R can use all the cores on a machine for matrix operations if the basic linear algebra subprograms libraries are optimized for parallel processing. In addition, R has a set of packages that enable parallel computing (e.g., `foreach`, `parallel`), which allow the simultaneous execution of operations on multiple threads in parallel and thereby reducing computation time. The `foreach` package supports parallel execution of the `for` command, which can execute repeated operations on multiple processors (cores) on a single computer or on multiple nodes of a network-connected computer cluster. The `parallel` package also supports this type of parallel computation through an implementation of the `apply` functions provided in the core package in R.

Graphical outputs

In addition to specialized analyses, decision analysts often use graphics to communicate the results of their analyses. One of the main advantages of using R is the ability to produce publication-quality figures and the wealth of information and support available on how to customize these figures to achieve the desired results. For example, the package `ggplot2` provides a high degree of flexibility to create polished and complex plots. An example of a CEAC plotted with `ggplot2` is shown in Figure 3. In addition, the package `lattice` improves on base R graphics by providing better defaults and by simplifying the visualization of multivariate relationships. Users can find many resources online (e.g., http://www.cookbook-r.com/Graphs/) that assist in learning the plotting capabilities of R.

Figure 3



Cost-effectiveness acceptability curves of a generic cost-effectiveness analysis plotted with ggplot2.

Discussion

In this article, we illustrated the increasing use of R in health decision sciences and provided an overview of R packages and user-written functions that can be useful in decision analysis. Although many of these packages are not specific to decision analysis, they are nonetheless highly relevant and, in most circumstances, can be applied without modification. We grouped relevant R packages into 3 broad stages of decision analysis: model development, input parameter estimation, and output analyses. In addition, we highlighted the integration of R with other software, including BUGS and NetLogo.

Decision analysts increasingly adopt a programming approach. Although learning a new programming language can impose an initial barrier, the expected payoffs could outweigh the entry costs. For example, using a programming language allows for modeling complex situations, integration with primary data analysis, documentation, transparency, automation of repeated analyses, troubleshooting, reproducibility, and visualization of the results with advanced graphical tools.

In addition to R, there are other high-level languages that are commonly used in decision analyses, such as C++, MATLAB, and Python, which are sometimes superior to R, particularly in computational efficiency. However, R is unique because it is supported by a large network of users and a culture of sharing open-source code that is already proven useful among statisticians. In fact, most of R functionality is provided by user-written packages and online support, allowing R to be highly customizable so that users can develop their own functions that fit their needs. R packages are often of high quality and written by specialized professionals and academicians. These packages tend to be well scrutinized by the community, which helps improve the code and remove the bugs. This is facilitated by the open-source nature of the R code, which allows anyone to view the underlying code and potentially contribute to improving it.

There are some drawbacks in using R or any other high-level programming languages in health decision sciences. One major limitation is the lack of a graphical user interface that helps decision analysts to visualize their mathematical models. Such visualization is often desired because it facilitates the development and communication of the model structure. For example, decision trees often involve conditional probabilities that have an intuitive representation with a tree diagram. Although these trees can be represented with mathematical formulae, these formulae tend to become nested and very complicated to visualize, interpret, or debug without a graphical user interface. The package `ArvoRe` is an effort to design cost-effectiveness analyses with decision trees and simple Markov models in R.

Furthermore, to address this limitation, we are currently developing an open-source tool (OpenTree) that provides an interface for building and editing decision models. OpenTree can translate graphical trees to R code. Users can then use this generated R code independently from the graphical representation of the model to conduct their analyses.

Another limitation of open-source software, in general, relates to package dependencies and version control. For example, it is important to maintain backward compatibility when new packages are released. Although R has a mechanism to decrease the risk of package incompatibility, this task becomes particularly challenging as the number of user-written packages increases. The package packrat is an effort to overcome this challenge. This package allows each R project to have its own library of packages, thus preventing version conflict and broken package dependencies among various projects.

The purpose of this article is to provide an overview of the use of R in health decision sciences compared with several other software packages and to provide a sample of R packages that we believe are useful in various stages of model design and development.

There are other repositories of R packages in addition to CRAN, including GitHub, Bioconductor, and R-Forge, that host additional R packages that may be of use in health decision sciences. In addition, the list of articles, journals, software, R packages, and repositories included in this overview is to provide only an overview and is not meant to be comprehensive. For example, there are other software that were not included in this overview, such as simul8 (simul8, Boston, MA) and @risk (Palisade, Ithaca, NY). In addition, we recognize that our sample was limited in Google Scholar, and there are other important journals (e.g., Health Technology Assessment) that were not included. However, exhaustively listing every R package, function, and repository available, especially when new packages are continuously added, is challenging. To address this challenge, we created a Wiki page (R in Health Decision Sciences) available at https://en.wikiversity.org/wiki/R_in_Health_Decision_Sciences. This resource allows the medical decision sciences community to freely add new information, share their experiences, and update the existing resources. We prefer this approach over a published list as it is less likely to be outdated.

R is becoming increasingly popular as a teaching tool for health decision sciences at universities and in various short courses and workshops in professional society meetings in recent years. Furthermore, we decided to create the Decision Analysis in R for Technologies in Health (DARTH) workgroup. The group is an international endeavor composed of members with shared interests modeling in R with the goal of providing tutorials and courses to familiarize scientists, especially new investigators with using R in health decision sciences.

In summary, R has rapidly become one of the most widely used tools in statistical analyses. Many of the R packages can be applied to decision analysis with minimal or no modifications. Thus, the purpose of this overview is to provide a baseline reference point for users interested in navigating the large number of R packages relevant to decision analyses. In subsequent tutorials, and as part of the DARTH's group efforts, we will detail how to use OpenTree and conduct specific

tasks that are commonly used in decision analysis and simulation modeling, for example, model building, calibration, and VOI analysis among others.

Supplemental material

Appendix A and B of this article can be found in the online version at https://journals.sagepub.com/doi/suppl/10.1177/0272989X16686559

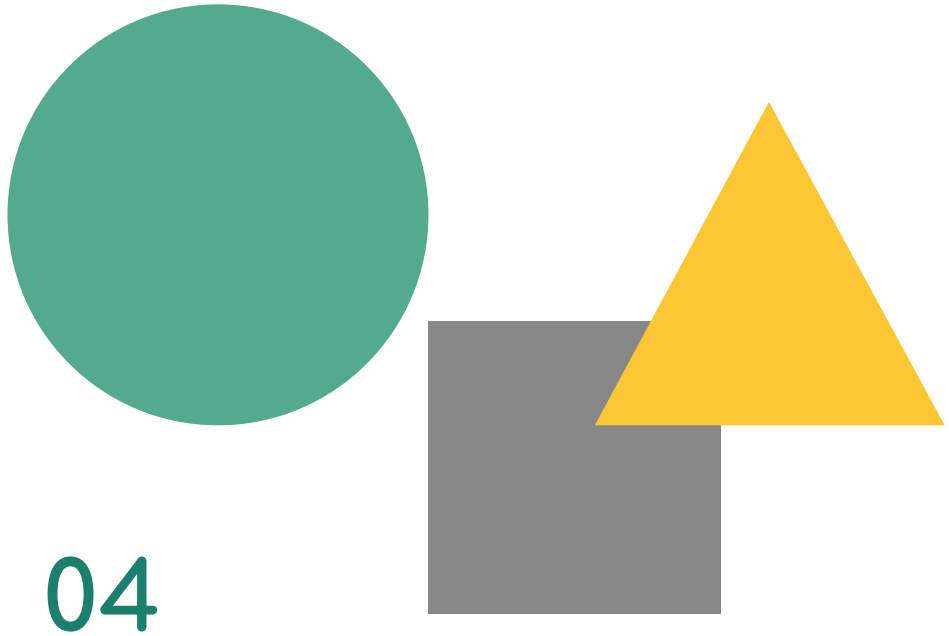
Acknowledgments

The authors would like to thank Professor Jon Baron and an anonymous reviewer of *Medical Decision Making* for their valuable suggestions.

References

1. Tosh J, Wailoo A. *Review of Software for Decision Modelling*. 2008.
2. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making*. 2017; 37(4):340–352.
3. Hawkins N, Epstein D, Drummond M, Wilby J, Kainth A, Chadwick D, Sculpher M. Assessing the cost-effectiveness of new pharmaceuticals in epilepsy in adults: the results of a probabilistic decision model. *Med Decis Making*. 2005; 25(5):493–510.
4. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. 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. Briggs AH, Weinstein MC, Fenwick E a L, Karnon J, Sculpher MJ, Paltiel a D. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012; 32(5):722–732.
6. Pitman R, Fisman D, Zaric G, Postma M, Kretzschmar M, Edmunds J, Brisson M, On Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. 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.
7. 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.
8. 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. *Med Decis Making*. 2012; 32(5):690–700.
9. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012; 32(5):678–689.
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. Beck JR, Pauker SG. The Markov Process in Medical Prognosis. *Med Decis Making*. 1983; 3(4):419–458.
12. Iskandar R, Alarid-Escudero F, Kuntz KM. Does Size Matter? A Method for Adaptively Determining Cycle Length in a State-Tranistion Model. In: *The 37th Annual Meeting of the Society for Medical Decision Making. 2015*, p. 2.
13. Shechter SM, Bailey MD, Schaefer AJ, Roberts MS. The Optimal Time to Initiate HIV Therapy Under Ordered Health States. Operations Research. *Oper Res*. 2008; 56(1):20–33.

14. Wilenskly U. NetLogo: Center for Connected Learning and Computer-Based Modeling, https://ccl.northwestern.edu/netlogo/ (1999).
15. Latimer NR. Survival analysis for economic evaluations alongside clinical trials - Extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013; 33(6):743–754.
16. Welton NJ, Sutton AJ, Cooper N, Abrams KR AA. *Evidence Synthesis for Decision Making in Healthcare*. West Sussex: A John Wiley & Sons, Ltd., Publication, 2012.
17. Burchard EG, Oh SS, Foreman MG, Celedón JC. Moving toward True inclusion of racial/ethnic minorities in federally funded studies: A key step for achieving respiratory health equality in the United States. *Am J Respir Crit Care Med*. 2015; 191(5):514–521.
18. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. *Med Decis Making*. 2013; 33(5):679–691.
19. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence Synthesis for Decision Making 1: Introduction. *Med Decis Making*. 2013; 33(5):597–606.
20. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials. *Med Decis Making*. 2013; 33(5):607–617.
21. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence Synthesis for Decision Making 3: Heterogeneity-Subgroups, Meta-Regression, Bias, and Bias-Adjustment. *Med Decis Making*. 2013; 33(5):618–640.
22. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence Synthesis for Decision Making 4: Inconsistency in Networks of Evidence Based on Randomized Controlled Trials. *Med Decis Making*. 2013; 33(5):641–656.
23. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making*. 2013; 33(5):657–670.
24. Dias S, Sutton AJ, Welton NJ, Ades AE. Evidence Synthesis for Decision Making 6: Embedding Evidence Synthesis in Probabilistic Cost-effectiveness Analysis. *Med Decis Making*. 2013; 33(5):671–678.
25. Shkolnikov VM ST. Human Fertility Database. *Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria)*, www.humanfertility.org.
26. Shkolnikov V, Barbieri M WJ. The Human Mortaliy Database. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*, http://www.mortality.org/.
27. National Institutes of Health. Surveillance, Epidemiology, and End Results (SEER) program populations 1969-2014. *Rockville, MD: National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch.*
28. Booth H, Hyndman RJ, Tickle L, de Jong P. Lee-Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demogr Res*. 2006; 15(9):289–310.
29. Goldhaber-Fiebert JD, Brandeau ML. Modeling and Calibration for Exposure to Time-Varying, Modifiable Risk Factors: The Example of Smoking Behavior in India. *Med Decis Making*. 2015; 35(2):196–210.
30. Taylor DCA, Pawar V, Kruzikas DT, Gilmore KE, Sanon M, Weinstein MC. Incorporating calibrated model parameters into sensitivity analyses: Deterministic and probabilistic approaches. *Pharmacoeconomics*. 2012; 30(2):119–126.
31. Chia YL, Salzman P, Plevritis SK, Glynn PW. Simulation-based parameter estimation for complex models: a breast cancer natural history modelling illustration. *Stat Methods Med Res*. 2004; 13(6):507–524.
32. Kim JJ, Kuntz KM, Stout NK, Mahmud S, Villa LL, Franco EL, Goldie SJ. Multiparameter calibration of a natural history model of cervical cancer. *Am J Epidemiol*. 2007; 166(2):137–150.
33. Taylor DCA, Pawar V, Kruzikas D, Gilmore KE, Pandya A, Iskandar R, Weinstein MC. Methods of Model Calibration: Observations from a Mathematical Model of Cervical Cancer. *Pharmacoeconomics*. 2010; 28(11):995–1000.
34. Stout NK, Knudsen AB, Kong CY (Joey), Mcmahon PM, Gazelle GS. Calibration Methods Used in Cancer Simulation Models and Suggested Reporting Guidelines. *Pharmacoeconomics*. 2009; 27(7):533–545.
35. Goldhaber-Fiebert JD, Jalal HJ. Some Health States Are Better Than Others: Using Health State Rank Order to Improve Probabilistic Analyses. *Med Decis Making*. 2016; 36(8):927–940.
36. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast, Nonparametric Regression-Based Method. *Med Decis Making*. 2015; 35(5):570–583.
37. Jalal H, Goldhaber-Fiebert JD, Kuntz KM. Computing expected value of partial sample information from probabilistic sensitivity analysis using linear regression metamodeling. *Med Decis Making*. 2015; 35(5):584–595.
38. Brennan A, Kharroubi SA. Efficient computation of partial expected value of sample information using Bayesian approximation. *J Health Econ*. 2007; 26(1):122–148.



04

AN INTRODUCTORY TUTORIAL ON COHORT STATE-TRANSITION MODELS IN R USING A COST-EFFECTIVENESS ANALYSIS

Fernando Alarid-Escudero
Eline Krijkamp
Eva A. Enns
Alan Yang
M.G. Myriam Hunink
Petros Pechlivanoglou
Hawre Jalal

Medical Decision Making – under review

“And now that you don’t have to be perfect, you can be good.”

John Steinbeck

AN INTRODUCTORY TUTORIAL ON COHORT STATE-TRANSITION MODELS IN R USING A COST-EFFECTIVENESS ANALYSIS

Abstract

Decision models can combine information from different sources to simulate the long-term consequences of alternative strategies in the presence of uncertainty. Cohort state-transition models (cSTM) are decision models commonly used in medical decision-making to simulate hypothetical cohorts’ transitions among various health states over time. This tutorial focuses on time-independent cSTM, where transition probabilities among health states remain constant over time. We implement time-independent cSTM in R, an open-source mathematical and statistical programming language. We construct time-independent cSTMs using a previously published decision model, calculate costs and effectiveness outcomes, conduct a cost-effectiveness analysis of multiple strategies, and conduct a probabilistic sensitivity analysis. We provide open-source code in R to facilitate wider adoption. In a second more advanced tutorial, we illustrate time-dependent cSTMs.

Introduction

Policymakers are often tasked with allocating limited healthcare resources under constrained budgets and uncertainty about future outcomes. Health economic evaluations might inform their final decisions. These economic evaluations often rely on decision models to synthesize evidence from different sources and project long-term outcomes of various alternative strategies. A commonly used decision model is the discrete-time cohort state-transition model (cSTM), often referred to as a Markov model.[1]

In a recent review, we illustrated the increased use of R's statistical programming framework in health decision sciences. We provided a summary of available resources to apply to medical decision making.[2] Many packages have been explicitly developed to estimate and construct cSTMs in R. For example, the `markovchain`[3] and `heemod`[4] packages are designed to build cSTMs using a pre-defined structure. The `markovchain` package focuses on simulating time-independent and time-dependent Markov chains but is not designed to conduct economic evaluations. `heemod` is a well-structured R package for economic evaluations. However, these packages are necessarily stylized and inflexible, requiring users to follow a specific cSTM structure. For example, users are still required to set up the cSTM structure, specify the parameters, and run analyses in a pre-specified approach, limiting the understanding of how cSTMs work and are constructed. If the desired cSTM does not fit within this structure, using these packages can be challenging.

This tutorial demonstrates how to conduct a full cost-effectiveness analysis (CEA) comparing multiple interventions and implementing probabilistic sensitivity analysis (PSA) without needing a specialized cSTM package. We first describe each of the components of a time-independent cSTM. Then, we illustrate the implementation of these components with an example. Our general conceptualization should apply to other programming languages (e.g., MATLAB, Python, C++, and Julia). The reader can find the most up-to-date R code of the time-independent cSTM and the R code to create the tutorial graphs in the accompanying GitHub repository (https://github.com/DARTH-git/cohort-modeling-tutorial-intro) to replicate and modify the example to fit their needs. We assume that the reader is familiar with the basics of decision modeling and has a basic understanding of programming. Thus, a prior introduction to R and linear algebra for decision modelers is recommended.

This introductory tutorial aims to (1) conceptualize time-independent cSTMs for implementation in a programming language and (2) provide a template for implementing these cSTMs in base R. We focus on using R base packages, ensuring modelers understand the concept and structure of cSTMs and avoid the limitation of constructing cSTMs in a package-specific structure. We used previously developed R packages for visualizing CEA results and checking cSTMs are correctly specified.

Cohort state-transition models (cSTMs)

A cSTM is a dynamic mathematical model in which a hypothetical cohort of individuals transition between different health states over time. In contrast, an individual-based state-transition model (iSTM) is a type of STM where simulated individuals transition between health states over time.[5] We have previously published a tutorial on the implementation of iSTM in R.[6]

A cSTM is most appropriate when the decision problem has a dynamic component (e.g., the disease process can vary over time) and can be described using a reasonable number of health states. cSTMs are often used because of their transparency, efficiency, ease of development, and debugging. cSTMs are usually computationally less demanding than iSTMs, providing the ability to conduct PSA and value-of-information (VOI) analyses that otherwise might not be computationally feasible with iSTMs.[5] cSTMs have been used to evaluate screening and surveillance programs,[7,8] diagnostic procedures,[9] disease management programs,[10] interventions,[11] and policies.[12]

A cSTM consists of a set of n_s mutually exclusive and collectively exhaustive health states. The cohort is assumed to be homogeneous within each health state. Individuals in the cohort residing in a particular health state are assumed to have the same characteristics and are indistinguishable from one another. The cohort transitions between health states with defined probabilities, which are called "transition probabilities". A transition probability represents the chance that individuals in the cohort residing in a state in a given cycle transition to another state or remain in the same state. In a cSTM, a one-cycle transition probability reflects a conditional probability of transitioning during the cycle, given that the person is alive at the beginning of the cycle.[13]

Transition probabilities only depend on the current health state in a given cycle. They do not depend on the history before that cycle. This property is often referred to as the "Markovian assumption".[14–16] This means that in a cSTM, transition probabilities do not depend on the history of past transitions or time spent in a given state. In the advanced tutorial, we illustrate how to incorporate time dependence in a cSTM.

cSTMs can be classified as either time-independent (time-homogeneous) or time-dependent (time-inhomogeneous). Time-independent cSTMs have constant transition probabilities (i.e., the probability of any state transition is independent of time). In contrast, time-dependent cSTMs have transition probabilities or rewards (e.g., costs or utilities associated with being in a particular health state) that vary over time. A reward refers to a value assigned to individuals for being in a given health state (state reward) or transitioning between health states (transition reward). Time-independent models are more straightforward to implement than time-dependent ones, but most problems in healthcare are best modeled with time-dependent cSTMs. For example, time-dependent cSTMs can capture the increasing age-specific background mortality as the cohort ages (age dependency) and dependency on the amount of time spent in a given state (state residence dependence). However, it is easiest to start implementing the structure of a time-independent model and expand it to account for various types of time dependency. Therefore, this tutorial focuses on time-independent cSTMs. The advanced tutorial illustrates how to construct time-dependent cSTMs.[17] We begin our exposition by explaining the difference between rates and probabilities and how to transform from one to the other under certain assumptions.

Rates versus probabilities

Discrete-time cSTMs use probabilities to determine transitions between states. However, these transitions might be reported in terms of rates, or probabilities may not always be available in the desired cycle length. For example, transition probabilities might be available from published literature in one time period (e.g., annual) and might differ from the model's cycle length time scale (e.g., monthly). Below, we illustrate a simple approach to converting from rates to probabilities and using rates to convert probabilities from one time scale to another.

While probabilities and rates are often numerically similar in practice, there is a subtle but important conceptual difference between them. A rate represents the *instantaneous* force of an event occurrence per unit time, while a probability represents the cumulative risk of an event over a defined period.

To illustrate this difference further, let us assume that after 10,000 person-years of observation of healthy individuals (e.g., 10,000 individuals observed for an average of 1 year, or 5,000 individuals observed for an average of 2 years, etc.), we observe 500 events of interest (e.g., becoming sick from some disease). The annual event rate of becoming sick, $\boldsymbol{\mu}_{yearly}$, is then equal to $\boldsymbol{\mu}_{yearly}$=500/10,000=0.05.

If we then wanted to know what proportion of an initially healthy cohort becomes sick at the end of the year, we can convert the

annual rate of becoming sick into an annual probability of becoming sick using the following equation:

$$p_{yearly} = 1 - \exp\left(-\mu_{yearly}\right), \tag{1}$$

resulting in $\boldsymbol{p}_{yearly}$ = 1 -exp (-0.05) = 0.0488. Equation (1) assumes that the rate of becoming sick is constant over the year, implying that the time until a healthy person becomes sick is exponentially distributed. The parameter $\boldsymbol{p}_{yearly}$ is the transition probability from healthy to sick in a cSTM when using an annual cycle length.

If we were concerned that an annual cycle length was too long to capture disease dynamics accurately, we could use a monthly cycle length. This change of time scale would require us to parameterize the model with monthly transition probabilities. To convert the annual probability of becoming sick to a monthly probability, we first calculate the monthly rate of becoming sick:

$$\mu_{yearly} = -\ln\left(1 - p_{yearly}\right). \tag{2}$$

Because rates are instantaneous, the monthly rate is then just the annual rate divided by 12:

$$\mu_{monthly} = \mu_{yearly}/12. \tag{3}$$

We divide by 12 because this is the number of months (desired cycle length) in a year (cycle length of the given data). If the original or desired cycle length were different, we would divide by a different factor (e.g., annual to weekly: 52; monthly to annual: 1/12; annual to daily: 365.25, etc.).

The monthly probability can then be calculated from the monthly rate using equation (1):

$$p_{monthly} = 1 - \exp\left(-\mu_{monthly}\right). \tag{4}$$

These equations are also useful for computing probabilities when studies (e.g., survival analyses) provide rates rather than transition probabilities.

Time-independent cSTM dynamics

A cSTM consists of three core components: (1) a state vector, $\boldsymbol{m}_{t,}$ that stores the distribution of the cohort across all health states in cycle t where t = 0, ⋯, n_T; (2) the cohort trace matrix, $\boldsymbol{M}$, that stacks $\boldsymbol{m}_t$ for all t and represents the distribution of the cohort in the various states over time; and (3) a transition probability matrix, $\boldsymbol{P}$.[18] If the cSTM is comprised of n_S discrete health states, $\boldsymbol{m}_t$ is a 1 × n_S vector and $\boldsymbol{P}$ is a n_S × n_S matrix. The i-th element of $\boldsymbol{m}_t$, where i = 1, ⋯, n_S, represents the proportion of the cohort in the i-th health state in cycle t, referred to as $m_{[t,i]}$. Thus, $\boldsymbol{m}_t$ is written as:

$$\mathbf{m}_t = \begin{bmatrix} m_{[t,1]} & m_{[t,2]} & \cdots & m_{[t,n_S]} \end{bmatrix}.$$

The elements of $\boldsymbol{P}$ are the transition probabilities of moving from state i to state j, $p_{[i,j]}$, where $\{i,j\}$ = 1, ⋯, n_S and all should have values between 0 and 1.

$$P = \begin{bmatrix} p_{[1,1]} & p_{[1,2]} & \cdots & p_{[1,n_S]} \\ p_{[2,1]} & p_{[2,2]} & \cdots & p_{[2,n_S]} \\ \vdots & \vdots & \ddots & \vdots \\ p_{[n_S,1]} & p_{[n_S,2]} & \cdots & p_{[n_S,n_S]} \end{bmatrix}.$$

For $\boldsymbol{P}$ to be a correctly specified transition probability matrix, each row of the transition probability matrix must sum to one, $\sum_{j=1}^{n_s} p_{[i,j]} = 1$ for all $i = 1, \cdots, n_s$.

The state vector at cycle $t + 1$ ($\boldsymbol{m}_{t+1}$) is then calculated as the matrix product of the state vector at cycle t, $\boldsymbol{m}_t$, and the transition probability matrix, $\boldsymbol{P}$, such that

$$\mathbf{m}_{t+1} = \mathbf{m}_t P \text{ for } t = 0, \ldots, (n_T - 1),$$

where $\boldsymbol{m}_1$ is computed from $\boldsymbol{m}_0$, the initial state vector with the distribution of the cohort across all health states at the start of the simulation (cycle 0). Then, we iteratively apply this equation through $t = (n_T - 1)$.

The cohort trace matrix, $\boldsymbol{M}$, is a matrix of dimensions $(n_T + 1) \times n_s$ where each row is a state vector ($\boldsymbol{m}_t$), such that

$$M = \begin{bmatrix} -\mathbf{m}_0- \\ -\mathbf{m}_1- \\ \vdots \\ -\mathbf{m}_{n_T}- \end{bmatrix}.$$

Note that the initial cycle (i.e., cycle 0) corresponds to $t = 0$, which is on the first row of $\boldsymbol{M}$. Thus, $\boldsymbol{M}$ stores the output of the cSTM, which could be used to compute various epidemiological, and economic outcomes, such as life expectancy, prevalence, cumulative resource use, and costs, etc. Table 1 describes the elements related to the core components of cSTM and their suggested R code names. For a more detailed description of the variable types, data structure, R name for all cSTM elements, please see the Supplementary Material.

Table 1

Components of a cSTM with their R name.

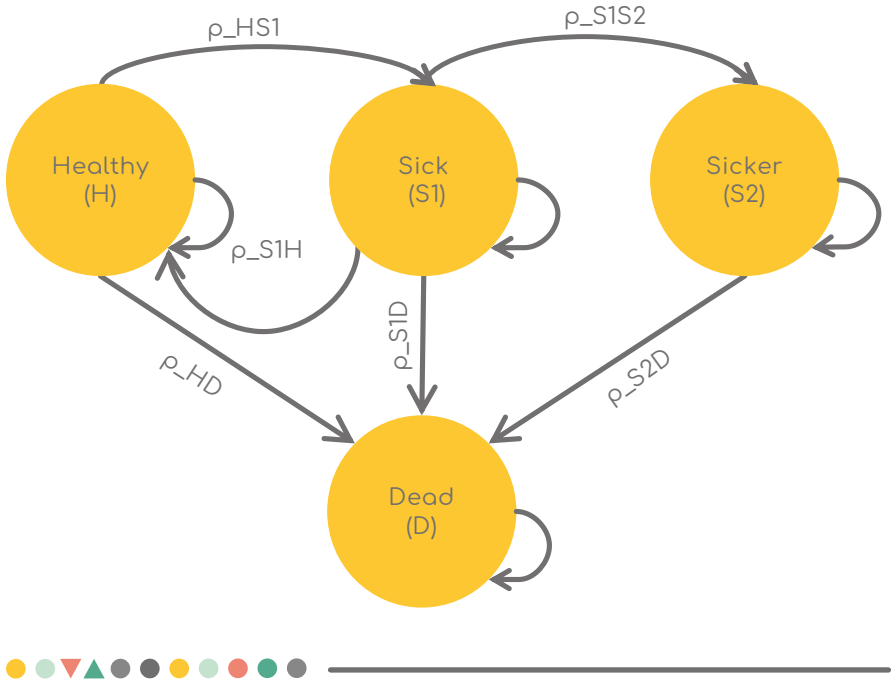
Element	Description	R name
n_S	Number of states	`n_states`
$\boldsymbol{m}_0$	Initial state vector	`v_m_init`
$\boldsymbol{m}_t$	State vector in cycle t	`v_mt`
$\boldsymbol{M}$	Cohort trace matrix	`m_M`
$\boldsymbol{P}$	Time-independent transition probability matrix	`m_P`

Case study: Sick-Sicker model

Here, we use the previously published 4-state "Sick-Sicker" model for conducting a CEA of multiple strategies to illustrate the various aspects of cSTM implementation in R.[6,19] Figure 1 represents the state-transition diagram of the Sick-Sicker model.

The model simulates a cohort at risk of a hypothetical disease with two stages, "Sick" and "Sicker", to compute the expected costs and quality-adjusted life years (QALYs) of the cohort over time. All the parameters of the Sick-Sicker model and the corresponding R variable names are presented in Table 2. The naming of these parameters and variables follows the notation described in the DARTH coding framework.[20] Briefly, we define variables by `<x>_<y>_<var_name>`, where `x` is the prefix that indicates the data type (e.g., scalar (no prefix), `v` for vector, `m` for matrix, `a` for array, `df` for data frame, etc.), `y` is the prefix indicating variable type (e.g., `p` for probability, `r` for rate, `hr` for hazard ratio, `lor` for log-odds ratio, `c` for cost, `u` for utility, etc.), and `var_name` is some description of the variable presented separated by underscores. For example, `v_p_HD` denotes the vector of transition probabilities from health state "H" to health state "D". In later sections we will define and name all the other parameters.

Figure 1



State-transition diagram of the time-independent Sick-Sicker cohort state-transition model, showing all possible states (labeled with state names) and transitions (labeled with transition probability variable names).

In this model, we simulate a hypothetical cohort of 25-year-olds in the "Healthy" state (denoted "H") until they reach a maximum age of 100 years. We will simulate the cohort dynamics in annual cycle lengths, requiring a total of 75 one-year cycles. The total number of cycles is denoted as n_T and defined in R as `n_cycles.` The model setup is as follows. Healthy individuals are at risk of developing the disease when they transition to the "Sick" state (denoted by "S1"). Sick individuals are at risk of further progressing to a more severe disease stage, the "Sicker" health state (denoted by "S2"). Individuals in S1 can recover and return to H. However, once individuals reach S2, they cannot recover; the probability of transitioning to S1 or H from S2 is zero. Individuals in H face constant background mortality. Individuals in S1 and S2 face an increased hazard of death, compared to healthy individuals, in the form of a hazard ratio (HR) of 3 and 10, respectively, relative to the background mortality rate. Individuals in S1 and S2 also experience increased healthcare costs and reduced quality of life (QoL) compared to individuals in H. When individuals die, they transition to the absorbing "Dead" state (denoted by "D"), where they remain. All transitions between non-death states are assumed to be conditional on surviving each cycle. We discount both costs and QALYs at an annual rate of 3%.

We are interested in evaluating the cost-effectiveness of four strategies: the standard of care (strategy SoC), strategy A, strategy B, and a combination of strategies A and B (strategy AB). Strategy A involves administering treatment A that increases the QoL of individuals in S1 from 0.75 (utility without treatment, `u_S1`) to 0.95 (utility with treatment A, `u_trtA`). Treatment A costs $12,000 per year (`c_trtA`).[6] This strategy does not impact the QoL of individuals in S2, nor does it change the risk of becoming sick or progressing through the sick states. Strategy B uses treatment B to reduce only the rate of Sick individuals progressing to the Sicker state with a hazard ratio (HR) of 0.6 (`hr_S1S2_trtB`), costs $13,000 per year (`c_trtB`), and does not affect QoL. Strategy AB involves administering both treatments A and B.

We assume that it is not possible to distinguish between Sick and Sicker patients; therefore, individuals in both disease states receive the treatment under the treatment strategies. After comparing the four strategies in terms of expected QALYs and costs, we calculate the incremental cost per QALY gained between non-dominated strategies.

Table 2

Description of parameters, their R variable name, base-case values and distribution.

Parameter	R name	Base-case	Distribution
Number of cycles (n_T)	`n_cycles`	75 years	-
Names of health states (n)	`v_names_states`	H, S1, S2, D	-
Annual discount rate for costs	`d_c`	3%	-
Annual discount rate for QALYs	`d_e`	3%	-
Number of PSA samples (K)	`n_sim`	1,000	-
Annual transition probabilities conditional on surviving			
- Disease onset (H to S1)	`p_HS1`	0.150	beta(30, 170)
- Recovery (S1 to H)	`p_S1H`	0.500	beta(60, 60)
- Disease progression (S1 to S2)	`p_S1S2`	0.105	beta(84, 716)
Annual mortality			
- Background mortality rate (H to D)	`r_HD`	0.002	
- Hazard ratio of death in S1 vs H	`hr_S1`	3.0	lognormal(log(3.0), 0.01)
- Hazard ratio of death in S2 vs H	`hr_S2`	10.0	lognormal(log(10.0), 0.02)
Annual costs			
- Healthy individuals	`c_H`	$2,000	gamma(100.0, 20.0)
- Sick individuals in S1	`c_S1`	$4,000	gamma(177.8, 22.5)
- Sick individuals in S2	`c_S2`	$15,000	gamma(225.0, 66.7)
- Dead individuals	`c_D`	$0	-
Utility weights			
- Healthy individuals	`u_H`	1.00	beta(200, 3)
- Sick individuals in S1	`u_S1`	0.75	beta(130, 45)
- Sick individuals in S2	`u_S2`	0.50	beta(230, 230)
- Dead individuals	`u_D`	0.00	-
Treatment A cost and effectiveness			
- Cost of treatment A, additional to state-specific healthcare costs	`c_trtA`	$12,000	gamma(576.0, 20.8)
- Utility for treated individuals in S1	`u_trtA`	0.95	beta(300, 15)
Treatment B cost and effectiveness			
- Cost of treatment B, additional to state-specific healthcare costs	`c_trtB`	$12,000	gamma(676.0, 19.2)
- Reduction in rate of disease progression (S1 to S2) as hazard ratio (HR)	`hr_S1S2_trtB`	log(0.6)	lognormal(log(0.6), 0.1)

The following sections include R code snippets. All R code is stored as a GitHub repository and can be accessed at https://github.com/DARTH-git/cohort-modeling-tutorial-intro. We initialize the input parameters in the R code below by setting the variables to their base-case values. We do this process as the first coding step, all in one place, so the updated value will carry through the rest of the code when a parameter value changes.

```r
## General setup
cycle_length <- 1          # cycle length equal one year
n_age_init   <- 25         # age at baseline
n_age_max    <- 100        # maximum age of follow up
n_cycles <- n_age_max - n_age_init # number of cycles
v_names_states <- c("H", "S1", "S2", "D") # the 4 health states of the model:
                           # Healthy (H), Sick (S1), Sicker (S2), Dead (D)

n_states <- length(v_names_states) # number of health states
d_e <- 0.03                # discount rate for QALYs of 3% per cycle
d_c <- 0.03                # discount rate for costs of 3% per cycle
v_names_str <- c("Standard of care", # store the strategy names
                 "Strategy A",
                 "Strategy B",
                 "Strategy AB")

## Transition probabilities (per cycle), hazard ratios and odds ratio (OR)
r_HD    <- 0.002           # constant rate of dying when Healthy (all-cause
                           mortality rate)
p_HS1   <- 0.15            # probability of becoming Sick when Healthy
p_S1H   <- 0.5             # probability of becoming Healthy when Sick
p_S1S2  <- 0.105           # probability of becoming Sicker when Sick
hr_S1   <- 3               # hazard ratio of death in Sick vs Healthy
hr_S2   <- 10              # hazard ratio of death in Sicker vs Healthy
```

```
# Effectiveness of treatment B
hr_S1S2_trtB <- 0.6        # hazard ratio of becoming Sicker when Sick under
                           treatment B
## State rewards
## Costs
c_H    <- 2000             # cost of being Healthy for one cycle
c_S1   <- 4000             # cost of being Sick for one cycle
c_S2   <- 15000            # cost of being Sicker for one cycle
c_D    <- 0                # cost of being dead for one cycle
c_trtA <- 12000            # cost of receiving treatment A for one cycle
c_trtB <- 13000            # cost of receiving treatment B for one cycle
# Utilities
u_H    <- 1                # utility of being Healthy for one cycle
u_S1   <- 0.75             # utility of being Sick for one cycle
u_S2   <- 0.5              # utility of being Sicker for one cycle
u_D    <- 0                # utility of being dead for one cycle
u_trtA <- 0.95             # utility when receiving treatment A for one cycle
```

To compute the background mortality risk, `p_HD`, from the background mortality rate for the same cycle length (i.e., `cycle_length` = 1), we apply Equation (1) to `r_HD`. To compute the mortality risks of the cohort in S1 and S2, we multiply the background mortality rate `r_HD` by the hazard ratios `hr_S1` and `hr_S2`, respectively, and then convert back to probabilities using Equation (1). These calculations are required because hazard ratios only apply to rates and not to probabilities. The code below performs the computation in R. In the `darthtools` package (https://github.com/DARTH-git/darthtools), we provide R functions that compute transformations between rates and probabilities since these transformations are frequently used.

```
## Mortality rates
r_S1D <- r_HD * hr_S1                      # rate of dying when Sick
r_S2D <- r_HD * hr_S2                      # rate of dying when Sicker
## Probabilities of dying
cycle_length <- 1
p_HD  <- 1 - exp(-r_HD * cycle_length)     # background mortality risk (i.e.,
                                           probability)
p_S1D <- 1 - exp(-r_S1D * cycle_length)    # probability of dying when Sick
p_S2D <- 1 - exp(-r_S2D * cycle_length)    # probability of dying when Sicker
```


To compute the risk of progression from S1 to S2 under treatment B, we first transform `p_S1S2` to a rate, `r_S1S2`, using Equation (2). Then, we multiply the hazard ratio of treatment B by the rate of progressing from S1 to S2 and transform it back to probabilities by applying Equation (1).

```
## Transition probability of becoming Sicker when Sick for treatment B
# transform probability to rate
r_S1S2 <- -log(1 - p_S1S2)/cycle_length
# apply hazard ratio to rate to obtain transition rate of becoming Sicker when
Sick
# for treatment B
r_S1S2_trtB <- r_S1S2 * hr_S1S2_trtB
# transform rate to probability
# probability to become Sicker when Sick
# under treatment B conditional on surviving
p_S1S2_trtB <- 1 - exp(-r_S1S2_trtB * cycle_length)
```

For the Sick-Sicker model, the entire cohort starts in the H state. Therefore, we create the $1 \times n_s$ initial state vector `v_m_init` with all of the cohort assigned to the H state:

```
v_m_init <- c(H = 1, S1 = 0, S2 = 0, D = 0)   # initial state vector
```

The variable `v_m_init` is used to initialize M represented by `m_M` for the cohort under strategy SoC. We also create a trace for each of the other treatment-based strategies.

```
## Initialize cohort trace for SoC
m_M <- matrix(NA,
              nrow = (n_cycles + 1), ncol = n_states,
              dimnames = list(0:n_cycles, v_names_states))
# Store the initial state vector in the first row of the cohort trace
m_M[1, ] <- v_m_init
## Initialize cohort trace for strategies A, B, and AB
# Structure and initial states are the same as for SoC
m_M_strA  <- m_M           # Strategy A
m_M_strB  <- m_M           # Strategy B
m_M_strAB <- m_M           # Strategy AB
```

Note that the initial state vector, `v_m_init`, can be modified to account for the cohort's distribution across the states at the start of the simulation. This distribution can also vary by strategy if needed.

Since the Sick-Sicker model consists of 4 states, we create a 4 × 4 transition probability matrix for strategy SoC, `m_P`. We initialize the matrix with default values of zero for all transition probabilities and then populate it with the corresponding transition probabilities. To access an element of `m_P`, we specify first the row name (or number) and then the column name (or number) separated by a comma. For example, we could access the transition probability from state Healthy (H) to state Sick (S1) using the corresponding row or column state-names as characters `m_P["H","S1"]`. We assume that all transitions to non-death states are conditional on surviving to the end of a cycle. Thus, we first condition on surviving by multiplying the transition probabilities times `1-p_HD`, the probability of surviving a cycle. For example, to obtain the probability of transitioning from H to S1, we multiply the transition probability from H to S1 conditional on being alive, `p_HS1` by `1-p_HD`. We create the transition probability matrix for strategy A as a copy of the SoC's transition probability matrix because treatment A does not alt er the cohort's transition probabilities.

```
## Initialize transition probability matrix for strategy SoC
m_P <- matrix(0,
              nrow = n_states, ncol = n_states,
              dimnames = list(v_names_states, v_names_states)) # add names
## Fill in matrix
# From H
m_P["H", "H"]   <- (1 - p_HD) * (1 - p_HS1)
m_P["H", "S1"]  <- (1 - p_HD) * p_HS1
m_P["H", "D"]   <- p_HD
# From S1
m_P["S1", "H"]  <- (1 - p_S1D) * p_S1H
m_P["S1", "S1"] <- (1 - p_S1D) * (1 - (p_S1H + p_S1S2))
m_P["S1", "S2"] <- (1 - p_S1D) * p_S1S2
m_P["S1", "D"]  <- p_S1D
# From S2
m_P["S2", "S2"] <- 1 - p_S2D
m_P["S2", "D"]  <- p_S2D
# From D
m_P["D", "D"]   <- 1

## Initialize transition probability matrix for strategy A as a copy of SoC's
m_P_strA <- m_P
```

Because treatment B alters progression from S1 to S2, we created a different transition probability matrix to model this treatment, `m_P_strB`. We initialize `m_P_strB` as a copy of `m_P` and update only the transition probabilities from S1 to S2 (i.e., `p_S1S2` is replaced with `p_S1S2_trtB`). Strategy AB also alters progression from S1 to S2 because it uses treatment B, so we create this strategy's transition probability matrix as a copy of the transition probability matrix of strategy B.

```
## Initialize transition probability matrix for strategy B
m_P_strB <- m_P
## Update only transition probabilities from S1 involving p_S1S2
m_P_strB["S1", "S1"] <- (1 - p_S1D) * (1 - (p_S1H + p_S1S2_trtB))
m_P_strB["S1", "S2"] <- (1 - p_S1D) * p_S1S2_trtB

## Initialize transition probability matrix for strategy AB as a copy of B's
m_P_strAB <- m_P_strB
```

Once all transition matrices are created, we verify they are valid by checking that each row sums to one and that each entry is between 0 and 1 using the functions `check_sum_of_transition_array` and `check_transition_probability`, respectively. These functions are part of the `darthtools` package and have been described previously.[20]

```
### Check if transition probability matrices are valid
## Check that transition probabilities are [0, 1]
check_transition_probability(m_P)
check_transition_probability(m_P_strA)
check_transition_probability(m_P_strB)
check_transition_probability(m_P_strAB)
```

```
## Check that all rows sum to 1
check_sum_of_transition_array(m_P,      n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(m_P_strB, n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(m_P_strA, n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(m_P_strAB, n_states = n_states, n_cycles = n_
cycles)
```

Next, we obtain the cohort distribution across the 4 states over 75 cycles using a time-independent cSTM under all four strategies. To achieve this, we iteratively compute the matrix product between each of the rows of `m_M` and `m_P`, and between `m_M_strB` and `m_P_strB`, respectively, using the `%*%` symbol in R at each cycle using a `for` loop

```
# Iterative solution of time-independent cSTM
for(t in 1:n_cycles){
  # For SoC
  m_M[t + 1, ] <- m_M[t, ] %*% m_P
  # For strategy A
  m_M_strA[t + 1, ] <- m_M_strA[t, ] %*% m_P_strA
  # For strategy B
  m_M_strB[t + 1, ] <- m_M_strB[t, ] %*% m_P_strB
  # For strategy AB
  m_M_strAB[t + 1, ] <- m_M_strAB[t, ] %*% m_P_strAB
}
```

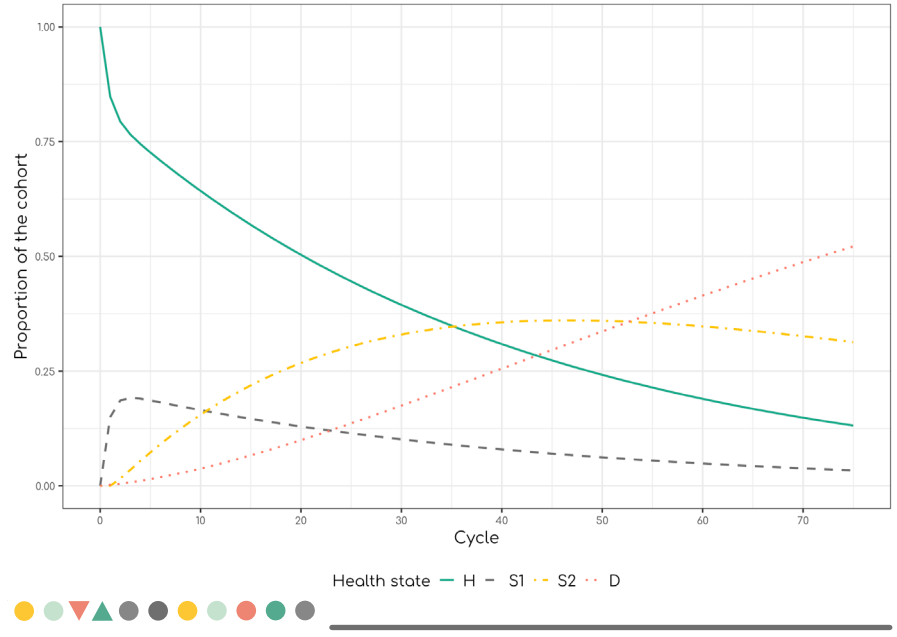
Table 3 shows the cohort trace matrix M of the Sick-Sicker model under strategies SoC and A for the first six cycles. The whole cohort starts in the H state and transitions to the rest of the states over time. Given that the D state is absorbing, the proportion in this state increases over time. A graphical representation of the cohort trace for all the cycles is shown in Figure 2.

Table 3

The distribution of the cohort under strategies SoC and A for the first six cycles of the time-independent Sick-Sicker model. The first row, labeled with cycle 0, contains the distribution of the cohort at time zero.

Cycle	H	S1	S2	D
0	1.000	0,000	1.000	0,000
1	0.848	0.150	0.000	0.002
2	0.794	0.186	0.016	0.005
3	0.766	0.192	0.035	0.008
4	0.745	0.190	0.054	0.011
5	0.726	0.186	0.073	0.015

Figure 2



Cohort trace of the time-independent cSTM under strategies SoC and A.

Cost and effectiveness outcomes

We are interested in computing the total QALYs and costs accrued by the cohort over a predefined time horizon for a CEA. In the advanced cSTM tutorial,[17] we describe how to compute epidemiological outcomes from cSTMs, such as survival, prevalence, and life expectancy.[5] These epidemiological outcomes are often used to produce other measures of interest for model calibration and validation.

State rewards

A state reward refers to a value assigned to individuals for being in a given health state. These could be either utilities or costs associated with remaining in a specific health state for one cycle in a CEA context. The column vector $\mathbf{y}$ of size $n_T + 1$ can represent the total expected reward of an outcome of interest for the entire cohort at each cycle. To calculate $\mathbf{y}$, we compute the matrix product of the cohort trace matrix times a vector of state rewards $\mathbf{r}$ of the same dimension as the number of states (n_S), such that

$$\mathbf{y} = M\mathbf{r}. \tag{5}$$

For the Sick-Sicker model, we create a vector of utilities and costs for each of the four strategies considered. The vectors of utilities and costs in `R`, `v_u_SoC` and `v_c_SoC`, respectively, represent the utilities and costs corresponding with being in each of the four health states under SoC, which are shown in Table 2.

```
# Vector of state utilities under SOC
v_u_SoC <- c(H = u_H, S1 = u_S1, S2 = u_S2, D = u_D)
# Vector of state costs under SoC
v_c_SoC <- c(H = c_H, S1 = c_S1, S2 = c_S2, D = c_D)
```

We account for the benefits and costs of both treatments individually and their combination to create the state-reward vectors under treatments A and B (strategies A and B, respectively) and when applied jointly (strategy AB). Only treatment A affects QoL, so we create a vector of utilities specific to strategies involving treatment A (strategies A and AB), `v_u_strA` and `v_u_strAB`. These vectors will have the same utility weights as for strategy SoC except for being in S1. We assign the utility associated with the benefit of treatment A in that state, `u_trtA`. Treatment B does not affect QoL, so the vector of utilities for strategy involving treatment B, `v_u_strB`, is the same as for SoC.

```
# Vector of state utilities for strategy A
v_u_strA <- c(H = u_H, S1 = u_trtA, S2 = u_S2, D = u_D)
# Vector of state utilities for strategy B
v_u_strB <- c(H = u_H, S1 = u_S1, S2 = u_S2, D = u_D)
# Vector of state utilities for strategy AB
v_u_strAB <- c(H = u_H, S1 = u_trtA, S2 = u_S2, D = u_D)
```

Both treatments A and B incur a cost. To create the vector of state costs for strategy A, `v_c_strA`, we add the cost of treatment A, `c_trtA`, to the state costs of S1 and S2. Similarly, when constructing the vector of state costs for strategy B, `v_c_str`B, we add the cost of treatment B, `c_trtB`, to the state costs of S1 and S2. Finally, for the vector of state costs for strategy AB, `v_c_strAB`, we add both treatment costs to the state costs of S1 and S2.

```
# Vector of state costs for strategy A
v_c_strA <- c(H  = c_H,
              S1 = c_S1 + c_trtA,
              S2 = c_S2 + c_trtA,
              D  = c_D)
# Vector of state costs for strategy B
v_c_strB <- c(H  = c_H,
              S1 = c_S1 + c_trtB,
              S2 = c_S2 + c_trtB,
              D  = c_D)
# Vector of state costs for strategy AB
v_c_strAB <- c(H  = c_H,
               S1 = c_S1 + (c_trtA + c_trtB),
               S2 = c_S2 + (c_trtA + c_trtB),
               D  = c_D)
```

To compute the expected QALYs and costs for the Sick-Sicker model under SoC and strategy A, we apply Equation (5) by multiplying the cohort trace matrix, `m_M`, times the corresponding strategy-specific state vectors of rewards. Similarly, to compute the expected rewards for strategies B and AB, we multiply the cohort trace matrix accounting for the effectiveness of treatment B, `m_M_strB`, times their corresponding state vectors of rewards.

```
# Vector of QALYs under SoC
v_qaly_SoC   <- m_M %*% v_u_SoC
# Vector of costs under SoC
v_cost_SoC   <- m_M %*% v_c_SoC
# Vector of QALYs for strategy A
v_qaly_strA  <- m_M_strA %*% v_u_strA
# Vector of costs for strategy A
v_cost_strA  <- m_M_strA %*% v_c_strA
# Vector of QALYs for strategy B
v_qaly_strB  <- m_M_strB %*% v_u_strB
# Vector of costs for strategy B
v_cost_strB  <- m_M_strB %*% v_c_strB
# Vector of QALYs for strategy AB
v_qaly_strAB <- m_M_strAB %*% v_u_strAB
# Vector of costs for strategy AB
v_cost_strAB <- m_M_strAB %*% v_c_strAB
```

Within-cycle correction

A discrete-time cSTM involves an approximation of continuous-time dynamics to discrete points in time. The discretization might introduce biases when estimating outcomes based on state occupancy.[21] One approach to reducing these biases is to shorten the cycle length, requiring simulating the model for a larger number of cycles, which can be computationally burdensome. Another approach is to use within-cycle corrections (WCC).[22] In this tutorial, we use Simpson's 1/3rd rule by multiplying the rewards (e.g., costs and effectiveness) by 1/3 in the first and last cycles, by 4/3 for the odd cycles, and by 2/3 for the even cycles.[23,24] We implement the WCC by generating a column vector $\mathbf{wcc}$ of size $n_T + 1$ with values corresponding to the first, $t = 0$, and last cycle, $t = n_T$, equal to 1/3, and the entries corresponding to the even and odd cycles with 2/3 and 4/3, respectively.

$$\mathbf{wcc} = \left[\frac{1}{3}, \frac{2}{3}, \frac{4}{3}, \cdots, \frac{1}{3}\right]$$

Since the within-cycle correction vector is the same for costs and QALYs, only one vector (`v_wcc`) is required.

```
# First, we define two functions to identify if a number is even or odd
is_even  <- function(x) x %% 2 == 0
is_odd   <- function(x) x %% 2 != 0
## Vector with cycles
v_cycles <- seq(1, n_cycles + 1)
## Generate 2/3 and 4/3 multipliers for even and odd entries, respectively
v_wcc    <- is_even(v_cycles)*(2/3) + is_odd(v_cycles)*(4/3)
## Substitute 1/3 in first and last entries
v_wcc[1] <- v_wcc[n_cycles + 1] <- 1/3
```

Discounting future rewards

We often discount future costs and benefits by a specific rate to calculate the net present value of these rewards. We then use this rate to generate a column vector with cycle-specific discount weights $\mathbf{d}$ of size $n_T + 1$ where its t-th entry represents the discounting for cycle t

$$\mathbf{d} = \left[1, \frac{1}{(1+d)^1}, \frac{1}{(1+d)^2}, \cdots, \frac{1}{(1+d)^{n_T}}\right],$$

where d is the cycle-length discount rate. At the end of the simulation, we multiply the vector of expected rewards, $\mathbf{y}$, by a discounting column vector. The total expected discounted outcome summed over the n_T cycles, y, is obtained by the inner product between $\mathbf{y}$ transposed, $\mathbf{y}'$, and $\mathbf{d}$,

$$y = \mathbf{y}'\mathbf{d}. \tag{6}$$

The discount vectors for costs and QALYs for the Sick-Sicker model with annual cycles, `v_dwc` and `v_dwe`, respectively, are

```
# Discount weight for effects
v_dwe <- 1 / ((1 + d_e) ^ (0:(n_cycles)))
# Discount weight for costs
v_dwc <- 1 / ((1 + d_c) ^ (0:(n_cycles)))
```

To compute the total expected discounted QALYs and costs under all four strategies accounting for both discounting and WCC, we incorporate $\mathbf{wcc}$ in equation (6) using an element-wise multiplication with $\mathbf{d}$, indicated by the $\odot$ sign.

$$y = \mathbf{y}^{'} \left(\mathbf{d} \odot \mathbf{wcc} \right). \tag{7}$$

To compute the total expected discounted and WCC-corrected QALYs under all four strategies in R, we apply Equation (7) to the reward vectors of each strategy.

```
## Expected discounted QALYs under SoC
n_tot_qaly_SoC   <- t(v_qaly_SoC)   %*% (v_dwe * v_wcc)
## Expected discounted costs under SoC
n_tot_cost_SoC   <- t(v_cost_SoC)   %*% (v_dwc * v_wcc)
## Expected discounted QALYs for strategy A
n_tot_qaly_strA  <- t(v_qaly_strA)  %*% (v_dwe * v_wcc)
## Expected discounted costs for strategy A
n_tot_cost_strA  <- t(v_cost_strA)  %*% (v_dwc * v_wcc)
## Expected discounted QALYs for strategy B
n_tot_qaly_strB  <- t(v_qaly_strB)  %*% (v_dwe * v_wcc)
## Expected discounted costs for strategy B
n_tot_cost_strB  <- t(v_cost_strB)  %*% (v_dwc * v_wcc)
## Expected discounted QALYs for strategy AB
n_tot_qaly_strAB <- t(v_qaly_strAB) %*% (v_dwe * v_wcc)
## Expected discounted costs for strategy AB
n_tot_cost_strAB <- t(v_cost_strAB) %*% (v_dwc * v_wcc)
```


Table 4

Total expected discounted QALYs and costs per average individual in the cohort of the Sick-Sicker model by strategy accounting for within-cycle correction.

	Costs	QALYs
Standard of care	$148,657	20.990
Strategy A	$275,937	21.717
Strategy B	$248,571	22.482
Strategy AB	$361,341	23.354

The total expected discounted QALYs and costs for the Sick-Sicker model under the four strategies accounting for within-cycle correction are shown in Table 4.

Incremental cost-effectiveness ratios (ICERs)

We combine the total expected discounted costs and QALYs for all four strategies into outcome-specific vectors, `v_cost_str` for costs and `v_qaly_str` for QALYs. So far, we have used base R to create and simulate cSTMs. For the CEA, we use the R package `dampack` (https://cran.r-project.org/web/packages/dampack/)[25] to calculate the incremental costs and effectiveness and the incremental cost-effectiveness ratio (ICER) between non-dominated strategies and create the data frame `df_cea` with this information. These outcomes are required inputs to conduct a CEA.

```
### Vector of costs
v_cost_str <- c(n_tot_cost_SoC, n_tot_cost_strA, n_tot_cost_strB, n_tot_cost_
strAB)
### Vector of effectiveness
v_qaly_str <- c(n_tot_qaly_SoC, n_tot_qaly_strA, n_tot_qaly_strB, n_tot_qaly_
strAB)

### Calculate incremental cost-effectiveness ratios (ICERs)
df_cea     <- dampack::calculate_icers(cost = v_cost_str,
                                       effect = v_qaly_str,
                                       strategies = v_names_str)
```

SoC is the least costly and effective strategy, followed by Strategy B producing an expected incremental benefit of 1.492 QALYs per individual for an additional expected cost of $99,913 with an ICER of $66,960/QALY followed by Strategy AB with an ICER $129,354/QALY. Strategy A is a dominated strategy (Table 5). Strategies SoC, B and AB form the cost-effectiveness efficient frontier (Figure 3).

Table 5

Cost-effectiveness analysis results for the Sick-Sicker model. ND: Non-dominated strategy; D: Dominated strategy.

Strategy	Costs ($)	QALYs	Incremental Costs ($)	Incremental QALYs	ICER ($/QALY)	Status
Standard of care	148,657	20.990	NA	NA	NA	ND
Strategy B	248,571	22.482	99,913	1.492	66,960	ND
Strategy AB	361,341	23.354	112,771	0.872	129,354	ND
Strategy A	275,937	21.717	NA	NA	NA	D

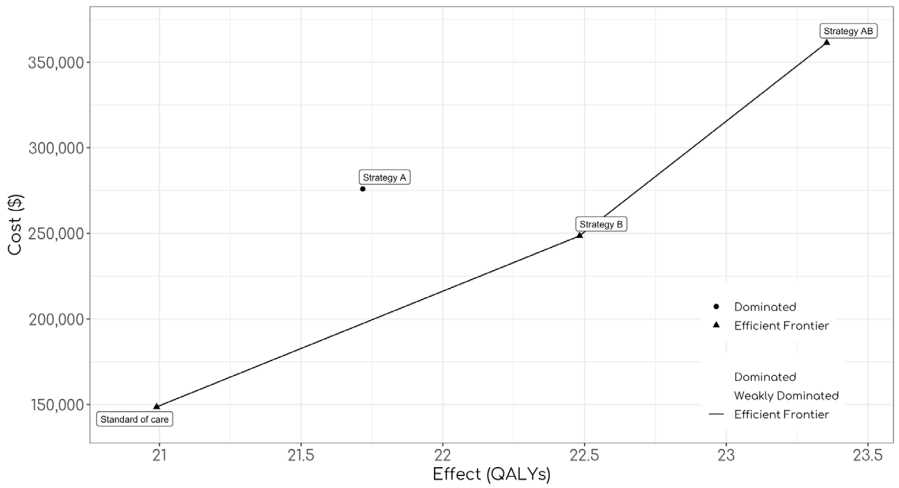
Probabilistic sensitivity analysis

To quantify the effect of model parameter uncertainty on cost-effectiveness outcomes, we conducted a probabilistic sensitivity analysis (PSA).[26] In a PSA, we randomly draw K parameter sets (`n_sim`) from distributions that reflect the current uncertainty in model parameter estimates. The distribution for all the parameters and their values are described in Table 2 and more detail in the Supplementary Material. For each sampled set of parameter values, we compute model outcomes (e.g., total discounted cost and QALYs) for each strategy. In a previously published manuscript, we describe the implementation of these steps in R.[20] Briefly, to conduct the PSA, we create three R functions:

1. `generate_psa_params(n_sim, seed)`: a function that generates a sample of size `n_sim` for the model parameters, `df_psa_input`, from their distributions defined in Table 2. The function input `seed` sets the seed of the pseudo-random number generator used in sampling parameter values, which ensures reproducibility of the PSA results.
2. `decision_model(l_params_all, verbose = FALSE)`: a function that wraps the R code of the time-independent cSTM described in the section "Time-independent cSTM dynamics". This function requires inputting a list of all model parameter values, `l_params_all` and whether the user wants print messages on whether transition probability matrices are valid via the verbose parameter.
3. `calculate_ce_out(l_params_all, n_wtp = 100000)`: a function that calculates total discounted costs and QALYs based on the `decision_model` function output. This function also computes the net monetary benefit (NMB) for a given willingness-to-pay threshold, specified by the argument `n_wtp`. These functions are provided in the accompanying GitHub repository of this manuscript.

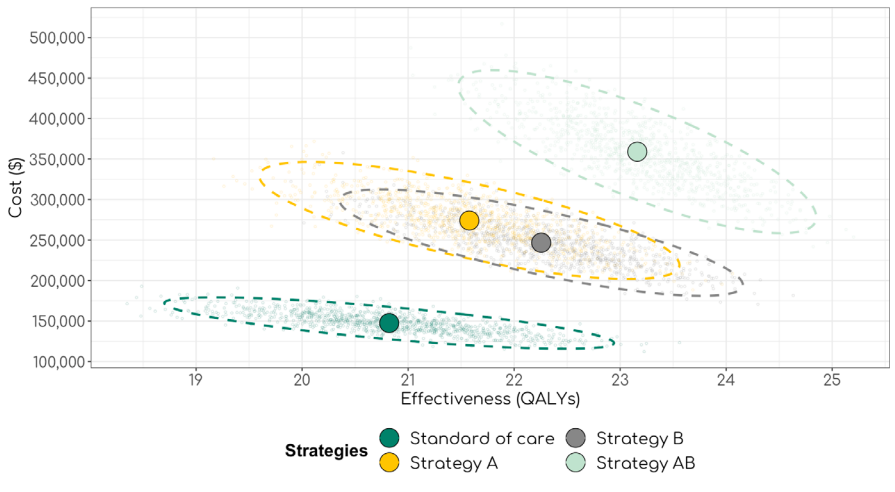
To conduct the PSA of the CEA using the time-independent Sick-Sicker cSTM, we sampled 1,000 parameter sets from their distributions. For each sampled parameter set, we simulated the cost and effectiveness of each strategy. Results from a PSA can be represented in various ways. For example, the joint distribution, 95% confidence ellipse, and the expected values of the total discounted costs and QALYs for each strategy can be plotted in a cost-effectiveness scatter plot (Figure 4),[27] where each of the 1,000 simulations are plotted as a point in the graph. The CE scatter plot for CEA using the time-independent model shows that strategy AB has the highest expected costs and QALYs. Standard of care has the lowest expected cost and QALYs. Strategy B is more effective and least costly than Strategy A. Strategy A is a strongly dominated strategy.

Figure 3



Cost-effectiveness efficient frontier of the cost-effectiveness analysis based on the time-independent Sick-Sicker model.

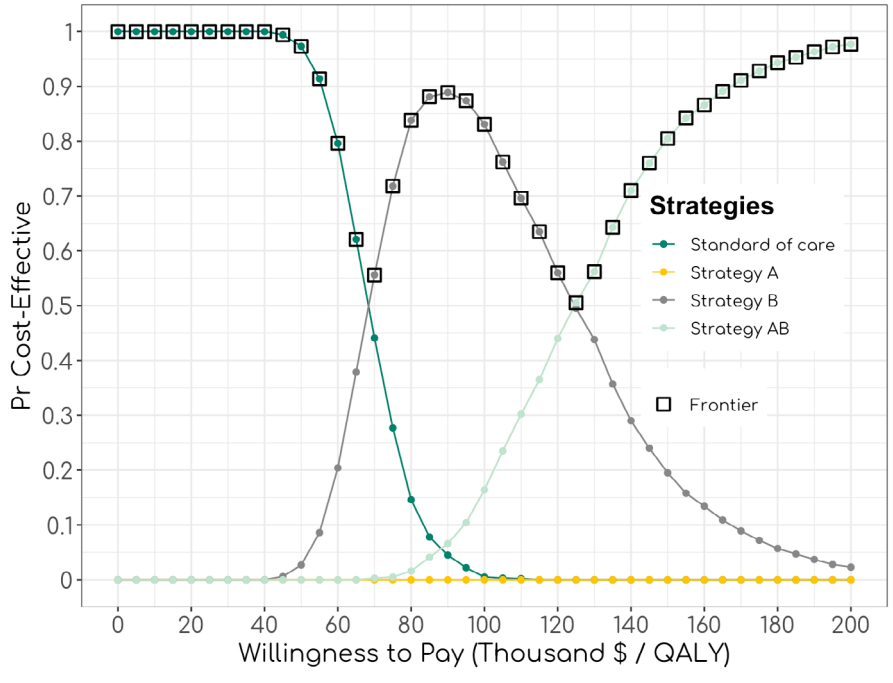
Figure 4



Cost-effectiveness scatter plot.

Figure 5 presents the cost-effectiveness acceptability curves (CEACs), which show the probability that each strategy is cost-effective, and the cost-effectiveness frontier (CEAF), which shows the strategy with the highest expected net monetary benefit (NMB), over a range of willingness-to-pay (WTP) thresholds. Each strategy's NMB is computed using NMB=QALY×WTP-Cost[28] for each PSA sample. At WTP thresholds less than $70,000 per QALY gained, strategy SoC has both the highest probability of being cost-effective and the highest expected NMB. This switches to strategy B for WTP thresholds between $70,000 and $125,000 per QALY gained and to strategy AB for WTP thresholds greater than or equal to $125,000 per QALY gained.

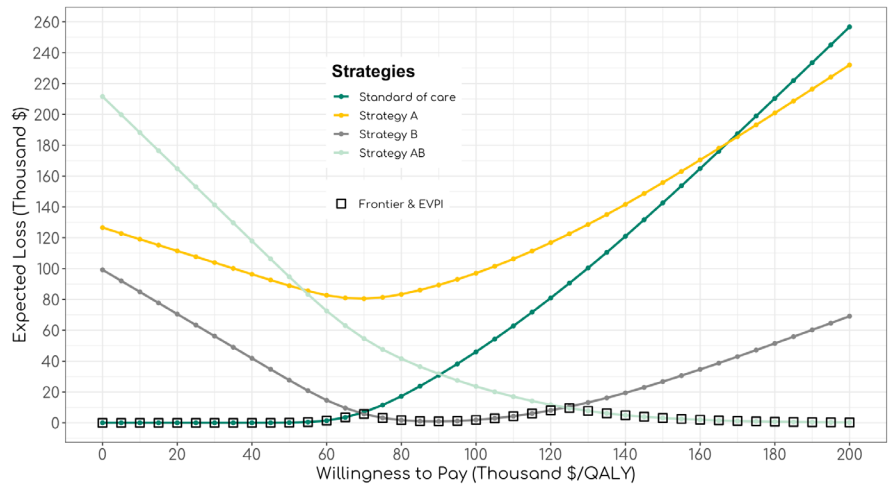
Figure 5



Cost-effectiveness acceptability curves (CEACs) and frontier (CEAF).

The CEAC and CEAF do not show the magnitude of the expected net benefit lost (i.e., expected loss) when the chosen strategy is not the cost-effective strategy in all the samples of the PSA. To complement these results, we quantify expected loss from each strategy over a range of WTP thresholds with the expected loss curves (ELCs). These curves quantify the expected loss from each strategy over a range of WTP thresholds (Figure 6). The expected loss considers both the probability of making the wrong decision and the magnitude of the loss due to this decision, representing the foregone benefits of choosing a suboptimal strategy. The expected loss of the optimal strategy represents the lowest envelope of the ELCs because, given current information, the loss cannot be minimized further. The lower envelope also represents the expected value of perfect information (EVPI), which quantifies the value of eliminating parameter uncertainty. At a WTP threshold of $125,000 per QALY, the EVPI is highest at $9,577. For a more detailed description of CEAC, CEAF, ELC and EVPI interpretations and the R code to generate them, we refer the reader to previously published literature.[29]

Figure 6



Expected loss curves (ELCs) and expected value of perfect information (EVPI).

Discussion

In this introductory tutorial, we provided a step-by-step mathematical conceptualization of time-independent cSTMs and a walk-through of their implementation in R using a hypothetical disease example with accompanying code throughout the tutorial. We used R as the programming language of choice because it is open-source, meaning that the source code for every function in R is freely available. In addition, R has a vast and increasing number of packages that can assist modelers in conducting most of their analyses in the same environment.

The parameterization of our example model assumes all parameters are known, or at least, the characterization of their uncertainty is known (i.e., we know their distributions). However, to construct a real-world cSTM, modelers must conduct a thorough synthesis of current evidence to determine these appropriate structures and inform all parameters based on the current evidence. For example, literature must be carefully considered when determining whether transitions between non-death health states are estimated conditional on being alive or are estimated as competing risks along with mortality risks.[26] Similarly, our PSA analysis simplifies reality where all model parameters are assumed to be independent of each other. However, parameters could be correlated or have a rank order, and appropriate statistical methods that simulate these correlations or rank order might be needed.[30] We encourage modelers to use appropriate statistical methods to synthesize and quantify model parameters' uncertainty accurately. In addition, modelers should appropriately specify all model parameters for the cycle length of the model.[22]

In general, cSTMs are recommended when the number of states is considered "not too large".[5] This recommendation arises because it becomes more challenging to keep track of their construction as the number of states increases. It is possible to build reasonably complex cSTMs in R as long as the computer's RAM can store the size of the transition probability matrix and outputs of interest. For time-independent cSTMs, in general, this should not be a problem with the capacity of current RAM in personal computers. With increasing model complexity and interdependency of functions to conduct various

analyses like PSA, it is essential to ensure all code and functions work as expected and all elements of the cSTM are valid. We can achieve this by creating functions that help with model debugging, validation, and thorough unit testing. In the accompanying GitHub repository, we provide functions to check that transition probability matrices and their elements are valid. These functions are an example of a broader standard practice in software development called unit testing that requires building functions to test and check that the model and model-based analysis perform as intended.[31] However, unit testing is beyond the scope of this tutorial. We refer the reader to a previously published manuscript that describes unit testing in more detail and provides accompanying code.[20]

In this tutorial, we implemented a cSTM using a (discrete-time) transition matrix. However, cSTM can also be implemented via (discrete-time) difference equations or (continuous-time) differential equations in R.[32,33] We refer readers interested in learning more on continuous-time cSTMs to previously published manuscripts[21,34–36] and a tutorial using R.[37] Finally, the variable names used in this paper reflect our coding style. While we provide standardized variable names, adopting these conventions is ultimately a personal preference.

In summary, this tutorial provides a conceptualization of time-independent cSTMs and a step-by-step guide to implement them in R. We aim to add to the current body of literature and material on building this type of decision model so that health decision scientists and health economists can develop cSTMs in a more flexible, efficient, open-source manner and to encourage increased transparency and reproducibility. In the advanced cSTM tutorial, we explore generalizing this framework to time-dependent cSTM, generating epidemiological outcomes, and incorporating transition rewards.

Supplemental material

Supplemental tables, data and model code are provided in the GitHub repository:
https://github.com/DARTH-git/cohort-modeling-tutorial-intro

Acknowledgements

We want to thank the anonymous reviewers of Medical Decision Making for their valuable suggestions and the students who took our classes where we refined these materials.

References

1. Kuntz, KM, Russell, BL, Owens, DK, Sanders, GD, Trikalions, TA, Salomon J. Decision Models in Cost-Effectiveness Analysis. In: *Cost-Effectiveness in Health and Medicine. New York: Oxford University* Press, 2017, pp. 105–136.
2. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making.* 2017; 37(7):735–746.
3. Spedicato GA. Discrete Time Markov Chains with R. R J. 2017; 9(2):84–104.
4. Filipović-Pierucci A, Zarca K, Durand-Zaleski I. *Markov Models for Health Economic Evaluations: The R Package heemod*, http://arxiv.org/abs/1702.03252 (2017).
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. *Med Decis Making.* 2012; 32(5):690–700.
6. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making.* 2018; 38(3):400–422.
7. Suijkerbuijk AWM, Van Hoek AJ, Koopsen J, De Man RA, Mangen MJJ, De Melker HE, Polder JJ, De Wit GA, Veldhuijzen IK. Cost-effectiveness of screening for chronic hepatitis B and C among migrant populations in a low endemic country. *PLoS One.* 2018; 13(11):1–16.
8. Sathianathen NJ, Konety BR, Alarid-Escudero F, Lawrentschuk N, Bolton DM, Kuntz KM. Cost-effectiveness Analysis of Active Surveillance Strategies for Men with Low-risk Prostate Cancer. *Eur Urol.* 2019; 75(6):910–917.
9. Lu S, Yu Y, Fu S, Ren H. Cost-effectiveness of ALK testing and first-line crizotinib therapy for non-small-cell lung cancer in China. *PLoS ONE* 13(10): e0205827.
10. Djatche LM, Varga S, Lieberthal RD. Cost-Effectiveness of Aspirin Adherence for Secondary Prevention of Cardiovascular Events. *PharmacoEconomics - Open.* 2018; 2(4):371–380.

11. Smith-Spangler CM, Juusola JL, Enns EA, Owens DK, Garber AM. Population Strategies to Decrease Sodium Intake and the Burden of Cardiovascular Disease: A Cost-Effectiveness Analysis. *Ann Intern Med*. 2010; 152(8):481–487.
12. Pershing S, Enns EA, Matesic B, Owens DK, Goldhaber-Fiebert JD. Cost-Effectiveness of Treatment of Diabetic Macular Edema. *Ann Intern Med*. 2014; 160(1):18–29.
13. Miller DK, Homan SM. Determining Transition Probabilities: Confusion and Suggestions. *Med Decis Making*. 1994; 14(1):52–58.
14. Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond MF, McGuire A (eds) *Economic Evaluation in Health Care: Merging Theory with Practice*. New York, NY: Oxford University Press, 2001, pp. 141–171.
15. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making*. 1993; 13(4):322–38.
16. Beck JR, Pauker SG. The Markov Process in Medical Prognosis. *Med Decis Making*. 1983; 3(4):419–458.
17. Alarid-Escudero F, Krijkamp EM, Enns EA, Yang A, Hunink MGM, Pechlivanoglou P, Jalal H. A Tutorial on Time-Dependent Cohort State-Transition Models in R using a Cost-Effectiveness Analysis Example. *arXiv Prepr*., http://arxiv.org/abs/2108.13552 (2021).
18. Iskandar R. A theoretical foundation for state-transition cohort models in health decision analysis. *PLoS One*. 2018; 13(12):e0205543.
19. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying Best-Fitting Inputs in Health-Economic Model Calibration: A Pareto Frontier Approach. *Med Decis Making*. 2015; 35(2):170–182.
20. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics*. 2019; 37(11):1329–1339.
21. van Rosmalen J, Toy M, O'Mahony JF. A mathematical approach for evaluating Markov models in continuous time without discrete-event simulation. *Med Decis Making*. 2013; 33(6):767–779.
22. Hunink, M., Weinstein, M., Wittenberg, E., Drummond, M., Pliskin, J., Wong, J., & Glasziou, P. (2014). *Decision Making in Health and Medicine*: Integrating Evidence and Values (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139506779
23. Elbasha EH, Chhatwal J. Theoretical foundations and practical applications of within-cycle correction methods. *Med Decis Making*. 2016; 36(1):115–131.
24. Elbasha EH, Chhatwal J. Myths and misconceptions of within-cycle correction: a guide for modelers and decision makers. *Pharmacoeconomics*. 2016; 34(1):13–22.
25. Alarid-Escudero F, Knowlton G, Easterly CA, Enns EA. Decision Analytic Modeling Package (dampack), https://cran.r-project.org/web/packages/dampack/ (2021).
26. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012; 32(5):722–732.
27. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making*. 2002; 22(4):290–308.
28. Stinnett AA, Mullahy J. Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. *Med Decis Making*. 1998; 18(2_suppl):S68–S80.
29. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. "Time Traveling Is Just Too Dangerous" But Some Methods Are Worth Revisiting: The Advantages of Expected Loss Curves Over Cost-Effectiveness Acceptability Curves and Frontier. Supplement. *Value Health*. 2019; 22(5):611–618.
30. Goldhaber-Fiebert JD, Jalal HJ. Some Health States Are Better Than Others: Using Health State Rank Order to Improve Probabilistic Analyses. *Med Decis Making*. 2015; 36(8):927–940.
31. Wickham H, Bryan J. Testing. *R Packages. Organize, test, document and share your code*. Sebastopol, CA: O'Reilly Media, 2021, pp. 1–13.
32. Grimmett G, Welsh D. Markov Chains. *Probability*: *An Introduction*. Oxford University Press, 2014, p. Chapter 12.
33. Axler S, Gehring FW, Ribet KA. Difference Equations. New York, NY: *Springer*. http://link.springer.com/10.1007/0-387-27645-9 (2005)
34. Cao Q, Buskens E, Feenstra T, Jaarsma T, Hillege H, Postmus D. Continuous-Time Semi-Markov Models in Health Economic Decision Making: An Illustrative Example in Heart Failure Disease Management. *Med Decis Making*. 2016; 36(1):59–71.
35. Begun A, Icks A, Waldeyer R, Landwehr S, Koch M, Giani G. Identification of a multistate continuous-time nonhomogeneous Markov chain model for patients with decreased renal function. *Med Decis Making*. 2013; 33(2):298–306.
36. Soares MO, Canto E Castro L. Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics*. 2012; 30(12):1101–1117.
37. Frederix GWJ, van Hasselt JGC, Severens JL, Hövels AM, Huitema ADR, Raaijmakers J a M, Schellens JHM. Development of a framework for cohort simulation in cost-effectiveness analyses using a multistep ordinary differential equation solver algorithm in R. *Med Decis Making*. 2013; 33(6):780–792.

05

A TUTORIAL ON TIME-DEPENDENT COHORT STATE-TRANSITION MODELS IN R USING A COST-EFFECTIVENESS ANALYSIS EXAMPLE

Fernando Alarid-Escudero
Eline Krijkamp
Eva A. Enns
Alan Yang
M.G. Myriam Hunink
Petros Pechlivanoglou
Hawre Jalal

Medical Decision Making – under review

A TUTORIAL ON TIME-DEPENDENT COHORT STATE-TRANSITION MODELS IN R USING A COST-EFFECTIVENESS ANALYSIS EXAMPLE

Abstract

This tutorial shows how to implement time-dependent cohort state-transition models (cSTMs) to conduct cost-effectiveness analyses (CEA) in R, where transition probabilities and rewards vary by time. We account for two types of time-dependency: time since the start of the simulation (simulation-time dependency) and time spent in a health state (state-residence dependency). We illustrate how to conduct a CEA of multiple strategies based on a time-dependent cSTM using a previously published cSTM, including probabilistic sensitivity analyses. We also demonstrate how to compute various epidemiological outcomes of interest from the outputs generated from the cSTM, such as survival probability and disease prevalence. We demonstrate both the mathematical notation and the R code to execute the calculations. This tutorial builds upon an introductory tutorial that introduces time-independent cSTMs using a CEA example in R. We provide an up-to-date public code repository for broader implementation.

"Let us postpone nothing. Let us balance life's account every day... One who daily puts the finishing touches to his life is never in want of time."

Seneca, Moral letters to Lucilius

Introduction

Cohort state-transition models (cSTMs), commonly known as Markov models, are decision models that simulate disease dynamics over time. cSTMs are widely used to evaluate various health policies and clinical strategies, such as screening and surveillance programs,[1,2] diagnostic procedures,[3] disease management programs,[4] and interventions[5,6]. The simplest cSTMs are time-independent (time-homogeneous), meaning that the transition probabilities and other model parameters remain fixed over the simulated time horizon. In an introductory tutorial, we described the implementation of time-independent cSTMs.[7] In many applications, a time-independent cSTM is limited because key parameters may vary over time. For example, background mortality changes as a cohort ages, the risk of cancer recurrence might change as a function of time since diagnosis, or the incidence of vector-borne diseases may have temporal or seasonal trends. The costs and utility of residing in a particular health state might also vary over time. For example, cancer-related healthcare costs and utility might depend on whether a patient is in their first year of remission or their fifth. Similarly, a person's last year of life is often their most expensive in health care spending. Allowing for flexibility in capturing time-varying dynamics is often essential in realistic models.

This tutorial expands the cSTM framework described in the introductory time-independent cSTM tutorial by allowing transition probabilities, costs, and utilities to vary over time and for one-time costs or utilities when the cohort experiences events when transitioning between states, often called transition rewards. We also demonstrate how to compute various epidemiological measures from cSTMs needed for model calibration or validation.

We distinguish between two types of time-dependency that require different approaches: (1) Simulation-time dependency, which represents the time since the start of the simulation, and (2) state-residence dependency, representing time spent in a health state. Simulation-time dependency affects parameters that vary with time for the entire cohort in the same way. The most common example of simulation-time dependency is age-specific background mortality over time as the cohort ages. Since all members of the cohort age at the same rate, we can implement this dependency by changing the mortality parameters as the simulation progresses.[8] Similarly, in a model simulating a cohort starting from disease diagnosis, any dependence on time since diagnosis can be implemented as a dependence on the time since simulation start. Seasonal or temporal variation in disease incidence can also be reflected through simulation-time dependency, where the risk of developing a disease can vary based on the current season or year in the simulation.

State-residence dependency captures time dependence on events that members of the cohort could experience at different times. For example, in a model simulating a cohort of healthy individuals, they may experience disease onset at different times. Thus, parameters that depend on the time since disease onset cannot be implemented based on simulation time; instead, we need to track cohort members *from their disease onset time*. We implement this type of time-dependency by expanding the model state space to include disease states that encode the time since an event has occurred. For example, instead of a single "Sick" state, individuals would transition first to the "Sick - cycle 1" state at disease onset, then "Sick - cycle 2" at the next cycle, and so on. In this way, each replicate of the "Sick" state can have different transition probabilities (e.g., mortality, risk of complications, etc.), costs, or utilities. We should note that we can also use this state expansion approach to time-dependency to model simulation-time dependency. However, as we see below, substantial state expansion can be cumbersome and potentially computationally burdensome. For simplicity, modelers should always consider simulation time-dependency first.

Besides describing the two forms of time-dependency, we also illustrate the concept and implementation of transition rewards, such as one-time costs or utilities applied when individuals experience events when transitioning between certain states. These transition

rewards reflect event-driven outcome impacts, such as a higher cost for the first cycle of disease onset due to increased diagnostic and management costs or the increased cost of transition to the dead state incurred from end-of-life interventions or management.[9]

In this tutorial, we describe how to implement simulation-time dependent cSTMs and then add state-residence dependency to account for both time dependencies. We illustrate the use of these cSTMs to conduct a cost-effectiveness analysis (CEA) in R, a statistical software with increasing use in health decision sciences.[10] We also illustrate the calculation of various epidemiological measures beyond the simple cohort trace, including survival, life expectancy, prevalence, and incidence. Some of these calculations leverage the state-transition array implementation that we will introduce, while others are manipulations of the cohort trace. Such epidemiological measures may be of interest for specific analyses but are also important for calibrating and validating the model against real-world data.

Readers can find the most up-to-date model code and code to create the tutorial graphs in the accompanying GitHub repository (https://github.com/DARTH-git/cohort-modeling-tutorial-timedep). We encourage the reader first to review the basics of decision modeling and how to develop time-independent cSTMs in R, as described in the introductory tutorial.

Simulation-time dependency

As described above, simulation-time dependency represents the time since the model starts and can be represented by defining the transition probability matrix as a function of time, P_t. The elements of P_t are the transition probabilities of moving from state i to state j in time t, $p_{[i,j,t]}$, where $\{i,j\} = 1, \cdots, n_S$ and $t = 0, \cdots, n_T$, n_S is the number of health states of the model and n_T is the number of cycles that represent total simulation time.

$$P_t = \begin{bmatrix} p_{[1,1,t]} & p_{[1,2,t]} & \cdots & p_{[1,n_S,t]} \\ p_{[2,1,t]} & p_{[2,2,t]} & \cdots & p_{[2,n_S,t]} \\ \vdots & \vdots & \ddots & \vdots \\ p_{[n_S,1,t]} & p_{[n_S,2,t]} & \cdots & p_{[n_S,n_S,t]} \end{bmatrix}.$$

Note that in each cycle t all rows of the transition probability matrix must sum to one, $\sum_{j=1}^{n_s} p_{[i,j,t]} = 1$ for all $i = 1, \cdots, n_s$ and $t = 0, \cdots, n_T$.

Next, we specify the initial distribution of the cohort at $t = 0$. We then define $\mathbf{m}_0$ as the vector that captures the distribution of the cohort among the states at $t = 0$ (i.e., the initial state vector). As illustrated in the introductory tutorial, we iteratively compute the cohort distribution among the health states in each cycle from the transition probability matrix and the cohort's distribution from the prior cycle. The state vector at the next cycle $t + 1$ ($\mathbf{m}_{t+1}$) is then calculated as the matrix product of the state vector at cycle t, $\mathbf{m}_t$, and the transition probability matrix that the cohort faces in cycle t, P_t,

$$\mathbf{m}_{t+1} = \mathbf{m}_t P_t \qquad \text{for} \qquad t = 0, \dots, (n_T - 1). \tag{1}$$

Equation (1) is iteratively evaluated until $t = n_T$.

The cohort trace matrix, M, is a matrix of dimensions $(n_T + 1) \times n_S$ where each row is a state vector ($-\mathbf{m}_t-$), such that

$$M = \begin{bmatrix} -\mathbf{m}_0 - \\ -\mathbf{m}_1 - \\ \vdots \\ -\mathbf{m}_{n_T} - \end{bmatrix}.$$

M stores the output of the cSTM, which we can use to compute various epidemiological measures, such as prevalence and survival probability over time, and economic measures, such as cumulative resource use and costs.

Time dependency on state-residence

The implementation of state-residence dependency is slightly more involved than simulation-time dependency. As described above, dependence on state-residence occurs in applications where transition probabilities or rewards depend on the time spent in a given state. To account for state-residence dependency, we expand the number of states with as many transient states as the number of cycles required for state-residency. These transient states are often called tunnel states, where the cohort stays for only one cycle in each tunnel state and either transitions to the next tunnel state or completely exits the tunnel. The total number of states for a cSTM with tunnels is $n_{S_{\text{tunnels}}}$ and equals $n_S + n_{\text{tunnel}} - 1$, where n_{tunnel} is the total number of times a health state needs to be expanded for. We subtract one because one of the original states is expanded to a tunnel state. The transition probability matrix also needs to be expanded to incorporate these additional transient states, resulting in a transition probability matrix of dimensions $n_{S_{\text{tunnels}}} \times n_{S_{\text{tunnels}}}$. If the transition probabilities are also dependent on simulation time, as described in the previous section, we add a third dimension such that the dimensions will become $n_{S_{\text{tunnels}}} \times n_{S_{\text{tunnels}}} \times n_T$. As state-residence time dependency extends in more health states, the total number of health states will increase, causing what has been referred to as "state explosion". Table 1 describes the core components of time-dependent cSTMs and their suggested R code names.

Case study: a cost-effectiveness analysis using a time-dependent Sick-Sicker model

We demonstrate simulation-time and state-residence dependency in R by expanding the time-independent 4-state "Sick-Sicker" model described in the introductory tutorial.[7] The state-transition diagram of the Sick-Sicker model is presented in Figure 1. We first modify this cSTM to account for simulation-time dependency by incorporating age-dependent background mortality and then expand it to account for state-residence dependency. We will use this cSTM to conduct a CEA of different treatment strategies accounting for transition rewards.

Table 1

Core components of time-dependent cSTMs with their R name.

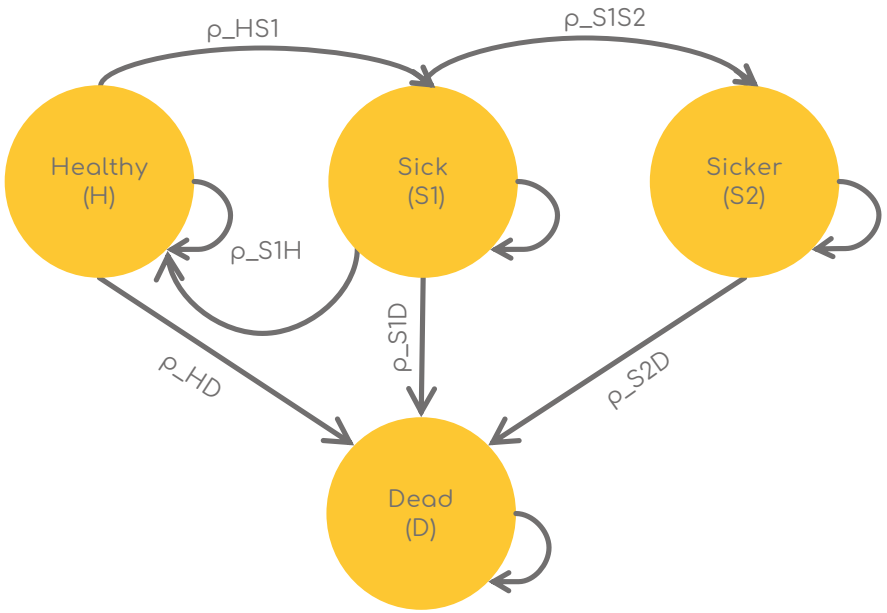
Element	Description	R name
n_S	Number of states	`n_states`
$\mathbf{m}_0$	Initial state vector	`v_s_init`
$\mathbf{m}_t$	State vector in cycle t	`v_mt`
M	Cohort trace matrix	`m_M`
$\mathbf{P}$	Time-dependent transition probability array	`a_P`
$\mathbf{A}$	Transition-dynamics array	`a_A`
n_{tunnel}	Number of tunnel states	`n_tunnel_size`
$n_{S_{\text{tunnel}}}$	Number of states including tunnel states	`n_states_tunnels`
$\mathbf{m}_{\text{tunnel}_0}$	Initial state vector for the model with tunnel states	`v_s_init_tunnels`

For a more detailed description of these components with their variable types, data structure, and R name, please see the Supplementary Materials.

As described in the introductory tutorial, this model simulates the healthcare costs and quality-adjusted life-years (QALYs) of a cohort of 25-year-old individuals at risk of developing a hypothetical disease with two stages: "Sick" and "Sicker".[11] We simulate the cohort of individuals who all start in the "Healthy" state (denoted by "H") over their lifetime, which means that we will simulate the cohort for 75 cycles. The total number of cycles is denoted as n_T and defined in R as `n_cycles`.

In the introductory tutorial, healthy individuals face a constant risk of death. To illustrate simulation-time dependency, here we assume that healthy individuals face age-specific background mortality. Healthy individuals who survive might become ill over time and transition to the "Sick" state (denoted by "S1"). In this tutorial, we consider that once healthy individuals get sick, they incur a one-time utility decrement of 0.01 (`du_HS1`, a disutility of transitioning from H to S1) and a transition cost of $1,000 (`ic_HS1`) that reflects the immediate costs of developing the illness. We demonstrate how to include these in the section transition rewards.

Figure 1



State-transition diagram of the Sick-Sicker cohort state-transition model. The circles represent the health states, and the arrows represent the possible transition probabilities. The labels next to the arrows represent the variable names for these transitions.

The S1 state is associated with higher mortality, higher healthcare costs, and lower quality of life than the H state. Once in the S1 state, individuals may recover (returning to the H state), die (move to the D state), or progress further to the "Sicker" state (denoted by "S2"), with further increases in mortality risk and healthcare costs and reduced quality of life. We assume that it is not clinically possible to distinguish between individuals in S1 and S2. Individuals in S1 and S2 face an increased hazard of death, compared to healthy individuals, in the form of a hazard ratio (HR) of 3 and 10, respectively, relative to the background age-specific mortality hazard rate. In the state-residence dependent model, the risk of progressing from S1 to S2 depends on the time spent in S1. Once simulated individuals die, they transition to the absorbing D state, where they remain, and incur a one-time cost of $2,000 (`ic_D`) for the expected acute care received before dying. All transitions between non-death states are assumed to be conditional on surviving each cycle. We simulated the evolution of the cohort in one-year discrete-time cycles.

We use this cSTM to evaluate the cost-effectiveness of four strategies: Strategy A, strategy B, a combination of A and B (strategy AB), and the standard of care (strategy SoC). Strategy A involves administering treatment A to individuals in S1 and S2. Treatment A increases the QoL of individuals only in S1 from 0.75 (utility without treatment, `u_S1`) to 0.95 (utility with treatment A, `u_trtA`) and costs $12,000 per year (`c_trtA`). This strategy does not impact the QoL of individuals in S2, nor does it change the risk of becoming sick or progressing through the sick states. Strategy B uses treatment B to reduce the rate of Sick individuals progressing to the S2 state with a hazard ratio of 0.6 (`hr_S1S2_trtB`) and costs $13,000 per year (`c_trtB`). Treatment B does not affect QoL. Strategy AB involves administering both treatments A and B, resulting in benefits in increased QoL in patients in the S1 state and reduced risk of progression from S1 to S2, accounting for the costs of both treatments. We discount both costs and QALYs at an annual rate of 3%. Model parameters and the corresponding R variable names are presented in Table 2 and follow the notation described in the DARTH coding framework.[12]

Note that for strategy A, the model has identical transition probabilities to SoC. The only difference is the added cost of the treatment for S1 and S2, and QoL increases for S1. After comparing the four strategies in terms of expected QALYs and costs, we calculate the incremental cost per QALY gained between non-dominated strategies.

Table 2

Description of parameters, their R variable name, base-case value and distribution.

Parameter	R name	Base-case	Distribution
Number of cycles (n_T)	`n_cycles`	75 years	constant
Names of health states	`v_names_states`	H, S1, S2, D	constant
Annual discount rate for costs	`d_c`	3%	constant
Annual discount rate for QALYs	`d_e`	3%	constant
Number of PSA samples (K)	`n_sim`	1,000	constant
Annual transition probabilities conditional on surviving			
- Disease onset (H to S1)	`p_HS1`	0.15	beta(30, 170)
- Recovery (S1 to H)	`p_S1H`	0.5	beta(60, 60)
- Time-independent disease progression (S1 to S2)	`p_S1S2`	0.105	beta(84, 716)
- Time-dependent disease progression (S1 to S2)	`v_p_S1S2_tunnels`		beta(84, 716)
Weibull parameters			
Scale (λ)	`p_S1S2_scale`	0.08	lognormal(log(0.08), 0.02)
Shape (γ)	`p_S1S2_shape`	1.10	lognormal(log(1.10), 0.05)
Annual mortality			
- Age-dependent background mortality rate (H to D)	`v_r_HDage`	age-specific	constant
- Hazard ratio of death in S1 vs H	`hr_S1`	3.0	lognormal(log(3.0), 0.01)
- Hazard ratio of death in S2 vs H	`hr_S2`	10.0	lognormal(log(10.0), 0.02)
Annual costs			
- Healthy individuals	`c_H`	$2,000	gamma(100.0, 20.0)
- Sick individuals in S1	`c_S1`	$4,000	gamma(177.8, 22.5)
- Sick individuals in S2	`c_S2`	$15,000	gamma(225.0, 66.7)
- Dead individuals	`c_D`	$0	-
- Cost of treatment A as an additional costs on individuals treated in S1 or S2	`c_trtA`	$12,000	gamma(576.0, 20.8)
- Cost of treatment B as an additional costs on individuals treated in S1 or S2	`c_trtB`	$13,000	gamma(676.0, 19.2)
Utility weights			
- Healthy individuals	`u_H`	1.00	beta(200, 3)
- Sick individuals in S1	`u_S1`	0.75	beta(130, 45)
- Sick individuals in S2	`u_S2`	0.50	beta(230, 230)
- Dead individuals	`u_D`	0.00	constant
Treatment A effectiveness			
- Utility for treated individuals in S1	`u_trtA`	0.95	beta(300, 15)
Treatment B effectiveness			
- Reduction in rate of disease progression (S1 to S2) as hazard ratio (HR)	`hr_S1S2_trtB`	log(0.6)	lognormal(log(0.6), 0.1)
Transition rewards			
- Utility decrement of healthy individuals when transitioning to S1	`du_HS1`	0.01	beta(11, 1088)
- Cost of healthy individuals when transitioning to S1	`ic_HS1`	$1,000	gamma(25, 40)
- Cost of dying when transitioning to D	`ic_D`	$2,000	gamma(100, 20)

The R code below describes the initialization of the input parameters.

```
## General setup
cycle_length <- 1   # cycle length equal one year
n_age_init   <- 25  # age at baseline
n_age_max    <- 100 # maximum age of follow up
n_cycles     <- n_age_max - n_age_init # number of cycles
# The 4 health states of the model:
v_names_states <- c("H",  # Healthy (H)
                    "S1", # Sick (S1)
                    "S2", # Sicker (S2)
                    "D")  # Dead (D)
n_states     <- length(v_names_states) # number of health states
d_e          <- 0.03 # discount rate for QALYs of 3% per cycle
d_c          <- 0.03 # discount rate for costs of 3% per cycle
v_names_str  <- c("Standard of care", # store the strategy names
                  "Strategy A",
                  "Strategy B",
                  "Strategy AB")
```

```
## Transition probabilities (per cycle), hazard ratios and odds ratio (OR)
p_HS1   <- 0.15  # probability of becoming Sick when Healthy
p_S1H   <- 0.5   # probability of becoming Healthy when Sick
p_S1S2  <- 0.105 # probability of becoming Sicker when Sick
hr_S1   <- 3     # hazard ratio of death in Sick vs Healthy
hr_S2   <- 10    # hazard ratio of death in Sicker vs Healthy

# Effectiveness of treatment B
hr_S1S2_trtB <- 0.6 # hazard ratio of becoming Sicker when Sick under treatment B

## State rewards
## Costs
c_H    <- 2000  # cost of being Healthy for one cycle
c_S1   <- 4000  # cost of being Sick for one cycle
c_S2   <- 15000 # cost of being Sicker for one cycle
c_D    <- 0     # cost of being dead for one cycle
c_trtA <- 12000 # cost of receiving treatment A for one cycle
c_trtB <- 13000 # cost of receiving treatment B for one cycle

# Utilities
u_H    <- 1     # utility of being Healthy for one cycle
u_S1   <- 0.75  # utility of being Sick for one cycle
u_S2   <- 0.5   # utility of being Sicker for one cycle
u_D    <- 0     # utility of being dead for one cycle
u_trtA <- 0.95  # utility when receiving treatment A for one cycle


## Transition rewards
du_HS1 <- 0.01  # one-time utility decrement when transitioning from Healthy to
Sick
ic_HS1 <- 1000  # one-time cost when transitioning from Healthy to Sick
ic_D   <- 2000  # one-time cost when dying
```

Incorporating simulation-time dependency

To illustrate simulation-time dependency in the Sick-Sicker cSTM, we model all-cause mortality as a function of age. We obtain all-cause mortality from life tables in the form of age-specific mortality hazard rates, $\mu(a)$, where a refers to age. For this example, we create a vector `v_r_mort_by_age` to represent age-specific background mortality hazard rates for 0 to 100 year-olds obtained from US life tables.[13] To compute the transition probability from state H to state D, corresponding to the cohort's age at each cycle, we transform the rate $\mu(a)$ to a transition probability assuming a constant exponential hazard rate within each year of age

$$p_{[H,D,t]} = 1 - \exp\{-\mu(a_0 + t)\},$$

where a_0 = 25 is the starting age of the cohort. Instead of iterating through the mortality hazard rates, we obtain a vector of background mortality hazard rates for the ages of interest between 25 through 100 by subsetting $\mu(a)$ (R variable name `v_r_mort_by_age`) for these ages. We transform the resulting R variable, `v_r_HDage`, to a probability.

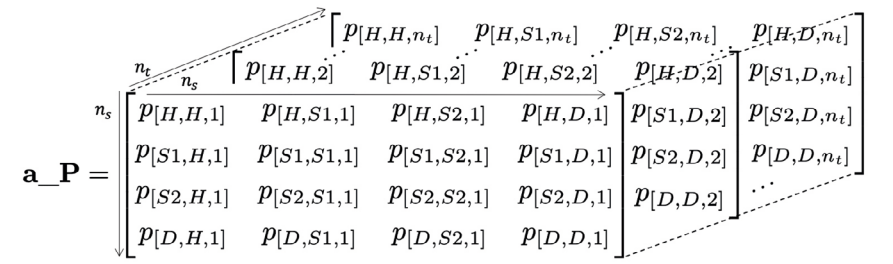
```
# Age-specific mortality rate in the Healthy state (background mortality)
v_r_HDage <- v_r_mort_by_age[(n_age_init + 1) + 0:(n_cycles - 1)]
# Transform to age-specific background mortality risk
v_p_HDage  <- 1 - exp(-v_r_HDage)
```

Because mortality in S1 and S2 are relative to background mortality which depends on age, mortality in S1 and S2 will also be age-dependent. To generate the age-specific mortality in S1 and S2, we multiply the age-specific background mortality rate, `v_r_HDage`, by the constant hazard ratios `hr_S1` and `hr_S2`, respectively. We then convert the resulting age-specific mortality rates to probabilities ensuring that the transition probabilities to D are bounded between 0 and 1.

```
## Age-specific mortality rates in the Sick and Sicker states
v_r_S1Dage <- v_r_HDage * hr_S1     # when Sick
v_r_S2Dage <- v_r_HDage * hr_S2     # when Sicker
## Age-specific probabilities of dying in the Sick and Sicker states
v_p_S1Dage <- 1 - exp(-v_r_S1Dage)  # when Sick
v_p_S2Dage <- 1 - exp(-v_r_S2Dage)  # when Sicker
```

To incorporate simulation-time dependency into the transition probability matrix, we expand the dimensions of the matrix and create a 3-dimensional transition probability array, $\mathbf{P}$ and named `a_P` in R, of dimensions $n_S \times n_S \times n_T$. The first two dimensions of this array correspond to transitions between states and the third dimension to time. The t-th element in the third dimension corresponds to the transition probability matrix at cycle t. A visual representation of `a_P` is shown in Figure 2.

Figure 2



A 3-dimensional representation of the transition probability array of the Sick-Sicker model with simulation-time dependency.

First, we initialize the transition probability array for SoC, `a_P_SoC`, with a default value of zero for all transition probabilities.

```
# Initialize the transition probability array
a_P_SoC <- array(0, dim = c(n_states, n_states, n_cycles),
                 dimnames = list(v_names_states, v_names_states, 0:(n_cycles - 1 )))
```

Filling `a_P_SoC` with the corresponding transition probabilities of the cohort under the SoC strategy is comparable with filling a transition probability matrix for a time-independent cSTM. However, this requires a slight modification of the code from the time-independent cSTM. Accounting for the time dimension is represented by the third dimension of the array. The code below illustrates how to assign age-dependent transition probabilities in the third dimension of the array. For constant transitions over time, we only need to provide one value for the transition probability. R replicates the value of such transitions as many times as the number of cycles (n_T+1 times in our example). We create the transition probability array for strategy A as a copy of SoC's because treatment A does not alter the cohort's transition probabilities.


```
### Fill in array
## From H
a_P_SoC["H", "H", ]   <- (1 - v_p_HDage) * (1 - p_HS1)
a_P_SoC["H", "S1", ]  <- (1 - v_p_HDage) * p_HS1
a_P_SoC["H", "D", ]   <- v_p_HDage
## From S1
a_P_SoC["S1", "H", ]  <- (1 - v_p_S1Dage) * p_S1H
a_P_SoC["S1", "S1", ] <- (1 - v_p_S1Dage) * (1 - (p_S1H + p_S1S2))
a_P_SoC["S1", "S2", ] <- (1 - v_p_S1Dage) * p_S1S2
a_P_SoC["S1", "D", ]  <- v_p_S1Dage
## From S2
a_P_SoC["S2", "S2", ] <- 1 - v_p_S2Dage
a_P_SoC["S2", "D", ]  <- v_p_S2Dage
## From D
a_P_SoC["D", "D", ]   <- 1

## Initialize transition probability matrix for strategy A as a copy of SoC's
a_P_strA <- a_P_SoC
```

As mentioned above, each slice along the third dimension of `a_P_SoC` corresponds to a transition probability matrix. For example, the transition matrix for 25-year-olds in the Sick-Sicker model under the SoC strategy can be retrieved by indexing the first slice of the array using:

```
a_P_SoC[, , 1]
##           H        S1        S2           D
## H  0.8491385 0.1498480 0.0000000 0.001013486
## S1 0.4984813 0.3938002 0.1046811 0.003037378
## S2 0.0000000 0.0000000 0.9899112 0.010088764
## D  0.0000000 0.0000000 0.0000000 1.000000000
```

For strategy B, we first initialize the three-dimensional array of transition probabilities, `a_P_strB` as a copy of `a_P_SoC` and update only the probability of remaining in S1 and the transition probability from S1 to S2 (i.e., `p_S1S2` is replaced with `p_S1S2_trtB`). Next, we create the transition probability array for strategy `AB, a_P_strAB`, as a copy of `a_P_strB` since the cSTMs for strategies B and AB have identical transition probabilities.

```
## Initialize transition probability array for strategy B
a_P_strB <- a_P_SoC
## Update only transition probabilities from S1 involving p_S1S2
a_P_strB["S1", "S1", ] <- (1 - v_p_S1Dage) * (1 - (p_S1H + p_S1S2_trtB))
a_P_strB["S1", "S2", ] <- (1 - v_p_S1Dage) * p_S1S2_trtB

## Initialize transition probability matrix for strategy AB as a copy of B's
a_P_strAB <- a_P_strB
```

Once we create the transition probability arrays, we check they are valid (i.e., ensuring transition probabilities are between 0 and 1, and transition probabilities from each state sum to 1) using the functions `check_sum_of_transition_array` and `check_transition_probability`, provided in the `darthtools` package (https://github.com/DARTH-git/darthtools).

```
### Check if transition probability matrices are valid
## Check that transition probabilities are [0, 1]
check_transition_probability(a_P_SoC)
check_transition_probability(a_P_strA)
check_transition_probability(a_P_strB)
check_transition_probability(a_P_strAB)
## Check that all rows sum to 1
check_sum_of_transition_array(a_P_SoC, n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(a_P_strA, n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(a_P_strB, n_states = n_states, n_cycles = n_
cycles)
check_sum_of_transition_array(a_P_strAB, n_states = n_states, n_cycles = n_
cycles)
```

In the Sick-Sicker model, the entire cohort starts in the Healthy state. Therefore, we create the $1 \times n_S$ initial state vector `v_s_init` with all of the cohort assigned to the H state:

```
v_s_init <- c(H = 1, S1 = 0, S2 = 0, D = 0) # initial state vector
v_s_init
#  H S1 S2  D
#  1  0  0  0
```


We use the variable `v_s_init` to initialize M represented by `m_M` for the cohort under SoC strategy. We also create a trace for each of the other treatment-based strategies. Note that the initial state vector, `v_s_init`, can be modified to account for the distribution of the cohort across the states at the start of the simulation and might vary by strategy. To simulate the cohort over the n_T cycles for the simulation-time-dependent cSTM, we initialize four cohort trace matrices, one for each strategy.

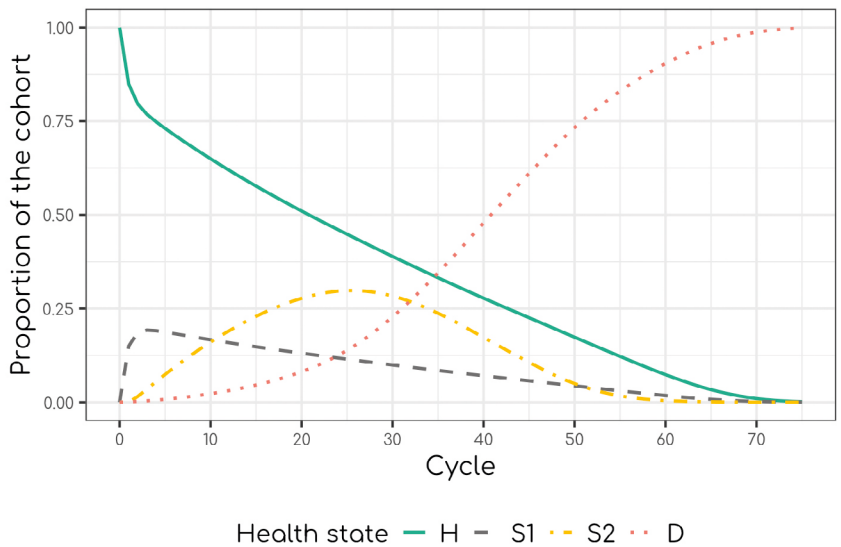
```
## Initialize cohort trace for age-dependent cSTM under SoC
m_M_SoC <- matrix(NA,
                  nrow = (n_cycles + 1), ncol = n_states,
                  dimnames = list(0:n_cycles, v_names_states))
# Store the initial state vector in the first row of the cohort trace
m_M_SoC[1, ] <- v_s_init
## Initialize cohort trace for strategies A, B, and AB
# Structure and initial states are the same as for SoC
m_M_strA  <- m_M_SoC # Strategy A
m_M_strB  <- m_M_SoC # Strategy B
m_M_strAB <- m_M_SoC # Strategy AB
```

We then use the matrix product to get the state vector of the cohort's distribution at each cycle t. This equation is similar to the one described for the time-independent model.[7] The only modification required is to index the transition probability arrays by t to obtain each strategy's cycle-specific transition probability matrices.

```
# Iterative solution of age-dependent cSTM
for(t in 1:n_cycles){
  # For SoC
  m_M_SoC[t + 1, ] <- m_M_SoC[t, ] %*% a_P_SoC[, , t]
  # For strategy A
  m_M_strA[t + 1, ] <- m_M_strA[t, ] %*% a_P_strA[, , t]
  # For strategy B
  m_M_strB[t + 1, ] <- m_M_strB[t, ] %*% a_P_strB[, , t]
  # For strategy AB
  m_M_strAB[t + 1, ] <- m_M_strAB[t, ] %*% a_P_strAB[, , t]
}
```

A graphical representation of the cohort trace for all cycles of the age-dependent cSTM under SoC is shown in Figure 3.

Figure 3



Cohort trace of the age-dependent cSTM under SoC.

Incorporating time-dependency on state-residence

Here we add the dependency on state-residence to the simulation-time dependent Sick-Sicker model defined above. We assume the risk of progression from S1 to S2 increases as a function of the time $\tau = 1, \cdots, n_{\text{tunnels}}$ the cohort remains in the S1 state. This increase follows a Weibull hazard function, $h(\tau)$, defined as

$$h(\tau) = \gamma\lambda(\lambda\tau)^{\gamma-1},$$

with a corresponding cumulative hazard, $H(\tau)$,

$$H(\tau) = (\lambda\tau)^{\gamma}, \tag{2}$$

where λ and γ are the scale and shape parameters of the Weibull hazard function, respectively.

To derive a transition probability from S1 to S2 as a function of the time the cohort spends in S1, $p_{[S1_\tau,S2,\tau]}$, we assume constant rates within each cycle interval (i.e., piecewise exponential transition times), where the cycle-specific probability of a transition is

$$p_{[S1_\tau,S2,\tau]} = 1 - \exp\left(-\mu_{[S1_\tau,S2,\tau]}\right), \tag{3}$$

where $\mu_{[S1_\tau,S2,\tau]}$ is the rate of transition from S1 to S2 in cycle τ defined as the difference in cumulative hazards between consecutive cycles[14]

$$\mu_{[S1_\tau,S2,\tau]} = H(\tau) - H(\tau - 1). \tag{4}$$

Substituting the Weibull cumulative hazard from Equation (2) into Equation (4) gives

$$\mu_{[S1_\tau,S2,\tau]} = (\lambda\tau)^{\gamma} - \left(\lambda(\tau - 1)\right)^{\gamma}, \tag{5}$$

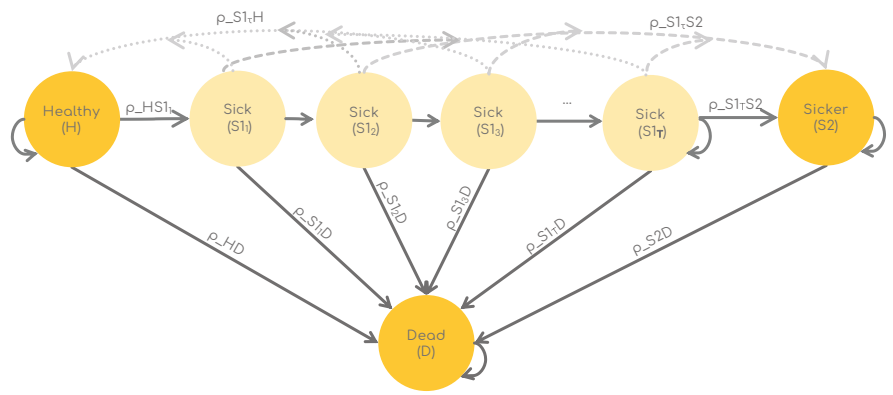
and the transition probability

$$p_{[S1_\tau,S2,\tau]} = 1 - \exp\left(-\left((\lambda\tau)^{\gamma} - \left(\lambda(\tau - 1)\right)^{\gamma}\right)\right). \tag{6}$$

We assume that state-residence dependency affects the cohort in the S1 state throughout the whole simulation (i.e., $n_{\text{tunnels}} = n_T$) and create a new variable called `n_tunnel_size` with the length of the tunnel equal to `n_cycles`. Thus, there will be 75 S1 tunnel states plus 3 more states (H, S2, D) resulting in a total of $n_{S_{\text{tunnels}}} = 78$.

Figure 4 shows the Sick-Sicker model's state-transition diagram with state-residence dependency with n_{tunnels} tunnel states for S1.

Figure 4



State-transition diagram of the Sick-Sicker model with tunnel states expanding the Sick state (S1_1, S1_2, ..., S1$_{n_{\text{tunnels}}}$).

To implement state-residence dependency in the Sick-Sicker cSTM, we create the vector variables `v_Sick_tunnel` and `v_names_states_tunnels` with the names of the Sick tunnel states' and all the states of the cSTM, including tunnels, respectively, and use the parameters listed in Table 1.

```
## Number of tunnels
n_tunnel_size <- n_cycles
## Vector with cycles for tunnels
v_cycles_tunnel <- 1:n_tunnel_size
## Vector with the names of the Sick tunnel state
v_Sick_tunnel <- paste("S1_", seq(1, n_tunnel_size), "Yr", sep = "")
## Create variables for model with tunnels
v_names_states_tunnels <- c("H", v_Sick_tunnel, "S2", "D") # state names
n_states_tunnels <- length(v_names_states_tunnels)         # number of states
## Initialize first cycle of Markov trace accounting for the tunnels
v_s_init_tunnels <- c(1, rep(0, n_tunnel_size), 0, 0)
```


Then, the transition rate and probability dependent on state-residence from Sick to Sicker, `v_r_S1S2_tunnels` and `v_p_S1S2_tunnels`, respectively, based on a Weibull hazard function are:

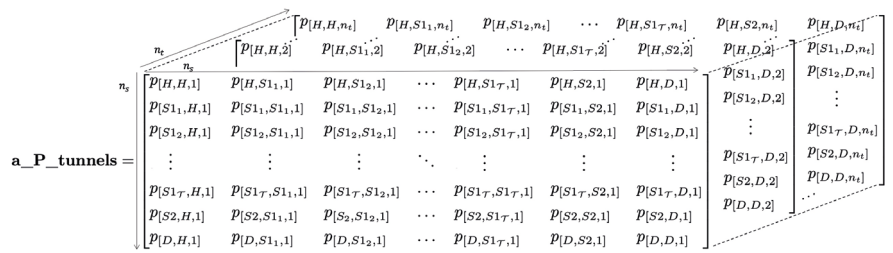
```
# Weibull parameters
p_S1S2_scale <- 0.08 # scale
p_S1S2_shape <- 1.10 # shape
# Weibull function
v_r_S1S2_tunnels <- (v_cycles_tunnel       * p_S1S2_scale) ^ p_S1S2_shape -
                    ((v_cycles_tunnel - 1) * p_S1S2_scale) ^ p_S1S2_shape

v_p_S1S2_tunnels <- 1 - exp(-v_r_S1S2_tunnels * cycle_length)
```

To adapt the 3-dimensional transition probability array to incorporate both age and state-residence dependency in the Sick-Sicker model under SoC, we first create an expanded 3-dimensional array accounting for tunnels, `a_P_tunnels_SoC`. The dimensions of this array are $n_{S_{\text{tunnels}}} \times n_{S_{\text{tunnels}}} \times n_T$. A visual representation of `a_P_tunnels_SoC` of the Sick-Sicker model with tunnel states expanding the Sick state is shown in Figure 5.

```
# Initialize array
a_P_tunnels_SoC <- array(0,
                         dim = c(n_states_tunnels, n_states_tunnels, n_cycles),
                         dimnames = list(v_names_states_tunnels,
                                         v_names_states_tunnels,
                                         0:(n_cycles - 1)))
```

Filling `a_P_tunnels_SoC` with the corresponding transition probabilities is similar to how it's done with `a_P_SoC` above, with the difference being that we now fill the transition probabilities from all the tunnel states by iterating through all the tunnel states and assigning the corresponding disease progression transition probabilities for each tunnel state.

Figure 5



The 3-dimensional transition probability array of the Sick-Sicker model expanded to account for simulation-time and state-residence dependency using τ tunnel states for S1.

```
### Fill in array
## From H
a_P_tunnels_SoC["H", "H", ]                 <- (1 - v_p_HDage) * (1 - p_HS1)
a_P_tunnels_SoC["H", v_Sick_tunnel[1], ]    <- (1 - v_p_HDage) * p_HS1
a_P_tunnels_SoC["H", "D", ]                 <- v_p_HDage
## From S1
for(i in 1:(n_tunnel_size - 1)){
  a_P_tunnels_SoC[v_Sick_tunnel[i], "H", ]  <- (1 - v_p_S1Dage) * p_S1H
  a_P_tunnels_SoC[v_Sick_tunnel[i],
                  v_Sick_tunnel[i + 1], ]   <- (1 - v_p_S1Dage) *
                                               (1 - (p_S1H + v_p_S1S2_tunnels[i]))
  a_P_tunnels_SoC[v_Sick_tunnel[i], "S2", ] <- (1 - v_p_S1Dage) * v_p_S1S2_
tunnels[i]
  a_P_tunnels_SoC[v_Sick_tunnel[i], "D", ]  <- v_p_S1Dage
}
# Repeat code for the last cycle to force the cohort stay in the last tunnel
state of Sick
a_P_tunnels_SoC[v_Sick_tunnel[n_tunnel_size], "H", ]   <- (1 - v_p_S1Dage) * p_S1H
a_P_tunnels_SoC[v_Sick_tunnel[n_tunnel_size],
                v_Sick_tunnel[n_tunnel_size], ]       <- (1 - v_p_S1Dage) *
                                 (1 - (p_S1H + v_p_S1S2_tunnels[n_tunnel_size]))
a_P_tunnels_SoC[v_Sick_tunnel[n_tunnel_size], "S2", ] <- (1 - v_p_S1Dage) *
                                                  v_p_S1S2_tunnels[n_tunnel_size]
a_P_tunnels_SoC[v_Sick_tunnel[n_tunnel_size], "D", ]  <- v_p_S1Dage
```



```
### From S2
a_P_tunnels_SoC["S2", "S2", ] <- 1 - v_p_S2Dage
a_P_tunnels_SoC["S2", "D", ]  <- v_p_S2Dage
# From D
a_P_tunnels_SoC["D", "D", ] <- 1
```

Next, we create the transition probability array for Strategy B. To implement the effectiveness of treatment B, we multiply the vector of transition rates, `v_r_S1S2_tunnels`, by the hazard ratio of treatment B, `hr_S1S2_trtB`. Then, we transform to a vector of transition probabilities that account for the duration of S1 state-residence under treatment B, `v_p_S1S2_tunnels_trtB`, following Equation (3).

```
# Apply hazard ratio to rate to obtain transition rate of becoming Sicker when
# Sick for treatment B
v_r_S1S2_tunnels_trtB <- v_r_S1S2_tunnels * hr_S1S2_trtB
# transform rate to probability to become Sicker when Sick under treatment B
# conditional on surviving
v_p_S1S2_tunnels_trtB <- 1 - exp(-v_r_S1S2_tunnels_trtB * cycle_length)
```

Then, we initialize the three-dimensional transition probability array for treatment B, `a_P_tunnels_trtB`, based on `a_P_tunnels_SoC`. The only difference here is that we update the transition probabilities from S1 involving `v_p_S1S2_tunnels` to `v_p_S1S2_tunnels_trtB` instead.

```
## Initialize transition probability array for treatment B
a_P_tunnels_trtB <- a_P_tunnels_SoC
## Update only transition probabilities from S1 involving v_p_S1S2_tunnels
for(i in 1:(n_tunnel_size - 1)){
  a_P_tunnels_trtB[v_Sick_tunnel[i], "H", ] <- (1 - v_p_S1Dage) * p_S1H
  a_P_tunnels_trtB[v_Sick_tunnel[i],
                   v_Sick_tunnel[i + 1], ] <- (1 - v_p_S1Dage) *
                                              (1 - (p_S1H + v_p_S1S2_tunnels_
trtB[i]))
  a_P_tunnels_trtB[v_Sick_tunnel[i], "S2", ] <- (1 - v_p_S1Dage) * v_p_S1S2_
tunnels_trtB[i]
  a_P_tunnels_trtB[v_Sick_tunnel[i], "D", ]  <- v_p_S1Dage
}
```

```
# repeat code for the last cycle to force the cohort stay in the last tunnel
state of Sick
a_P_tunnels_trtB[v_Sick_tunnel[n_tunnel_size], "H", ] <- (1 - v_p_S1Dage) *
p_S1H
a_P_tunnels_trtB[v_Sick_tunnel[n_tunnel_size],
           v_Sick_tunnel[n_tunnel_size], ] <- (1 - v_p_S1Dage) *
                                             (1 - (p_S1H +v_p_S1S2_tunnels_
trtB[n_tunnel_size]))
a_P_tunnels_trtB[v_Sick_tunnel[n_tunnel_size], "S2", ] <- (1 - v_p_S1Dage) *
                                                          v_p_S1S2_tunnels_
trtB[n_tunnel_size]
a_P_tunnels_trtB[v_Sick_tunnel[n_tunnel_size], "D", ]  <- v_p_S1Dage
```

Once we create both three-dimensional transition probability arrays with tunnels, we check they are valid (i.e., between 0 and 1 and transition probabilities from each state sum to 1).

```
### Check if transition probability matrices are valid
## Check that transition probabilities are [0, 1]
check_transition_probability(a_P_tunnels_SoC)
check_transition_probability(a_P_tunnels_trtB)
## Check that all rows sum to 1
check_sum_of_transition_array(a_P_tunnels_SoC, n_states = n_states_tunnels,
                              n_cycles = n_cycles)
check_sum_of_transition_array(a_P_tunnels_trtB, n_states = n_states_tunnels,
                              n_cycles = n_cycles)
```

To simulate the cohort and store its state occupation over the n_T cycles for the cSTM accounting for state-residence dependency, we initialize two new cohort trace matrices for the SoC and treatment B, `m_M_tunnels_SoC` and `m_M_tunnels_trtB`, respectively. The dimensions of both matrices are $(n_T + 1) \times n_{S_{\text{tunnels}}}$.

```
# Initialize cohort for state-residence cSTM under SoC
m_M_tunnels_SoC <- matrix(0,
                     nrow = (n_cycles + 1), ncol = n_states_tunnels,
                     dimnames = list(0:n_cycles, v_names_states_tunnels))
# Store the initial state vector in the first row of the cohort trace
m_M_tunnels_SoC[1, ] <- v_s_init_tunnels
## Initialize cohort trace under treatment B
m_M_tunnels_trtB <- m_M_tunnels_SoC
```

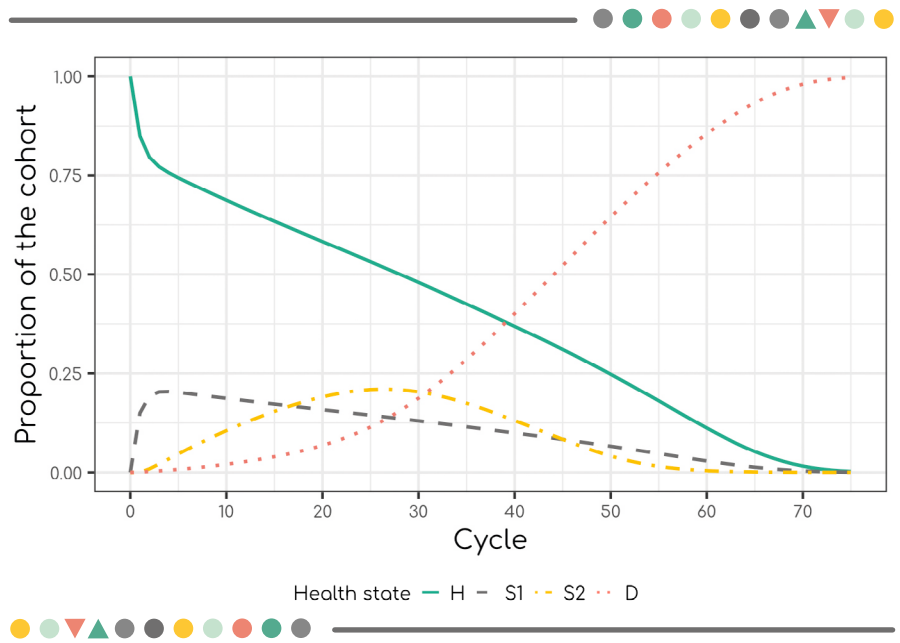
We then use the matrix product, similar to the simulation-time-dependent cSTM, to generate the full cohort trace.


```
# Iterative solution of state-residence dependent cSTM
for(t in 1:n_cycles){
  # For SoC
  m_M_tunnels_SoC[t + 1, ] <- m_M_tunnels_SoC[t, ] %*% a_P_tunnels_SoC[, , t]
  # Under treatment B
  m_M_tunnels_trtB[t + 1,] <- m_M_tunnels_trtB[t, ] %*% a_P_tunnels_trtB[, , t]
}
```

To compute a summarized cohort trace to capture occupancy in the H, S1, S2, D states under SoC, we aggregate over the tunnel states in each cycle (Figure 6).

```
# Create aggregated trace
m_M_tunnels_SoC_sum <- cbind (H = m_M_tunnels_SoC[, "H"],
                              S1 = rowSums(m_M_tunnels_SoC[,
                              which(v_names_states=="S1"): (n_tunnel_size +1)]),
                              S2 = m_M_tunnels_SoC[, "S2"],
                               D = m_M_tunnels_SoC[, "D"])
```

Figure 6



Cohort trace of the age-dependent cSTM accounting for state-residence dependency under SoC.

Epidemiological and economic measures

cSTMs can be used to generate different epidemiological and economic outputs. In a CEA, the main outcomes are typically the total discounted expected QALYs and total costs accrued by the cohort over the predefined time horizon. However, epidemiological outcomes can be helpful for calibration and validation. Some common epidemiological outcomes include survival, prevalence, incidence, the average number of events, and lifetime risk of events.[15] We show how to obtain some of these outcomes from the trace and transition probability objects.

Epidemiological measures

We provide the epidemiological definition of some of these outcomes and how they can be generated from a cSTM using the simulation-time dependent Sick-Sicker cSTM under SoC. In the GitHub repository, we provide the code to generate these outcomes from the state-residence-dependent cSTM.

Survival probability

The survival probability, $S(t)$, captures the proportion of the cohort remaining alive by cycle t. To estimate $S(t)$ from the simulated cohort of the simulation-time dependent Sick-Sicker model, shown in Figure 7, we sum the proportions of the non-death states for all n_T cycles in `m_M_SoC`.

```
v_S_SoC <- rowSums(m_M_SoC[, -which(v_names_states == "D")]) # vector with
survival curve
```

Life expectancy

Life expectancy (LE) refers to the expected number of time units remaining to be alive.[16] In continuous-time, LE is the area under the entire survival curve.[17]

$$LE = \int_{t=0}^{\infty} S(t)dt.$$

Figure 7 + 8

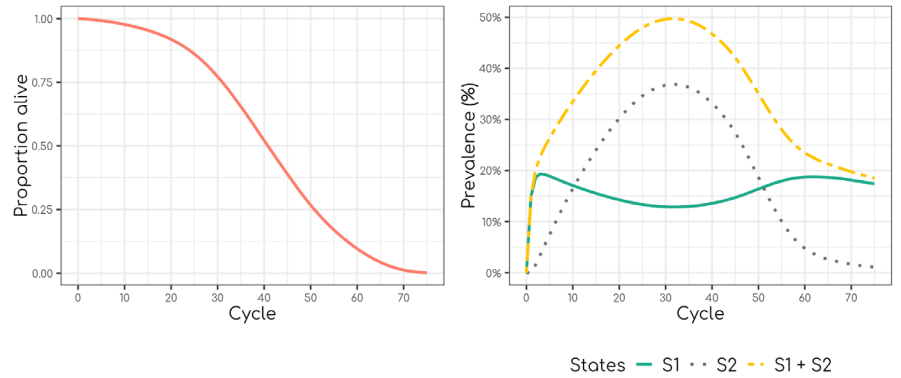


Figure 7: Survival curve of the age-dependent cSTM

Figure 8: Prevalence of sick states in age-dependent cSTM

In discrete-time using cSTMs, we often calculate restricted LE over a fixed time horizon (e.g., n_T) at which most of the cohort has transitioned to the Dead state and is defined as

$$LE = \sum_{t=0}^{n_T} S(t).$$

In the simulation-time-dependent Sick-Sicker model, where we simulated a cohort over n_T = 75 cycles, life expectancy `le_SoC` is 41.2 cycles, which is calculated as

```
le_SoC <- sum(v_S_SoC)     # life expectancy
```

Note that this equation expresses LE in the units of t. We use an annual cycle length; thus, the resulting LE will be in years. Analysts can also use other cycle lengths (e.g., monthly or daily), but the LE must be correctly converted to the desired unit if different than the cycle length units.

Prevalence

Prevalence is defined as the proportion of the population or cohort with a specific condition (or being in a particular health state) among those alive.[18] To calculate the prevalence of S1 at cycle t, prev $(t)_i$, we compute the ratio between the proportion of the cohort in S1 and the proportion alive at that cycle.[19] The proportion of the cohort alive is given by the survival probability $S(t)$ defined above. The individual prevalence of the S1 and S2 health states and the overall prevalence of sick individuals (i.e., S1 + S2) of the age-dependent Sick-Sicker cSTM at each cycle t is computed as follows and are shown in Figure 8.

```
v_prev_S1_SoC   <- m_M_SoC[, "S1"] / v_S_SoC  # vector with prevalence of Sick
v_prev_S2_SoC   <- m_M_SoC[, "S2"] / v_S_SoC  # vector with prevalence of Sicker
v_prev_S1S2_SoC <- rowSums(m_M_SoC[, c("S1", "S2")])/v_S_SoC  # prevalence of
                                                              Sick and Sicker
```

Economic measures

In CEA, we can calculate economic outcomes from state and transition rewards. A "state reward" refers to a value (e.g., cost, utility) assigned to individuals for remaining in a given health state for one cycle. A "transition reward" refers to the increase or decrease in either costs or utilities of transitioning from one health state to another, which may be associated with a one-time cost or utility impact. In the accompanying tutorial, we describe how to incorporate state rewards in CEA in detail.[7] Here, we describe and illustrate how to implement both state and transition rewards together using a transition array.

State rewards

As shown in the introductory tutorial, to add state rewards to the Sick-Sicker model, we first create a vector of utilities and costs for each of the four strategies considered. The vectors of utilities and costs, `v_u_SoC` and `v_c_SoC`, respectively, contain the utilities and costs corresponding to being in each of the four health states under SoC, shown in Table 2.


```
# Vector of state utilities under SoC
v_u_SoC <- c(H = u_H, S1 = u_S1, S2 = u_S2, D = u_D)
# Vector of state costs under SoC
v_c_SoC <- c(H = c_H, S1 = c_S1, S2 = c_S2, D = c_D)
```

We account for the benefits and costs of both treatments individually and their combination to create the state-reward vectors under treatments A and B (strategies A and B, respectively) and when applied jointly (strategy AB). Only treatment A affects QoL, so we create a vector of utilities for strategy A, `v_u_strA`, where we substitute the utility of being in S1 under SoC, `u_S1`, with the utility associated with the benefit of treatment A in being in that state, `u_trtA`. Treatment B does not affect QoL, so the vector of utilities for strategy B, `v_u_strB`, is the same as SoC's vector. However, when both treatments A and B are applied jointly (strategy AB), the resulting vector of utilities `v_u_strAB` equals that of strategy A.

```
# Vector of state utilities for strategy A
v_u_strA <- c(H = u_H, S1 = u_trtA, S2 = u_S2, D = u_D)
# Vector of state utilities for strategy B
v_u_strB <- v_u_SoC
# Vector of state utilities for strategy AB
v_u_strAB <- v_u_strA
```

Both treatments A and B incur a cost. To create the vector of state costs for strategy A, `v_c_strA`, we add the cost of treatment A, `c_trtA`, to S1 and S2 state costs. Similarly, when constructing the vector of state costs for strategy B, `v_c_strB`, we add the cost of treatment B, `c_trtB`, to S1 and S2 state costs. Finally, for the vector of state costs for strategy AB, `v_c_strAB`, we add both treatment costs to the state costs of S1 and S2.

```
# Vector of state costs for strategy A
v_c_strA <- c(H  = c_H,
              S1 = c_S1 + c_trtA,
              S2 = c_S2 + c_trtA,
              D  = c_D)
```

```
# Vector of state costs for strategy B
v_c_strB <- c(H  = c_H,
              S1 = c_S1 + c_trtB,
              S2 = c_S2 + c_trtB,
              D  = c_D)
# Vector of state costs for strategy AB
v_c_strAB <- c(H  = c_H,
               S1 = c_S1 + (c_trtA + c_trtB),
               S2 = c_S2 + (c_trtA + c_trtB),
               D  = c_D)
```

Transition rewards

As previously mentioned, dying (i.e., transitioning to the Dead state) incurs a one-time cost of $2,000 that reflects the acute care that might be received immediately preceding death. We also have a disutility and a cost increment on the transition from H to S1. Incorporating transition rewards requires keeping track of the proportion of the cohort that transitions between health states in each cycle while capturing the origin and destination states for each transition. The cohort trace, M, does not capture this information. However, obtaining this information is relatively straightforward in a cSTM and described in detail by Krijkamp et al.[9] Briefly, this approach involves changing the core computation in a traditional cSTM, from $m_t P_t$ to diag $(m_t) P_t$t. This simple change allows us to compute the proportion of the cohort that transitions between any two states in a cycle t. The result is no longer a cohort trace matrix, but rather a three-dimensional array that we refer to as a transition-dynamics array $(\mathbf{A})$ with dimensions $n_S \times n_S \times [n_T + 1]$. The t-th slice of $\mathbf{A}$, A_t, is a matrix that stores the proportion of the population that transitioned between any two states from cycles t - 1 to t. Similarly, we define the transition rewards by the states of origin and destination.

To account for both state and transition rewards, we create a *matrix* of rewards R_t of dimensions $n_S \times n_S$. The off-diagonal entries of R_t store the transition rewards, and the diagonal of R_t stores the state rewards for cycle t and assumes that rewards occur at the beginning of the cycle.[9] Finally, we multiply this matrix by A_t, the t-th slice of A, and apply discounting, within-cycle correction, and compute the overall reward for each strategy outcome. Below, we illustrate these computations in R.

To compute $\mathbf{A}$ for the simulation-time-dependent Sick-Sicker model under SoC, we initialize a three-dimensional array `a_A_SoC` of dimensions $n_S \times n_S \times [n_T + 1]$ and set the diagonal of the first slice to the initial state vector `v_s_init`. Next, we create a three-dimensional array for each of the strategies as a copy of the array under SoC.

```
# Initialize transition-dynamics array under SoC
a_A_SoC <- array(0,
                 dim = c(n_states, n_states, (n_cycles + 1)),
                 dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Set first slice to the initial state vector in its diagonal
diag(a_A_SoC[, , 1]) <- v_s_init
# Initialize transition-dynamics array for strategies A, B, and AB
# Structure and initial states are the same as for SoC
a_A_strA  <- a_A_SoC
a_A_strB  <- a_A_SoC
a_A_strAB <- a_A_SoC
```

We then compute a matrix multiplication between a diagonal matrix of each of the t-th rows of the cohort trace matrix under SoC and treatment B, denoted as `diag(m_M_SoC[t, ])` and `diag(m_M_strB[t, ])`, by the t-th matrix of the array of transition matrices, `a_P_SoC[, , t]` and `a_P_strB[, , t]`, respectively, over all n_T cycles.

```
# Iterative solution to produce the transition-dynamics array
for (t in 1:n_cycles){
  # For SoC
  a_A_SoC[, , t + 1]   <- diag(m_M_SoC[t, ]) %*% a_P_SoC[, , t]
  # For strategy A
  a_A_strA[, , t + 1]  <- diag(m_M_strA[t, ]) %*% a_P_strA[, , t]
  # For strategy B
  a_A_strB[, , t + 1]  <- diag(m_M_strB[t, ]) %*% a_P_strB[, , t]
  # For strategy AB
  a_A_strAB[, , t + 1] <- diag(m_M_strAB[t, ]) %*% a_P_strAB[, , t]
}
```

To create the arrays of rewards for costs and utilities for the simulation-time dependent Sick-Sicker cSTM, we create strategy-specific three-dimensional arrays of rewards and fill each of their rows across the third dimension with the vector of state rewards.

```
# Arrays of state and transition rewards
# Utilities under SoC
a_R_u_SoC <- array(matrix(v_u_SoC, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Costs under SoC
a_R_c_SoC <- array(matrix(v_c_SoC, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Utilities under Strategy A
a_R_u_strA <-  array(matrix(v_u_strA, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Costs under Strategy A
a_R_c_strA <- array(matrix(v_c_strA, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Utilities under Strategy B
a_R_u_strB <-  array(matrix(v_u_strB, nrow = n_states, ncol = n_states, byrow = T),
                    dim = c(n_states, n_states, n_cycles + 1),
                    dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Costs under Strategy B
a_R_c_strB <- array(matrix(v_c_strB, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Utilities under Strategy AB
a_R_u_strAB <-  array(matrix(v_u_strAB, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
# Costs under Strategy AB
a_R_c_strAB <- array(matrix(v_c_strAB, nrow = n_states, ncol = n_states, byrow = T),
                  dim = c(n_states, n_states, n_cycles + 1),
                  dimnames = list(v_names_states, v_names_states, 0:n_cycles))
```

To account for the transition rewards, we either add or subtract them in the corresponding location of the reward matrix representing the transitions of interest. Thus, for example, to account for the disutility of transitioning from H to S1 under strategy A, we subtract the disutility to the entry of the array of rewards corresponding to the transition from H to S1 across all cycles.


```
# Add disutility due to transition from Healthy to Sick
a_R_u_strA["H", "S1", ] <- a_R_u_strA["H", "S1", ] - du_HS1
```

In a similar approach, we add the costs of transitioning from H to S1 and the cost of dying under strategy A.

```
# Add transition cost due to transition from Healthy to Sick
a_R_c_strA["H", "S1", ] <- a_R_c_strA["H", "S1", ] + ic_HS1
# Add transition cost of dying from all non-dead states
a_R_c_strA[-n_states, "D", ] <- a_R_c_strA[-n_states, "D", ] + ic_D
a_R_c_strA[, , 1]
##       H    S1    S2    D
## H  2000 17000 27000 2000
## S1 2000 16000 27000 2000
## S2 2000 16000 27000 2000
## D  2000 16000 27000    0
```

Below, we show how to add the transition rewards to the reward matrices under SoC and strategies B and AB.

```
## SoC
# Add disutility due to transition from H to S1
a_R_u_SoC["H", "S1", ] <- a_R_u_SoC["H", "S1", ] - du_HS1
# Add transition cost due to transition from H to S1
a_R_c_SoC["H", "S1", ] <- a_R_c_SoC["H", "S1", ] + ic_HS1
# Add transition cost of dying from all non-dead states
a_R_c_SoC[-n_states, "D", ] <- a_R_c_SoC[-n_states, "D", ] + ic_D

## Strategy B
# Add disutility due to transition from Healthy to Sick
a_R_u_strB["H", "S1", ] <- a_R_u_strB["H", "S1", ] - du_HS1
# Add transition cost due to transition from Healthy to Sick
a_R_c_strB["H", "S1", ] <- a_R_c_strB["H", "S1", ] + ic_HS1
# Add transition cost of dying from all non-dead states
a_R_c_strB[-n_states, "D", ] <- a_R_c_strB[-n_states, "D", ] + ic_D

## Strategy AB
# Add disutility due to transition from Healthy to Sick
a_R_u_strAB["H", "S1", ] <- a_R_u_strAB["H", "S1", ] - du_HS1
# Add transition cost due to transition from Healthy to Sick
a_R_c_strAB["H", "S1", ] <- a_R_c_strAB["H", "S1", ] + ic_HS1
# Add transition cost of dying from all non-dead states
a_R_c_strAB[-n_states, "D", ] <- a_R_c_strAB[-n_states, "D", ] + ic_D
```

The state and transition rewards are applied to the model dynamics by element-wise multiplication between $\mathbf{A}$ and $\mathbf{R}$, indicated by the $\odot$ sign, which produces the array of outputs for all n_T cycles, $\mathbf{Y}$. Formally,

$$\mathbf{Y} = \mathbf{A} \odot \mathbf{R} \tag{7}$$

To obtain $\mathbf{Y}$ for QALYs and costs for all four strategies, we apply Equation (7) by the element-wise multiplication of the transition array `a_A_SoC` by the corresponding array of rewards.

```
# For SoC
a_Y_c_SoC        <- a_A_SoC * a_R_c_SoC
a_Y_u_SoC        <- a_A_SoC * a_R_u_SoC
# For Strategy A
a_Y_c_strA       <- a_A_strA * a_R_c_strA
a_Y_u_strA       <- a_A_strA * a_R_u_strA
# For Strategy B
a_Y_c_strB       <- a_A_strB * a_R_c_strB
a_Y_u_strB       <- a_A_strB * a_R_u_strB
# For Strategy AB
a_Y_c_strAB      <- a_A_strAB * a_R_c_strAB
a_Y_u_strAB      <- a_A_strAB * a_R_u_strAB
```

The total rewards for each health state at cycle t, $\mathbf{y}_t$, is obtained by summing the rewards across all j = 1, ⋯, n_S health states for all $n_T$cycles.

$$\mathbf{y}_t = \mathbf{1}^T Y_t = \left[\sum_{i=1}^{n_S} Y_{[i,1,t]}, \sum_{i=1}^{n_S} Y_{[i,2,t]}, \dots, \sum_{i=1}^{n_S} Y_{[i,n_S,t]}\right]. \tag{8}$$

To obtain the expected costs and QALYs per cycle for each strategy, $\mathbf{y}$, we apply Equation (8) again across all the matrices of the third dimension of $\mathbf{Y}$ for all the outcomes.


```
# Vectors of rewards
# QALYs under SoC
v_qaly_SoC       <- rowSums(t(colSums(a_Y_u_SoC)))
# Costs under SoC
v_cost_SoC       <- rowSums(t(colSums(a_Y_c_SoC)))
# QALYs under Strategy A
v_qaly_strA      <- rowSums(t(colSums(a_Y_u_strA)))
# Costs under Strategy A
v_cost_strA      <- rowSums(t(colSums(a_Y_c_strA)))
# QALYs under Strategy B
v_qaly_strB      <- rowSums(t(colSums(a_Y_u_strB)))
# Costs under Strategy B
v_cost_strB      <- rowSums(t(colSums(a_Y_c_strB)))
# QALYs under Strategy AB
v_qaly_strAB     <- rowSums(t(colSums(a_Y_u_strAB)))
# Costs under Strategy AB
v_cost_strAB     <- rowSums(t(colSums(a_Y_c_strAB)))
```

Within-cycle correction and discounting future rewards

Following the steps in the introductory cSTM tutorial,[7] here we use Simpson's 1/3rd rule for within-cycle correction (WCC),[20,21] and use exponential discounting for costs and QALYs. In our example, the WCC vector, $\mathbf{wcc}$, is the same for both costs and QALYs; thus, only one vector, `v_wcc`, is required.

```
## Vector with cycles
v_cycles <- seq(1, n_cycles + 1)
## Generate 2/3 and 4/3 multipliers for even and odd entries, respectively
v_wcc    <- ((v_cycles %% 2)== 0 )*(2/3) + ((v_cycles %% 2)!= 0 )*(4/3)
## Substitute 1/3 in first and last entries
v_wcc[1] <- v_wcc[n_cycles + 1] <- 1/3
```

The discount vectors, $\mathbf{d}$, for costs and QALYs for the Sick-Sicker model, `v_dwc` and `v_dwe`, respectively, are

```
# Discount weight for effects
v_dwe    <- 1 / ((1 + d_e) ^ (0:(n_cycles)))
# Discount weight for costs
v_dwc    <- 1 / ((1 + d_c) ^ (0:(n_cycles)))
```

To account for both discounting and WCC, we incorporate $\mathbf{wcc}$ in equation (9) using an element-wise multiplication with $\mathbf{d}$, indicated by the $\odot$ symbol, such that

$$y = \mathbf{y}'(\mathbf{d} \odot \mathbf{wcc}). \tag{9}$$

The total expected discounted costs and QALYs under all four strategies accounting for WCC, y, is obtained by applying Equation (9) to the expected outcomes accounting for transition rewards.

```
### For SoC
## QALYs
n_tot_qaly_SoC    <- t(v_qaly_SoC) %*% (v_dwe * v_wcc)
## Costs
n_tot_cost_SoC    <- t(v_cost_SoC) %*% (v_dwc * v_wcc)
### For Strategy A
## QALYs
n_tot_qaly_strA   <- t(v_qaly_strA) %*% (v_dwe * v_wcc)
## Costs
n_tot_cost_strA   <- t(v_cost_strA) %*% (v_dwc * v_wcc)
### For Strategy B
## QALYs
n_tot_qaly_strB   <- t(v_qaly_strB) %*% (v_dwe * v_wcc)
## Costs
n_tot_cost_strB   <- t(v_cost_strB) %*% (v_dwc * v_wcc)
### For Strategy AB
## QALYs
n_tot_qaly_strAB  <- t(v_qaly_strAB) %*% (v_dwe * v_wcc)
## Costs
n_tot_cost_strAB  <- t(v_cost_strAB) %*% (v_dwc * v_wcc)
```

The total expected discounted QALYs and costs for the simulation-time-dependent Sick-Sicker model under the four strategies accounting for WCC are shown in Table 3.

Table 3

Total expected discounted QALYs and costs per average individual in the cohort of the simulation-time-dependent Sick-Sicker model by strategy accounting for within-cycle correction.

	Costs	QALYs
Standard of care	$114,560	19.142
Strategy A	$211,911	19.840
Strategy B	$194,481	20.480
Strategy AB	$282,370	21.302

Resource prioritization in healthcare

To conduct the cost-effectiveness analysis, we follow the coding approach described in the introductory cSTM tutorial.[7] We combine the total expected discounted costs and QALYs for all four strategies into outcome-specific vectors, `v_cost_str` for costs and `v_qaly_str` for QALYs. We use the R package `dampack` (https://cran.r-project.org/web/packages/dampack/)[22] to calculate the incremental costs and effectiveness and the incremental cost-effectiveness ratio (ICER) between non-dominated strategies. `dampack` organizes and formats the results as a data frame, `df_cea`, that can be printed as a formatted table.

```
### Vector of costs
v_cost_str <- c(n_tot_cost_SoC, n_tot_cost_strA,
                n_tot_cost_strB, n_tot_cost_strAB)
### Vector of effectiveness
v_qaly_str <- c(n_tot_qaly_SoC, n_tot_qaly_strA,
                n_tot_qaly_strB, n_tot_qaly_strAB)

### Calculate incremental cost-effectiveness ratios (ICERs)
df_cea     <- dampack::calculate_icers(cost = v_cost_str,
                                       effect = v_qaly_str,
                                       strategies = v_names_str)
```

In terms of their costs and effectiveness, SoC is the least costly and effective strategy, followed by Strategy B producing an expected benefit of 1.338 QALYs per individual for an additional expected cost of $79,920 with an ICER of $59,726/QALY followed by Strategy AB with an ICER $106,927/QALY. Strategy A is a dominated strategy. The results of the CEA of the simulation-time-dependent Sick-Sicker model are presented in Table 4. The non-dominated strategies, SoC, B, and AB, form the cost-effectiveness efficient frontier of the CEA based on the simulation-time-dependent Sick-Sicker model (Figure 9).

Table 4

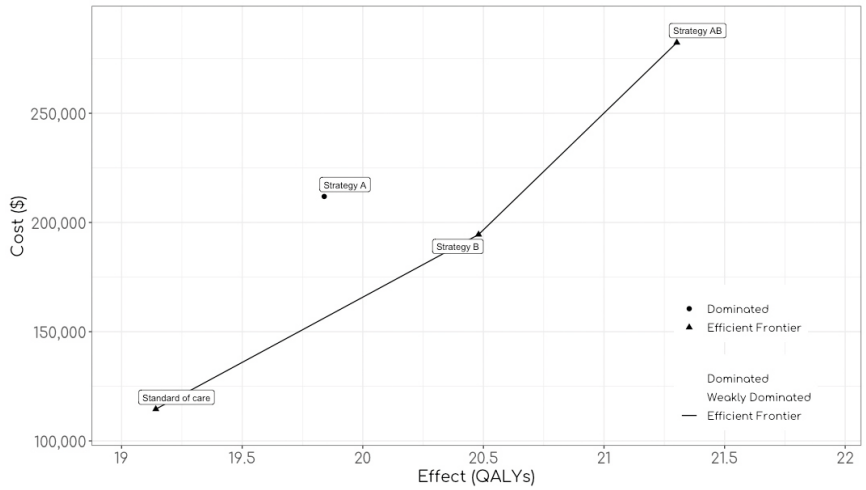
Cost-effectiveness analysis results for the simulation-time dependent Sick-Sicker model. ND: Non-dominated strategy; D: Dominated strategy.

Strategy	Costs ($)	QALYs	Incremental Costs ($)	Incremental QALYs	ICER ($/QALY)	Status
Standard of care	114,560	19.142	NA	NA	NA	ND
Strategy B	194,481	20.480	79,920	1.338	59,726	ND
Strategy AB	282,370	21.302	87,890	0.822	106,927	ND
Strategy A	211,911	19.840	NA	NA	NA	D

Probabilistic sensitivity analysis

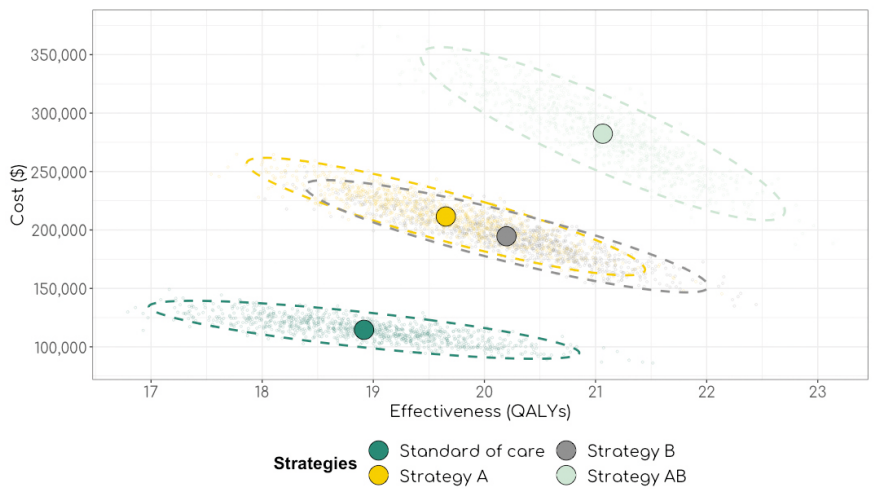
We conducted a probabilistic sensitivity analysis (PSA) to quantify the effect of model parameter uncertainty on cost-effectiveness outcomes.[23] In a PSA, we randomly draw parameter sets from distributions that reflect the current uncertainty in model parameter estimates. The parameters' distributions and their values are described in Table 2 and more detail in the Supplementary Material. We compute model outcomes for each sampled set of parameter values (e.g., total discounted cost and QALYs) for each strategy. We follow the steps to conduct a PSA from a previously published article describing the Decision Analysis in R for technologies in Health (DARTH) coding framework.[12]

Figure 9



Cost-effectiveness efficient frontier of all four strategies for the simulation-time-dependent Sick-Sicker model.

Figure 10



Cost-effectiveness scatter plot.

To conduct the PSA of the CEA using the simulation time-dependent Sick-Sicker cSTM, we sampled 1,000 parameter sets. For each set, we computed the total discounted costs and QALYs of each simulated strategy. Results from a PSA can be represented in various ways. For example, the joint distribution, 95% confidence ellipse, and the expected values of the total discounted costs and QALYs for each strategy can be plotted in a cost-effectiveness (CE) scatter plot (Figure 10),[24] where each of the 4,000 simulations (i.e., 1,000 combinations of total discounted expected costs and QALYs for each of the four strategies) are plotted as a point in the graph. The CE scatter plot for the CEA using the simulation-time-dependent model shows that strategy AB has the highest expected costs and QALYs. Standard of care has the lowest expected cost and QALYs. Strategy B is more effective and least costly than strategy A. And therefore, strategy A is strongly dominated by strategy B.

In Figure 11, we present the cost-effectiveness acceptability curves (CEACs) showing the probability that each strategy is cost-effective, and the cost-effectiveness frontier (CEAF), which shows the strategy with the highest expected net monetary benefit (NMB), over a range of willingness-to-pay (WTP) thresholds. Each strategy's NMB is computed using NMB=QALY×WTP−Cost[25] for each PSA sample. At WTP thresholds less than $65,000 per QALY, SoC is the strategy with the highest probability of being cost-effective and the highest expected NMB. Strategy B has the highest probability of being cost-effective and the highest expected NMB for WTP thresholds greater than $65,000 and smaller than $105,000 per QALY. Strategy AB, has the highest expected NMB for WTP thresholds greater than or equal to $105,000 and is the strategy with the highest probability of being cost-effective.

The CEAC and CEAF do not show the magnitude of the expected net benefit lost (i.e., expected loss) when the chosen strategy is not the cost-effective strategy in all the samples of the PSA. We quantify expected loss from each strategy over a range of WTP thresholds with

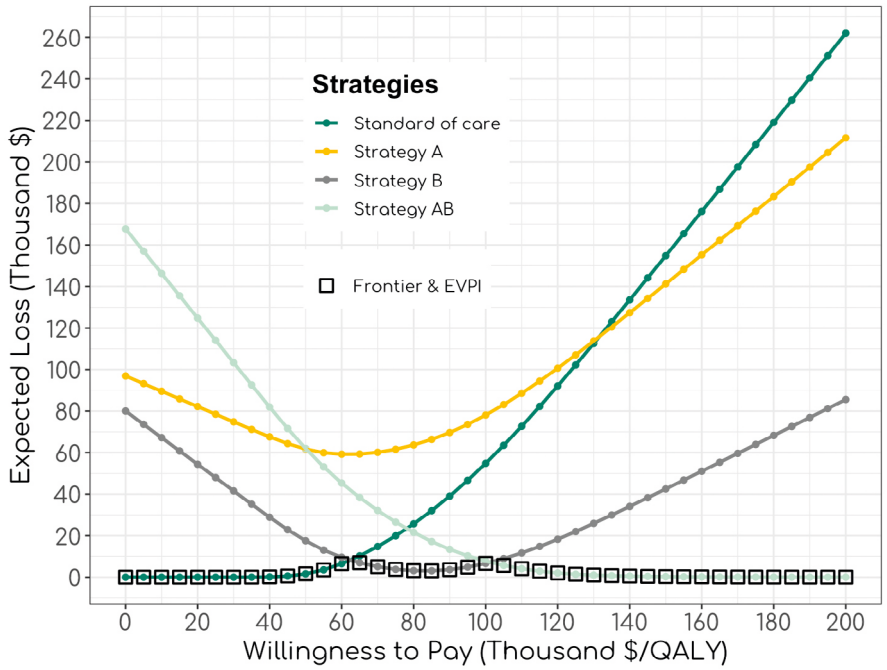
Figure 11

Cost-effectiveness acceptability curves (CEACs) and frontier (CEAF).

the expected loss curves (ELCs) to complement these results (Figure 12). The expected loss considers both the probability of making the wrong decision and the magnitude of the loss due to this decision, representing the foregone benefits of choosing a suboptimal strategy. The expected loss of the optimal strategy represents the lowest envelope of the ELCs because, given current information, the loss cannot be minimized further. The lower envelope also represents the expected value of perfect information (EVPI), which quantifies the value of eliminating parameter uncertainty. The strategy SoC has the lowest expected loss for WTP thresholds less than $60,000

per QALY, strategy B has the lowest expected loss for WTP threshold greater than or equal to $65,000 and less than $105,000. Strategy AB has the lowest expected loss for WTP threshold greater than or equal to $105,000 per QALY. At a WTP threshold of $65,000 per QALY, the EVPI is highest at $6,953. For a more detailed description of these outputs and the R code to generate them, we refer the reader to a previous publication by our group.[26]

Figure 12



Expected loss curves (ELCs) and expected value of perfect information (EVPI) generated from the probabilistic sensitivity analysis (PSA) output.

Discussion

In this tutorial, we conceptualize time-dependent cSTMs with their mathematical description and a walk-through of their implementation for a CEA in R using the Sick-Sicker example. We described two types of time-dependency: dependence on the time since the start of the simulation (simulation-time dependency) or on time spent in a health state (state-residence dependency). We also illustrate how to generate various epidemiological measures from the model, incorporate transition rewards in CEAs, and conduct a PSA.

We implemented simulation-time dependency by expanding the transition probability matrix into a transition probability array, where the third dimension captures time. However, there are alternative implementations of simulation-time dependency in cSTMs. For example, the model could be coded such that the time-varying elements of the transition probability matrix P_t are updated at each time point t as the simulation is run. This would eliminate the need for the transition probability array `a_P`, reducing computer memory requirements. But this comes at the expense of increasing the number of operations at every cycle, potentially slowing down the simulation.

We incorporated state-residence dependency using tunnel states by expanding the corresponding health states on the first and second dimensions of the 3-dimensional array to account for time spent in the current state in addition to simulation-time dependence. Another approach to account for state-residence dependency is to use a 3-dimensional transition probability array with dimensions for the current state, future state, and time in the current state.[27] However, in examples combining simulation-time and state-residence dependencies, this would necessitate a 4-dimensional array, which may be challenging to index.

It is the case that any time-varying feature in a discrete-time model can most generally be implemented as tunnel states with, at the

extreme, every state having a different tunnel state for each time step. The cohort would then progressively move through these tunnel states to capture their progression through time, and the model features (e.g., transition probabilities, costs or utilities) that change over time. Using time-varying transition probabilities is a shortcut that is possible when the cohort experiences these time-varying processes simultaneously as a function of the time from the simulation start. Even if the time-varying process has a different periodicity than the cycle length, either tunnel states or time-varying transition probabilities can be used to capture these effects. However, this time-varying process needs to be represented or approximated over an integer number of cycle lengths.

As described in the introductory tutorial, cSTMs are recommended when the number of states is considered "not too large".[15] Incorporating time-dependency in cSTMs using the provided approaches requires expanding the number of states by creating a multidimensional array for simulation-time dependency and/or creating tunnel states for state-residence dependency, increasing the amount of computer memory (RAM) required. It is possible to build reasonably complex time-dependent cSTMs in R as long as there is sufficient RAM to store the transition probability array and outputs of interest. For example, a typical PC with 8GB of RAM can handle a transition probability array of about 1,000 states and 600 time-cycle slices. However, if a high degree of granularity is desired, the dimensions of these data objects can grow quickly; if the required number of states gets too large and difficult to code, it may be preferable to use a stochastic (Monte Carlo) version of the state-transition model – often called individual-based state-transition models (iSTM) or microsimulation models – rather than a cohort simulation model.[15] In an iSTM, the risks and rewards of simulated individuals need not depend only on a current health state; they may also depend on an individual's characteristics and attributes. In addition, modelers can store health state history and other events over time for each individual to determine the risk of new events and corresponding costs and effects. Thus, we recommend carefully considering the required model structure before implementing it. An iSTM will also require additional functions to describe the dependency of transition probabilities and rewards on individuals' history. In a previous tutorial, we showed how to construct these functions for an iSTM using the Sick-Sicker example.[28]

We also described the concept and implementation of transition rewards. While these event-driven impacts could instead be captured by expanding the model state-space to include an initial disease state or a pre-death, end-of-life state, we can avoid state-space expansion by calculating incurred rewards based on the proportion of the cohort making the relevant transition. For implementation, this requires storing not just the cohort trace, which reflects how many individuals are in each state at a given cycle, but also a cohort state-transition array, which records how many individuals are making each possible transition at a given cycle.[9]

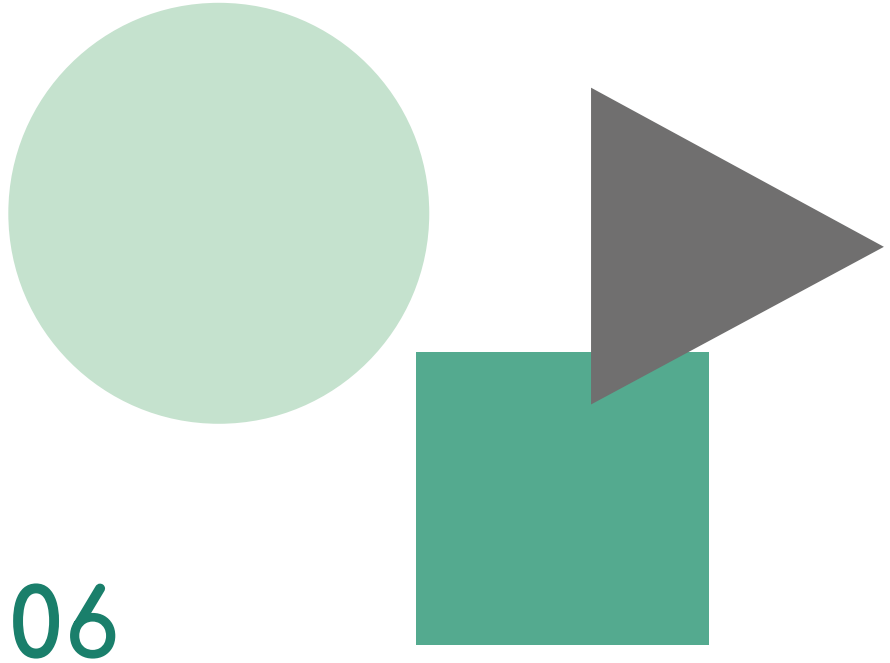
In summary, this tutorial extends our conceptualization of time-independent cSTMs to allow for time-dependency. It provides a step-by-step guide to implementing time-dependent cSTMs in R to generate epidemiological and economic measures, account for transition rewards, and conduct a CEA and a corresponding PSA. We hope that health decision scientists and health economists find this tutorial helpful in developing their cSTMs in a more flexible, efficient, and open-source manner. Ultimately, our goal is to facilitate R in health economic evaluations research with the overall aim to increase model transparency and reproducibility.

Supplemental material

Supplemental tables, data and model code are provided in the GitHub repository: https://github.com/DARTH-git/cohort-modeling-tutorial-timedep

References

1. Suijkerbuijk AWM, Van Hoek AJ, Koopsen J, De Man RA, Mangen MJJ, De Melker HE, Polder JJ, De Wit GA, Veldhuijzen IK. Cost-effectiveness of screening for chronic hepatitis B and C among migrant populations in a low endemic country. *PLoS One*. 2018; 13(11):1–16.
2. Sathianathen NJ, Konety BR, Alarid-Escudero F, Lawrentschuk N, Bolton DM, Kuntz KM. Cost-effectiveness Analysis of Active Surveillance Strategies for Men with Low-risk Prostate Cancer. *Eur Urol*. 2019; 75(6):910–917.
3. 3Lu S, Yu Y, Fu S, Ren H. Cost-effectiveness of ALK testing and first-line crizotinib therapy for non-small-cell lung cancer in China. *PLoS One*.; 13(10). Epub ahead of print October 1, 2018. DOI: 10.1371/journal.pone.0205827.
4. Djatche LM, Varga S, Lieberthal RD. Cost-Effectiveness of Aspirin Adherence for Secondary Prevention of Cardiovascular Events. *PharmacoEconomics - Open*. 2018; 2(4):371–380.
5. Pershing S, Enns EA, Matesic B, Owens DK, Goldhaber-Fiebert JD. Cost-Effectiveness of Treatment of Diabetic Macular Edema. *Ann Intern Med*. 2014; 160(1):18–29.
6. Smith-Spangler CM, Juusola JL, Enns EA, Owens DK, Garber AM. Population Strategies to Decrease Sodium Intake and the Burden of Cardiovascular Disease: A Cost-Effectiveness Analysis. *Ann Intern Med*. 2010; 152(8):481–487.
7. Alarid-Escudero F, Krijkamp E, Enns EA, Yang A, Hunink MGGM, Pechlivanoglou P, Jalal H. An Introductory Tutorial on Cohort State-Transition Models in R Using a Cost-Effectiveness Analysis Example. *arXiv:200107824v3*. 2021:1–26.
8. Snowsill T. A New Method for Model-Based Health Economic Evaluation Utilizing and Extending Moment-Generating Functions. *Med Decis Making*. 2019; 39(5):523–539.
9. Krijkamp EM, Alarid-Escudero F, Enns EA, Pechlivanoglou P, Hunink MGM, Yang A, Jalal HJ. A Multidimensional Array Representation of State-Transition Model Dynamics. *Med Decis Making*. 2020; 40(2):242–248.
10. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making*. 2017; 37(7):735–746.
11. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying Best-Fitting Inputs in Health-Economic Model Calibration: A Pareto Frontier Approach. *Med Decis Making*. 2015; 35(2):170–182.
12. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics*. 2019; 37(11):1329–1339.
13. Arias E, Heron M, Xu J. United States Life Tables, 2014. *Natl Vital Stat Reports*. 2017; 66(4):63.
14. Diaby V, Adunlin G, Montero AJ. Survival modeling for the estimation of transition probabilities in model-based economic evaluations in the absence of individual patient data: A tutorial. *Pharmacoeconomics*. 2014; 32(2):101–108.
15. 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.
16. Lee ET, Wang JW. *Statistical methods for Survival Data Analysis*. 3rd ed. Hoboken, NJ: Wiley, 2003.
17. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. Springer-Verlag, http://www.springer.com/statistics/life+sciences,+medicine+&+health/book/978-0-387-95399-1 (2003).
18. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Lippincott Williams & Wilkins, 2008.
19. Keiding N. Age-Specific Incidence and Prevalence: A Statistical Perspective. *J R Stat Soc*. 1991; 154(3):371–412.
20. Elbasha EH, Chhatwal J. Theoretical foundations and practical applications of within-cycle correction methods. *Med Decis Making*. 2016; 36(1):115–131.
21. Elbasha EH, Chhatwal J. Myths and misconceptions of within-cycle correction: a guide for modelers and decision makers. *Pharmacoeconomics*. 2016; 34(1):13–22.
22. Alarid-Escudero F, Knowlton G, Easterly CA, Enns EA. Decision Analytic Modeling Package (dampack), https://cran.r-project.org/web/packages/dampack/ (2021).
23. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, ISPOR-SMDM Modeling Good Research Practices Task Force. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. Med Decis Making. 2012; 32(5):722–732.
24. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Med Decis Making*. 2002; 22(4):290–308.
25. Stinnett AA, Mullahy J. Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. *Med Decis Making*. 1998; 18(2_suppl):S68–S80.
26. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. "Time Traveling Is Just Too Dangerous" But Some Methods Are Worth Revisiting: The Advantage of Expected Loss Curves Over Cost-Effectiveness Acceptability Curves and Frontier. *Value Health*. 2018; 22(5):611–618.
27. Hawkins N, Schulpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. *Med Decis Making*. 2005; 25(5):511–519.
28. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making*. 2018; 38(3):400–422.



06

A MULTIDIMENSIONAL ARRAY REPRESENTATION OF STATE-TRANSITION MODEL DYNAMICS

Eline Krijkamp*
Fernando Alarid-Escudero*
Eva A. Enns
Petros Pechlivanoglou
M.G. Myriam Hunink
Alan Yang
Hawre Jalal

*These authors contributed equally to this work.

Medical Decision Making 2020

Wayne Dyer

"If you change the way you look at things, the things you look at change"

A MULTIDIMENSIONAL ARRAY REPRESENTATION OF STATE-TRANSITION MODEL DYNAMICS

Abstract

Cost-effectiveness analyses often rely on cohort state-transition models (cSTMs). The cohort trace is the primary outcome of cSTMs, which captures the proportion of the cohort in each health state over time (state occupancy). However, the cohort trace is an aggregated measure that does not capture information about the specific transitions among health states (transition dynamics). In practice, these transition dynamics are crucial in many applications, such as incorporating transition rewards or computing various epidemiological outcomes that could be used for model calibration and validation (e.g., disease incidence and lifetime risk). In this article, we propose an alternative approach to compute and store cSTMs outcomes that capture both state occupancy and transition dynamics. This approach produces a multidimensional array from which both the state occupancy and the transition dynamics can be recovered. We highlight the advantages of the multidimensional array over the traditional cohort trace and provide potential applications of the proposed approach with an example coded in R to facilitate the implementation of our method.

Introduction

State-transition models (STM) are decision models commonly used in cost-effectiveness analysis (CEA) to estimate economic and health outcomes of different strategies over time in discrete time cycles.[1,2] In a cohort STM (cSTM), the disease dynamics are captured by distributing a closed cohort among a mutually exclusive and collectively exhaustive set of health states.[2–4] The cohort trace is the primary outcome of cSTMs, which comprises the proportion of the cohort in each health state over time (i.e., it summarizes state occupancy).[1,5] A limitation of the cohort trace is that it does not keep track of the transitions among health states over time (i.e., the transition dynamics of the cohort). As a consequence, it can only be used to capture outcomes that result from residing in a state for a full cycle by applying the so-called *state rewards* and does not contain a mechanism to assign *transition rewards*, which are applied only when specific transitions occur. It also limits the type of epidemiological outcomes that can be obtained from cSTMs. For example, obtaining incidence of a disease requires knowledge of the proportion of the population transitioning from a subset of states without disease to the state(s) representing the disease of interest.[6]

To overcome the limitations of the cohort trace, we propose a multidimensional array-based approach that serves as a full summary of cSTM dynamics that complements the already useful cohort trace. The proposed approach, called the dynamics-array approach, allows modelers to efficiently calculate all measures of interest that rely on transition dynamics and at the same time to aggregate all model dynamics into a standard cohort trace.

We start by providing a formal definition of cSTM components and the cohort trace. We complement this standard notation with a description of the detailed transition dynamics. Then, we introduce the multidimensional-array structure and show how it can be easily generated. In addition, we illustrate its use to compute a measure of interest that depends on transitions among health states. Finally, we demonstrate the dynamics-array approach with an illustrative example of a cSTM programmed in R[7,8] and compare this implementation with the traditional cohort trace approach in a simulation study. The R code is provided in the supplementary material and in GitHub (https://github.com/DARTH-git/state-transition-model-dynamics).

Traditional cohort trace approach

We denote the distribution of the cohort across n_s health states in a cSTM at the beginning of cycle t for all t in 0, ⋯, n_t as the state vector $\boldsymbol{m}_t$ of dimensions 1 × n_s. That is, each element in $\boldsymbol{m}_t$ represents the proportion of the cohort in health state i = 1, ⋯, n_s at time t. Thus, $\boldsymbol{m}_t$ is written as follows:

$$\mathbf{m}_t = \begin{bmatrix} m_{[t,1]} & m_{[t,2]} & \cdots & m_{[t,n_s]} \end{bmatrix}, \tag{1}$$

where the initial state vector $\boldsymbol{m}_0$ contains the distribution of the cohort across all n_s health states at the start of the simulation. The simulation begins at cycle 0 and all transitions are assumed to happen at the end of each cycle. The probability of transitioning from health state i to health state j at the end of cycle t is denoted as $p_{ij,t}$. This means that $p_{ij,t}$ determines how the population will be distributed in cycle t + 1. The collection of transition probabilities across the model states over the time horizon forms the time-dependent state transition probability matrix, P_t of dimensions n_s × n_s:

$$P_t = \begin{bmatrix} p_{[1,1,t]} & p_{[1,2,t]} & \cdots & p_{[1,n_s,t]} \\ p_{[2,1,t]} & p_{[2,2,t]} & \cdots & p_{[2,n_s,t]} \\ \vdots & \vdots & \ddots & \vdots \\ p_{[n_s,1,t]} & p_{[n_s,2,t]} & \cdots & p_{[n_s,n_s,t]} \end{bmatrix}. \tag{2}$$

For any t, all rows of P_t must sum to one. Note that if P_t is equal for all t times, equation (2) becomes a time-homogeneous transition probability matrix, where $P_t = P$.

The state vector at cycle $t+1$, $\boldsymbol{m}_{t+1}$, is then obtained by the inner product between the state vector at cycle t, $\boldsymbol{m}_t$, and the corresponding transition probability matrix P_t, such that

$$\mathbf{m_{t+1}} = \mathbf{m_t} P_t \quad \text{for } t = 0, \ldots, (n_t - 1). \tag{3}$$

Stacking the state vectors by rows for all t = 0, ⋯, n_t results in the full cohort trace matrix, M, of dimensions $(n_t+1) \times n_s$, where each row is a state vector ($\boldsymbol{m}_t$), resulting in

$$M = \begin{bmatrix} -\mathbf{m}_0- \\ -\mathbf{m}_1- \\ \vdots \\ -\mathbf{m}_{n_t}- \end{bmatrix}. \tag{4}$$

Together, the state vectors $\boldsymbol{m}_t$, the transition probability matrices P_t and the cohort trace M in equations (1) and (4), respectively, represent the three main components of a cSTM.

Dynamics-array approach

The trace matrix M aggregates transitions from all the states to a specific state, thus loses details of the transition dynamics. We propose to use a multidimensional array, $\boldsymbol{A}$, of dimensions $n_s \times n_s \times (n_t+1)$ to store the proportion of the cohort that transitions between any two health states in each cycle over the time horizon. This array can be thought of as a set of 2-dimensional matrices stacked along a third dimension that represents time. Below, we illustrate how to compute $\boldsymbol{A}$ from the three main components, the state vector $\boldsymbol{m}_t$, the transition probability matrices P_t, and the cohort trace M, of cSTMs.

$\boldsymbol{A}_0$ represents the first "slice" of $\boldsymbol{A}$. We compute $\boldsymbol{A}_0$ as a matrix containing the initial state vector $\boldsymbol{m}_0$ in its diagonal and 0s in the off-diagonal, such that

$$A_0 = \text{diag}(\mathbf{m}_0) = \begin{bmatrix} m_{[0,1]} & 0 & \cdots & 0 \\ 0 & m_{[0,2]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{[0,n_s]} \end{bmatrix}. \tag{5}$$

Each subsequent $(t+1)^{\text{th}}$ "slice" of $\boldsymbol{A}$ is obtained by multiplying a diagonal matrix of $\boldsymbol{m}_t$, denoted as diag($\boldsymbol{m}_t$), by P_t, such that

$$\mathbf{A}_{t+1} = \text{diag}(\mathbf{m}_t) \cdot P_t \quad \text{for } t = 0, \ldots, (n_t - 1). \tag{6}$$

The resulting elements of the t^{th} slice of $\boldsymbol{A}$, A_t for t>0, are

$$A_t = \begin{bmatrix} a_{[1,1,t]} & a_{[1,2,t]} & \cdots & a_{[1,n_s,t]} \\ a_{[2,1,t]} & a_{[2,2,t]} & \cdots & a_{[2,n_s,t]} \\ \vdots & \vdots & \ddots & \vdots \\ a_{[n_s,1,t]} & a_{[n_s,2,t]} & \cdots & a_{[n_s,n_s,t]} \end{bmatrix}, \tag{7}$$

where $a_{i,j,t}$ is the proportion of the cohort that transitions from state i to state j between cycles t - 1 and t, generated via

$$a_{[i,j,t]} = m_{[t-1,i]} p_{[i,j,t-1]} \quad \text{for } t = 1, \ldots, (n_t - 1), \tag{8}$$

where $\boldsymbol{m}_{t-1,j}$ is the proportion of the cohort in state i at cycle t - 1 and $p_{ij,t-1}$ the corresponding transition probability of transitioning from state i to state j at the end of cycle t - 1. In other words, $\boldsymbol{A}$ stores the transition dynamics of a simulated cohort in a cSTM.

Figure 1 presents graphically the computation involved in both (a) the traditional cohort trace approach and (b) the dynamics-array approach and shows the structures of (c) the resulting cohort trace M and dynamics-array $\boldsymbol{A}$.

Figure 1

(a) Cohort trace approach

$$\begin{bmatrix} m_{[t+1,1]} & m_{[t+1,2]} & \cdots & m_{[t+1,n_s]} \end{bmatrix} = \begin{bmatrix} m_{[t,1]} & m_{[t,2]} & \cdots & m_{[t,n_s]} \end{bmatrix} \begin{bmatrix} p_{[1,1,t]} & p_{[1,2,t]} & \cdots & p_{[1,n_s,t]} \\ p_{[2,1,t]} & p_{[2,2,t]} & \cdots & p_{[2,n_s,t]} \\ \vdots & \vdots & \ddots & \vdots \\ p_{[n_s,1,t]} & p_{[n_s,2,t]} & \cdots & p_{[n_s,n_s,t]} \end{bmatrix}$$

$$\boldsymbol{m}_{t+1} = \boldsymbol{m}_t \; P_t$$

(b) Dynamics-array approach

$$\begin{bmatrix} a_{[1,1,t+1]} & a_{[1,2,t+1]} & \cdots & a_{[1,n_s,t+1]} \\ a_{[2,1,t+1]} & a_{[2,2,t+1]} & \cdots & a_{[2,n_s,t+1]} \\ \vdots & \vdots & \ddots & \vdots \\ a_{[n_s,1,t+1]} & a_{[n_s,2,t+1]} & \cdots & a_{[n_s,n_s,t+1]} \end{bmatrix} = \begin{bmatrix} m_{[t,1]} & 0 & \cdots & 0 \\ 0 & m_{[t,2]} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m_{[t,n_s]} \end{bmatrix} \begin{bmatrix} p_{[1,1,t]} & p_{[1,2,t]} & \cdots & p_{[1,n_s,t]} \\ p_{[2,1,t]} & p_{[2,2,t]} & \cdots & p_{[2,n_s,t]} \\ \vdots & \vdots & \ddots & \vdots \\ p_{[n_s,1,t]} & p_{[n_s,2,t]} & \cdots & p_{[n_s,n_s,t]} \end{bmatrix}$$

$$A_{t+1} = diag(\boldsymbol{m}_t)\, P_t$$

(c) Final results

Matrix M

$$M = \begin{bmatrix} m_{[0,1]} & m_{[0,2]} & \cdots & m_{[0,n_s]} \\ m_{[1,1]} & m_{[1,2]} & \cdots & m_{[1,n_s]} \\ \vdots & \vdots & \ddots & \vdots \\ m_{[n_t,1]} & m_{[n_t,2]} & \cdots & m_{[n_t,n_s]} \end{bmatrix}$$

Array $\boldsymbol{A}$

$$\mathbf{A} = \begin{bmatrix} a_{[1,1,0]} & a_{[1,2,0]} & \cdots & a_{[1,n_s,0]} \\ a_{[2,1,0]} & a_{[2,2,0]} & \cdots & a_{[2,n_s,0]} \\ \vdots & \vdots & \ddots & \vdots \\ a_{[n_s,1,0]} & a_{[n_s,2,0]} & \cdots & a_{[n_s,n_s,0]} \end{bmatrix} \begin{bmatrix} a_{[1,1,1]} & a_{[1,2,1]} & \cdots & a_{[1,n_s,1]} \\ a_{[2,1,1]} & a_{[2,2,1]} & \cdots & a_{[2,n_s,1]} \\ \vdots & \vdots & \ddots & \vdots \\ a_{[n_s,1,1]} & a_{[n_s,2,1]} & \cdots & a_{[n_s,n_s,1]} \end{bmatrix} \cdots \begin{bmatrix} a_{[1,1,n_t]} & a_{[1,2,n_t]} & \cdots & a_{[1,n_s,n_t]} \\ a_{[2,1,n_t]} & a_{[2,2,n_t]} & \cdots & a_{[2,n_s,n_t]} \\ \vdots & \vdots & \ddots & \vdots \\ a_{[n_s,1,n_t]} & a_{[n_s,2,n_t]} & \cdots & a_{[n_s,n_s,n_t]} \end{bmatrix}$$

▶ *(a) The cohort trace approach computes each row vector $\boldsymbol{m}_{t+1}$ of the cohort trace M, where $\boldsymbol{m}_{t+1}$ describes the distribution of the simulated cohort among different health states at time t + 1. $\boldsymbol{m}_{t+1}$ results from multiplying the state vector $\boldsymbol{m}_t$ (green) by the transition probability matrix P_t. (b) The dynamics-array approach computes matrix $\boldsymbol{A}_{t+1}$ containing information regarding the transition dynamics of the simulated cohort at time t + 1. The state vector $\boldsymbol{m}_t$ is highlighted (green) to emphasize that the information in both approaches is identical. (c) The resulting matrix $\boldsymbol{M}$ and array $\boldsymbol{A}$ of the approaches (a) and (b), respectively.*

In R, it takes only a few lines of code to generate $\boldsymbol{A}$ complementary to M (Box 1).

Box 1

R code to iteratively generate the cohort trace M and the dynamics array $\boldsymbol{A}$

```
for(t in 1:n_t){     # loop through the number of cycles
# estimate the state vector for the next cycle (t + 1)
   m_M[t + 1, ]   <- m_M[t, ] %*% m_P
# estimate the transition dynamics at t + 1
   a_A[, , t + 1] <- diag(m_M[t, ]) %*% m_P
}
```

The cohort trace M can be computed from $\boldsymbol{A}$ by obtaining the t^{th} row of M, $\boldsymbol{m}_t$, summing each of the columns of $\boldsymbol{A}_t$ as follows:

$$\mathbf{m}_t = \mathbf{1}^T \mathbf{A}_t = \left[\sum_{i=1}^{n_s} a_{[i,1,t]}, \sum_{i=1}^{n_s} a_{[i,2,t]}, \ldots, \sum_{i=1}^{n_s} a_{[i,n_s,t]} \right], \quad (9)$$

where $\mathbf{1}$ is a vector of ones of dimension $n_s \times 1$. Although M can be obtained from $\boldsymbol{A}$, we prefer to compute both M and $\boldsymbol{A}$ simultaneously. Once generated, both M and $\boldsymbol{A}$ can be exported as data objects that contain all the information about the cSTM dynamics, which can then be used in future calculation of outcomes of interest.

Applying state and transition rewards

One of the main advantages of $\boldsymbol{A}$ over M is the ability to incorporate transition rewards. Here, we demonstrate how to apply both state and transition rewards (e.g., utilities or cost) to the cSTM by using the dynamics array, $\boldsymbol{A}$. Let R_t be a reward matrix of dimensions $n_s \times n_s$ that contains both state and transition rewards:

$$R_t = \begin{bmatrix} r_{[1,1,t]} & r_{[1,2,t]} & \cdots & r_{[1,n_s,t]} \\ r_{[2,1,t]} & r_{[2,2,t]} & \cdots & r_{[2,n_s,t]} \\ \vdots & \vdots & \ddots & \vdots \\ r_{[n_s,1,t]} & r_{[n_s,2,t]} & \cdots & r_{[n_s,n_s,t]} \end{bmatrix}, \quad (10)$$

where $r_{i,j,t}$ is the reward associated with transitioning from state i to state j at the end of cycle t. When $j = i$, $r_{i,j,t}$ is the reward associated with staying in the i^{th} health state at cycle t. That is, the off-diagonal entries of R_t store the transition rewards and the diagonal of R_t stores the state rewards for cycle t. Note that if R_t is equal for all t times (i.e., neither state nor transition rewards vary over time), equation (10) becomes a time-homogeneous rewards matrix, where $R_t = R$.

The state and transition rewards can be applied to the model dynamics by element-wise multiplication between A_t and R_t, indicated by the $\odot$ sign, which produces the matrix of outputs at cycle t, Y_t. Formally,

$$Y_t = A_t \odot R_t. \quad (11)$$

With this approach, the state rewards are accounted for at the beginning of the cycle and the transition rewards -assumed to happen at the end of the cycle- are accounted for at the next cycle.

In R, applying these rewards required one additional line of code compared to Box 1, as shown in line 5 of Box 2.

Box 2

R code to apply time-invariant state and transition rewards to the model dynamics stored in array $\boldsymbol{A}$

```
for(t in 1:n_t){                          # loop through the number of cycles
# estimate the state vector for the next cycle (t + 1)
   m_M[t + 1, ]   <- m_M[t, ] %*% m_P
# estimate the transition dynamics at t + 1
   a_A[, , t + 1] <- diag(m_M[t, ]) %*% m_P
# element-wise-multiplication of array A with the rewards matrices to
apply both state and transition rewards
   a_Y[, , t + 1] <- a_A[, , t + 1] * m_R
}
```

The total rewards for each health state at cycle t, $\boldsymbol{r}_t$, is obtained by summing the rewards across all $j = 1, \cdots, n_s$ health states:

$$\mathbf{r}_t = \mathbf{1}^T Y_t = \left[\sum_{i=1}^{n_s} Y_{[i,1,t]}, \sum_{i=1}^{n_s} Y_{[i,2,t]}, \ldots, \sum_{i=1}^{n_s} Y_{[i,n_s,t]} \right]. \quad (12)$$

Implementation in R using an illustrative example

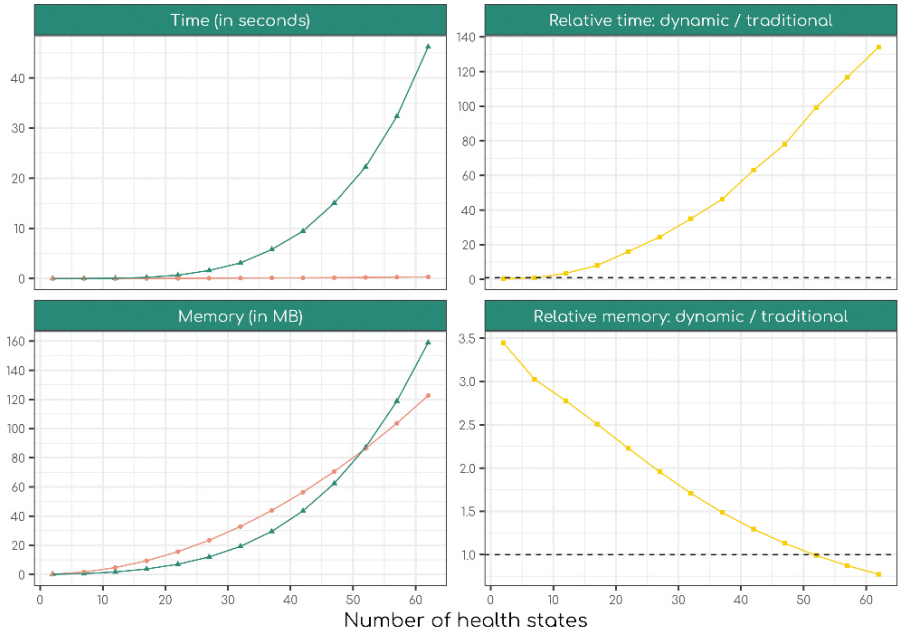
To facilitate the implementation of the dynamics-array approach, we demonstrate its use with a stylistic healthy-sick-dead 3-state time-homogeneous cSTM example coded in R.[8] The model is used to simulate a cohort of 70-year-old individuals to compute their expected costs and quality-adjusted life years (QALYs) accrued over

their remaining lifetime accounting for several transition rewards. Accounting for transition rewards with the traditional cohort trace approach is possible, however this requires creating additional temporary health states that keep track of the transitions. For our simple 3-state, model this already requires two additional temporary states. In more complex models, accounting for transition rewards will result in state explosion, and consequently, it is more likely to make errors while coding these models. The explanation of the state expansion and the R code for the traditional cohort trace approaches for our stylistic model can be found in the supplementary material and on GitHub (https://github.com/DARTH-git/state-transition-model-dynamics). GitHub also provides some code that shows that both approaches give identical results.

Comparison of methods using a simulation study

We conducted a simulation study to compare the computation time and memory requirements of the dynamics-array and traditional cohort trace approaches. We created a full factorial (13×110) design of experiment with the number of health states (from 2 to 62 with increments of 5 states) and the number of cycles (from 12 to 1320 with increments of 12 cycles) in a cSTM as the factors of the simulation study. We ran this full factorial experiment 10 times and took the average of the required time and memory to smooth out the variations in the computation time of R. Correcting for variation in R computation time is important, because the total required time is small (<50 seconds) and even a small variation (e.g., 3-5 seconds) could affect our results. Figure 2 shows that the dynamics-array approach is both faster (140-times with 62 health states, top part Figure 2) and requires less memory than the traditional cohort trace approach as the number of states increases (bottom part Figure 2). A more detailed description of the output of the simulation study is included in the supplementary material and the code is available on GitHub.

Figure 2



*Computation time and memory storage of the 2 approaches as a function of the number of states when running the model for 1,320 cycles. The **top left panel** shows the absolute computation time in seconds of both approaches. The **top right panel** shows the relative speedup of the dynamics-array approach compared to the traditional cohort trace approach. The horizontal dashed line at y-axis equals 0 indicates when the 2 approached are equally fast. The **bottom left panel** shows the absolute memory storage in megabytes (MB) of the 2 approached, while the **bottom right panel** shows the relative required memory of the dynamics-array approach compared to the traditional cohort trace approach. The horizontal dashed line at y-axis equals 1 indicate when both approached required the same memory storage. Above the line the traditional cohort trace requires less memory, while below the line the dynamics-array approach requires less memory. All results are based on the average of 10 simulations. This was done to smooth out the variations caused by the computation time of R.*

Estimation of epidemiological measures

By obtaining $\boldsymbol{A}$, it is possible to compute epidemiological outcomes that otherwise would not have been easily derived from M. For example, obtaining incidence and lifetime risk from M would require creating additional steps, variables or health states. Epidemiological outcomes could be used as outputs of simulation models for calibration or validation purposes. A full exposition of computing epidemiological measures from $\boldsymbol{A}$ is case-specific and is beyond the scope of this brief report. However, we illustrate the potential application of our approach with an example below.

Consider a cSTM with n_s > 3 health states. We are interested in calculating a ratio e_t of those that transition from health state 2 to health state n_s at cycle t to those that make the transition to health state n_s from health states 1, 2 and 3. Using the dynamics-array approach, the ratio e_t can be computed as follows:

$$e_t = \frac{a_{[2,n_s,t]}}{a_{[1,n_s,t]} + a_{[2,n_s,t]} + a_{[3,n_s,t]}} \quad \text{for } t = 1, \ldots, n_t. \quad (13)$$

With the transitional cohort trace approach, calculating this ratio would require adding three temporary health states to distinguish those that transition to n_s from the health states 1, 2 and 3.

Discussion

We propose a multidimensional-array approach to overcome a limitation of the cohort trace produced by cSTM in not being able to store transition dynamics. The practical application of our approach involves adding a simple step to the traditional cohort trace approach that stores all transitions among health states over time in multidimensional array $\boldsymbol{A}$. We described the multidimensional-array approach for a general cSTM where transitions are allowed from any state to others within a cycle, but our approach can also be applied to models that only allow one-state transition in a cycle (i.e., where the $p_{i,j,t}$ are 0 for those transitions that are not allowed).

Traditionally, researchers have dealt with this limitation of the cSTM cohort trace by creating temporary health states that collect the state-to-state transition information. However, as we showed in our "Comparison of methods using a simulation study", this solution can quickly complicate a model and result in an explosion of the number of health states. Using an individual-based microsimulation STM is another alternative[1], with considerable implications on computational time.[9]

Another method that explicitly keeps track of state-to-state transitions is through a discretely integrated condition-event (DICE) simulation.[10,11] DICE is a modeling technique that can free up some of the Markov restrictions that makes it possible to explicitly include many events occurring at various times. Although DICE simulation is a well-structured method, and the authors of the DICE papers provided very useful supplementary files to apply the method, we see the dynamics-array approach as a relatively simpler method to compute than DICE to overcome the limitation of the cohort trace on applying transition rewards and generating all the epidemiological outcomes of interest.

A potential limitation of the use of $\boldsymbol{A}$ is the additional computation needed when building the model. However, for many applications this may be a minor limitation given the matrix-based computational

efficiency of current computers. Another potential limitation is the additional storage memory required to store $\boldsymbol{A}$, which could become a limitation in systems with limited memory. This could be an issue for computationally complex models with multiple states. However, the benefit of using $\boldsymbol{A}$ for large models is that all the complexity in the model dynamics is summarized into a compact structure which makes it relatively simple to extract information or to apply new rewards without re-running the model.

In conclusion, structuring the output of cSTMs using the dynamics-array method approach is an efficient, simple and convenient method of summarizing the model dynamics. This simple structure allows applying state and transition rewards and obtaining epidemiological measures, while still being able to obtain and display the conventional cohort trace.

Supplemental material

Supplemental material for this article is available on the *Medical Decision Making* Website at https://journals.sagepub.com/doi/suppl/10.1177/0272989X19893973 and the GitHub repository mentioned in the manuscript.

References

1. 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.
2. Kuntz, KM, Russell, BL, Owens, DK, Sanders, GD, Trikalions, TA, Salomon J. Decision Models in Cost-Effectiveness Analysis. In: *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 2017, pp. 105–136.
3. Beck JR, Pauker SG. The Markov Process in Medical Prognosis. *Med Decis Mak*. 1983; 3(4):419–458.
4. Iskandar R. A theoretical foundation for state-transition cohort models in health decision analysis. *PLoS One*. 2018; 13(12):e0205543–e0205543.
5. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Mak*. 1993; 13(4):322–38.
6. Bouter LM, Zielhuis GA, Zeegers MPA. *Textbook of Epidemiology*. 1st ed. Houten: Bohn Stafleu van Loghum, 2018.
7. Jalal H, Pechlivanoglou P, Krijkamp EM, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Mak*. 2017; 37(7):735–746.
8. R Core Team. R: A language and Environment for Statistical Computing, https://www.r-project.org (2013).
9. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Mak*. 2018; 38(3):400–422.
10. Caro JJ. Discretely Integrated Condition Event (DICE) Simulation for Pharmacoeconomics. *Pharmacoeconomics*. 2016; 34(7):665–672.
11. Caro JJ, Möller J. Adding Events to a Markov Model Using DICE Simulation. *Med Decis Mak*. 2018; 38(2):235–245.

07

MICROSIMULATION MODELING IN HEALTH DECISION SCIENCES USING R: A TUTORIAL

Eline Krijkamp
Fernando Alarid-Escudero
Eva A. Enns
Hawre Jalal
M.G. Myriam Hunink
Petros Pechlivanoglou

Medical Decision Making 2018

"It does not matter how slowly you go as long as you do not stop."

Confucius

MICROSIMULATION MODELING IN HEALTH DECISION SCIENCES USING R: A TUTORIAL

Abstract

Clinical trials require participation of numerous patients, enormous research resources and substantial public funding. Time-consuming trials lead to delayed implementation of beneficial interventions and to reduced benefit to patients. This manuscript discusses two methods for thd treatment decisions. Estimating the net value of collecting further information, prior to undertaking a trial, informs a decision maker whether a clinical or health policy decision can be made with current information or if collection of extra evidence is justified. Additionally, estimating the net value of new information guides study design, data collection choices, and sample size estimation. The value-driven approach ensures the efficient use of research resources, reduces unnecessary burden to trial participants, and accelerates implementation of beneficial healthcare interventions.

Introduction

Healthcare policy makers are often faced with decisions on how to allocate healthcare resources given constrained budgets and continuous development of new, expensive technologies. Policy makers increasingly rely on health decision modeling tools to guide their decisions, as such models can synthesize evidence from different sources to give indications on the long-term implications and the uncertainty around a decision.[1]

One of the most common types of decision models used is that of state-transition cohort models.[2,3] Cohort models investigate a hypothetical homogeneous cohort of individuals as they transition across health states. In a deterministic cohort model the result is precisely determined given a set of initial conditions and parameters. As the complexity of decisions increases, deterministic cohort models become inadequate in reflecting the decision problem and more complex models are needed. For example, an assumption commonly made in cohort models is that the transition probabilities only depend on the current health state at any given cycle and cannot depend on the history prior to that cycle. This is often referred to as the "Markov" assumption and is an inherent limitation of cohort models.[4] Although this assumption can be relaxed by creating additional states that can capture the cohort's history, this can result in difficult to manage models due to "state explosion".[4]

Individual-based state-transition (or microsimulation) models address many of the limitations of deterministic cohort models because they can more accurately reflect individual clinical pathways, incorporate the impact of history on future events, and more easily capture the variation in patients' characteristics at baseline.[5,6] Microsimulation models differ from the more traditional cohort-based models since they simulate the impact of interventions or policies on individual trajectories rather than the deterministic mean response of homogeneous cohorts.[3,7–11] In a microsimulation model, outcomes are generated for each individual and are used to estimate the distribution of an outcome for a sample of potentially heterogeneous individuals. This individual-level simulation allows the inclusion of stochastic variation in disease progression as well as variation due to individual characteristics. Microsimulation models do not require the Markov assumption and so can introduce "memory" to the models' structure. This "memory" characteristic makes microsimulation models more popular in diseases where the severity of the disease or the cost and health outcomes vary with duration spent in a diseased state.[12]

There have been numerous microsimulation models developed for health applications. For example, the Population Health Model (POHEM) simulates the lifecycle of the Canadian population and assesses the impact of policy and program interventions, on the health status of Canadians;[13–15] the Future Elderly Model (FEM) predicts the impact of changes in healthcare status on future health and costs;[16] and the Prostate Cancer and Policy (PCOP) model evaluates the benefits and harms of prostate cancer screening.[17]

A drawback of the additional functionality of microsimulation models is numerical and computational complexity. In order to accommodate the computational demands of microsimulation models, health decision scientists increasingly adopt high level programming languages. The increasing availability of patient-level data also requires more advanced statistical analyses to be integrated with microsimulation models. In addition, computational efficiency can be achieved in decision modeling when relying on numerical and statistical computing software, such as Matlab or R, versus either spreadsheet software (e.g., Microsoft Excel) or common specialized software, such as TreeAge.[18] Recently, we illustrated the increased utilization of the statistical software R in health decision sciences, and provided a collection of resources for its application in medical decision making.[19]

R is an open-source and freely available software package, where statistical analyses can be combined with decision models within the same framework and the results can be presented in publication-ready tabular and graphical forms.[20]

The popularity of R in decision analysis follows similar trends in other fields of science, which have established R as the second most often used statistical software[21] and among the most often used programming languages.[22] Implementing a microsimulation model using a programming language provides the ability to use version control capabilities and repositories, such as GitHub. This is particularly useful when collaborating in teams where multiple members can contribute to different components of the model. Additionally, with R it is possible to use tools to generate web-based apps and graphical user interfaces (e.g., Shiny).[19]

R is not a programming language that was developed with a focus on computational performance. Nevertheless, there are some characteristics of the R programming language that can be leveraged to allow R-coded programs to perform efficiently. A key attribute of R is its ability to process problems more efficiently when presented in a vector rather than in a scalar format. Vectorization can reduce the need to rely on iterative "for" loops, which, in general, increase computation time significantly. Vectorization is particularly important in decision analysis, since processes are frequently repetitive and relatively straightforward to vectorize.

Despite the growing popularity of statistical programming for building simulation models in health decision sciences, few educational tutorials on how to perform these simulations are available.[12,23,24] This tutorial offers both a theoretical and applied introduction on building microsimulation models in R and has been created for beginner/intermediate users of R who are interested in developing an R-based microsimulation model.

This tutorial is structured as follows: First, we introduce a conceptual algorithm for model implementation, which can be applied to any high-level programming language, such as R. Subsequently, we illustrate how to build a microsimulation model in R using a simple, hypothetical decision problem. We guide the reader through the necessary steps and provide the R code used in the illustration.

We additionally explain a microsimulation solution that uses a vectorization approach. We then conclude with a discussion of the advantages and disadvantages of using R for microsimulation modeling and guidance on how to extend the presented methods to more complex applications. In Supplementary Appendices and on GitHub (https://github.com/DARTH-git/Microsimulation-tutorial), we provide the full R code used to run the microsimulation model and the vectorized approach.

Resource prioritization in healthcare

Microsimulation models are structured around a set of mutually exclusive and collectively exhaustive health states.[9,10,25,26] Each hypothetical individual can only be in one health state at any given cycle. Based on individual-specific transition probabilities, individuals are simulated through the model one at a time, while keeping track of their individual trajectories. In an economic evaluation framework, each health state, and potentially each transition between health states, is associated with a particular cost and health outcome value. For example, in the context of a cost-utility analysis, the cost and health outcome values would respectively represent the cost and the quality-adjusted life years (QALYs) associated with remaining in each health state for one cycle. Both state and transition values can depend on the individual's characteristics as well as on the past transitions of the individual.

State-transition microsimulation algorithm

The flexibility of a programming language like R implies that a microsimulation model can be implemented in multiple ways depending on the experience and programming style of the modeler and the research question that needs to be addressed. Below, we outline an algorithm that could be used when implementing a microsimulation model. This algorithm serves our goal to provide a generic approach with a reasonable balance between clarity and efficiency. First, we outline the steps to simulate one hypothetical individual over time.

1. The individual starts the simulation in an initial health state and is assigned a cost and health outcome value associated with staying in this initial state for one cycle, while taking into consideration any individual level characteristics.
2. During each cycle, the individual's probability to transition to a different health state (or remain in the current health state) in the next cycle is calculated based on the health states previously occupied and the individual's characteristics.
3. The health state to which the individual will transition to in the next cycle is sampled from a categorical distribution based on the probability of transitioning to each possible health state. The individual will then either transition to a new health state or remain in the same health state at the end of the cycle.
4. Each health state is associated with a particular cost or health outcome value attributed to remaining in the health state for one cycle. This could represent the costs associated with a health state and in a cost-utility context, the utility of remaining in a certain health state for one cycle. State-specific costs and health outcome values can depend on the individual's characteristics (age, gender etc.) as well as on past transitions of the individual. Transition values, that is one-time costs and one-time changes in health outcomes associated with the transition, may apply.
5. By aggregating all the state and transition values over the model's cycles (and applying discounting if needed) we can estimate the (discounted) total model outcomes for that individual's lifetime.

Repeating these steps for all individuals in the simulation allows us to describe the distribution of total cost and health outcomes for the population of interest.

To formalize the previous steps, we define the following notation

- nt as the number of discrete time cycles.
- t as the current cycle, ranging from 0 to nt.
- cl as the cycle length.
- ni as the number of individuals.
- i as a specific individual from the group of individuals.
- nx as the number of individual characteristics.
- ns as the number of health states.
- n as the ns character vector of health state names.
- p as the ns vector containing the transition probabilities for individual i at the end of cycle t.
- M as the $ni \times (nt + 1)$ matrix capturing the states occupied by all individuals during all cycles.
- X as the $ni \times (nt + 1) \times nx$ 3-dimensional array of individuals' characteristics for all cycles (e.g., age, gender, disease severity, co-morbidities).
- $Probs(M_i, X_i)$, $Costs(M_i, X_i)$, $Effs(M_i, X_i, cl)$ as functions assigning individual-specific probabilities, cost and health outcome values conditional on M_i, X_i and cl. M_i and X_i represent the vectors of health states and individual characteristics respectively for all time points for individual i.
- C as the $ni \times (nt + 1)$ matrix capturing the costs for all individuals during all cycles.
- E as the $ni \times (nt + 1)$ matrix capturing the health outcomes for all individuals during all cycles.
- dwc as the $nt \times 1$ vector of discount weights for costs, where $dwc = 1 / (1 + dc)^t$, with dc being the discount rate for the costs at any cycle t. The first cycle is not discounted, since the first time point t is equal to zero.
- dwe as the $nt \times 1$ vector of discount weights for health outcomes, where $dwe = 1 / (1 + de)^t$, with de being the discount rate for the health outcome values at any cycle t. The first cycle is not discounted, since the first time point t is equal to zero.
- tc as the $ni \times 1$ vector of total discounted costs for all individuals.
- te as the $ni \times 1$ vector of total discounted health outcomes for all individuals.

Below we present an algorithm that can be used to perform a state-transition microsimulation. The enumerated areas on the left correspond to the steps of the microsimulation algorithm outlined before.

1 for i = 1 to ni do

M_{i0} is assigned the initial health state for individual i.

$C_{i0} = Costs(M_{i0}, X_{i0})$ {Cost value for individual i during cycle 0 as function of the initial health state and individual's characteristics}

$E_{i0} = Effs(M_{i0}, X_{i0}, cl)$ {Health outcomes for individual i during cycle 0 as function of the initial health state, individual's characteristics and cycle length}

Then,

for t = 1 to nt do

2 $= Probs(M_{i[0:t]}, Xi_{[0:t]})$ {State-transition probabilities for individual i at cycle t as function of the complete history of states and individual characteristics up to the current cycle t}

3 $M_{it} \sim Cat(n, p)$ {Sample the state individual i will transition to during cycle t from a categorical distribution (Cat) of ns states n with probabilities p}

Assign state values for costs and health outcomes using the M matrix and individual characteristics X.

4 $C_{it} = Costs(M_{i[0:t]}, X_{i[0:t]})$ {Costs for individual i during cycle t as function of the complete history of states and individual characteristics up to the current cycle t}

$E_{it} = Effs(M_{i[0:t]}, X_{i[0:t]}, cl)$ {Health outcomes for individual i during cycle t as function of the complete history of states and individual characteristics up to the current cycle t and the cycle length}

end

end

5 $tc = C \bullet dwc$ {Total (discounted) costs per individual for all individuals, "•" denotes inner product multiplication}

$te = E \bullet dwe$ {Total (discounted) health outcomes per individual for all individuals, "•" denotes inner product multiplication}

The implementation of the microsimulation algorithm using R syntax is illustrated in Box 1. The notation in Box 1 follows the algorithm outlined before with inner product multiplication indicated in R by the "`%*%`" symbol. Since t starts at 0, we increment t by 1 to correspond to the element index in the vectors and matrices in R. Variable names for matrices are preceded by the prefix "m." while for vectors the prefix "v." is used. At each individual and each cycle, the vector of probabilities `v.p` is replaced with a new set of transition probabilities. The `Probs()` function returns a vector of state transition probabilities. Finally, both `Costs()` and `Effs()` functions yield a single value per subject per cycle that captures the cost and health outcomes per individual per cycle.

The algorithm in Box 1 illustrates the microsimulation process for one treatment strategy. In cases where two or more strategies are compared, the algorithm should be completed separately for each strategy with strategy-specific values. The vectors of undiscounted total costs and health outcomes per individual can be calculated by setting dc and de equal to zero. The approach above can be readily expanded or simplified. For example, in the context of a cost-utility economic evaluation, the QALYs accumulated within a cycle would be captured in the health outcome matrix E and consequently te would capture the total QALYs across all cycles. In extension, $\overline{tc}$ and $\overline{te}$, the mean of individual-specific costs and QALYs for each strategy, respectively, could be calculated to generate standard incremental cost-effectiveness tables.

Box 1

```r
# The matrices that will store information across the iterations need to be initialized first
m.M   <- m.C <- m.E <- matrix(nrow = n.i, ncol = n.t + 1) # create empty matrices
v.n   <- 1:n.s                                            # the vector of health state numbers
v.dwc <- 1 / (1 + dc) ^ (0:n.t)                           # cost discount weights based on the discount rate dc
v.dwe <- 1 / (1 + de) ^ (0:n.t)                           # health outcome discount weights based on the discount rate de
for (i in 1:n.i) {
  # Step 1
  m.M[i, 1] <- v.M_1[i]                                   # specify the initial health state per individual
  m.C[i, 1] <- Costs(m.M[i, 1], m.X[i, 1])                # costs per individual during cycle 0
  m.E[i, 1] <- Effs (m.M[i, 1], m.X[i, 1], cl)            # health outcome value per individual during cycle 0

  for (t in 1:n.t) {
    # Step 2
    v.p       <- Probs(m.M[i, 1:t], m.X[i, 1:t])        # calculate the transition probabilities
    # Step 3
    m.M[i, t + 1] <- sample(v.n, prob = v.p, size = 1) # sample health state at cycle t + 1 and store it in m.M
    # Step 4
    m.C[i, t + 1] <- Costs(m.M[i, 1:t + 1], m.X[i, 1:t + 1])  # costs per individual during cycle t + 1
    m.E[i, t + 1] <- Effs (m.M[i, 1:t + 1], m.X[i, 1:t + 1], cl) # health outcome per individual during cycle t + 1
  }                                                       # close the loop for time cycles
}                                                         # close the loop for the individuals

# Step 5
tc <- m.C %*% v.dwc                                       # total discounted cost per individual

te <- m.E %*% v.dwe                                       # total discounted health outcomes per individual
```

A generic implementation of the microsimulation algorithm outlined before using R syntax. Some of the R variable names are defined in Table 1.

Stochastic variation and random sampling in microsimulation models

Microsimulation models rely on Monte Carlo (MC) simulation methods to describe the stochastic transition process of individuals through the model.[27] Due to this stochastic process, the estimates of a microsimulation will be different between individuals. Therefore, executing the model for a sufficiently large number of individuals is important to achieve a representative distribution of population outcomes. The variability around the mean model estimate is called Monte Carlo Standard Error (MCSE). In a decision problem that could be described with either a microsimulation or a cohort model, the two model types should yield the same estimates as the number of simulated individuals in the microsimulation model increases.[28,29] In this tutorial we use graphical diagnostics to show how outcomes produced using a microsimulation model converge to those produced using a cohort model. One such example is a plot of the mean simulation estimates as a function of the number of individuals run in the microsimulation model (i.e., microsimulation sample size).

When designing a microsimulation model for comparative effectiveness or cost-effectiveness purposes, it is important that simulated individuals are as similar as possible across comparators, except for the intervention they are exposed to. One can think of the random sampling process as two parallel processes where the same hypothetical individual is exposed to both the intervention and the control. This can be achieved by using pre-sampled values from the distributions used in the model, by explicitly setting a seed number per individual.[30] This ensures that individuals across intervention scenarios share the same baseline characteristics and reduces any difference observed between the intervention scenarios that can be attributed to MC variability and not to intervention allocation. An additional benefit of setting the seed is that it allows the results to be reproducible despite the inherent stochastic process.

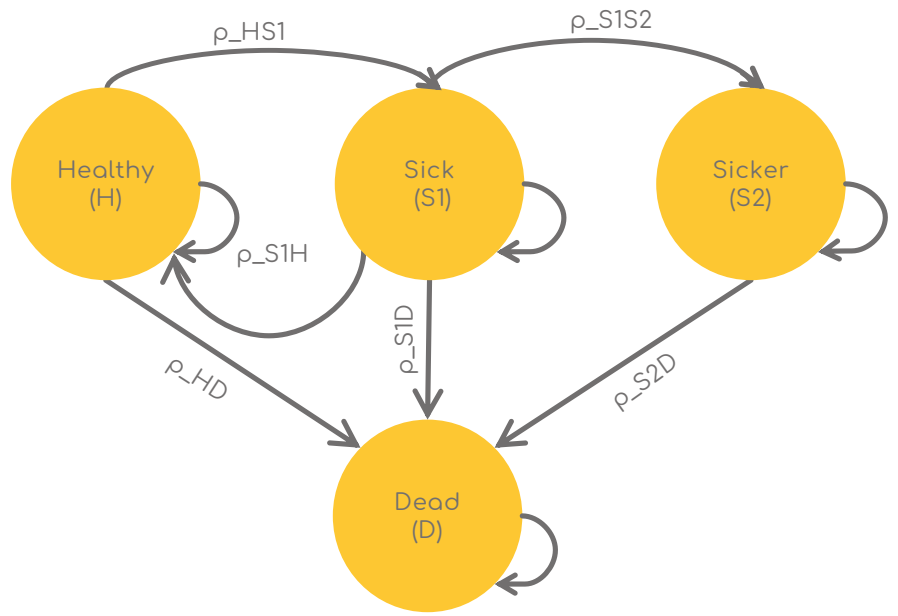
Example: the Sick-Sicker model

In this section, we illustrate the implementation of the algorithm presented above, using a previously published decision model.[31] This illustration provides an example of how R can be utilized for the development of a microsimulation model for a specific disease process. We intend for the R code below to be generic to allow it to be customizable. There are many opportunities to improve the efficiency of the code for specific applications, but the goal here is to provide a generic code as a starting point.

The illustration relies upon a state-transition cohort model of a hypothetical disease (the Sick-Sicker model), first described by Enns et al.[31] We initially create a microsimulation implementation of the Sick-Sicker cohort model and subsequently extend it to illustrate the advantages of microsimulation models in incorporating individual-level characteristics and "memory" in the disease process. The Sick-Sicker model consists of four health states: healthy (H), sick (S1), sicker (S2) and dead (D) (Figure 1). All individuals are assumed to be healthy at baseline. Over time, healthy individuals face a risk to develop the disease and progress to S1. Individuals in S1 can recover (return to state H), stay in S1, progress further to S2 or die. Individuals in S2 cannot recover (i.e. cannot transition to either S1 or H). Individuals in H are assumed to have a fixed mortality rate. Individuals in S1 and S2 have an increased mortality rate compared to healthy individuals. All individuals are followed for 30 years using annual cycles. This cycle length of one year is assumed to be constant throughout the model's time horizon.

We evaluate two alternative strategies: a no-treatment and a treatment strategy. The treatment improves utility for those individuals in the S1 state but has no impact on utility for those in the S2 state. However, due to the nature of this hypothetical disease, we are not able to distinguish those that are in S1 from those in S2. Therefore, under the treatment strategy, individuals who occupy states S1 or S2 receive treatment and continue doing so until they recover to the healthy state or die.

Figure 1



Schematic representation of the Sick-Sicker microsimulation model. In the extended microsimulation model the underlined probabilities (p.S1D and p.S2D) are adjusted to incorporate time-dependent transitions.

In the extension of the model that incorporates "memory" and individual-level characteristics, three modifications were made to the model. Firstly, we assumed that the benefits of treatment wane over time, with the utility of those in S1 on treatment decreasing by 0.03 every year they spend in S1. The second modification involved the dependency of mortality rates on the duration of remaining in the disease states (a 20% increase every cycle in S1 or S2, underlined probabilities). Finally, we also assumed that the improvement on quality of life by the treatment varies across individuals through a characteristic that acts as a treatment effect modifier. All model parameter values and R variable names are presented in Table 1.

Table 1

Input parameters for the illustrative microsimulation model.

Parameter	R name	Volume
Time horizon (nt)	`n.t`	30 years
Cycle length	`cl`	1 year
Number of simulated individuals (ni)	`n.i`	100,000
Names of health states (n)	`v.n`	H, S1, S2, D
Annual discount rate (costs/QALYs) (dc/de)	`d.c/d.e`	3%
Annual transition probabilities		
- Disease onset (H to S1)	`p.HS1`	0.15
- Recovery (S1 to H)	`p.S1H`	0.5
- Disease progression (S1 to S2)	`p.S1S2`	0.105
Annual risks of death		
- H to D	`p.HD`	0.005
- Rate ratio of death in S1 *vs* healthy	`rr.S1`	3
- Rate ratio of death in S2 *vs* healthy	`rr.S2`	10
Annual costs		
- Healthy individuals	`c.H`	$2,000
- Sick individuals in S1	`c.S1`	$4,000
- Sick individuals in S2	`c.S2`	$15,000
- Annual treatment cost per sick individual (S1 and S2)	`c.Trt`	$12,000
Utility weights		
- Healthy individuals	`u.H`	1.00
- Sick individuals in S1	`u.S1`	0.75
- Sick individuals in S2	`u.S2`	0.50
Intervention effect		
- Utility for treated individuals in S1 (SD)	`u.Trt`	0.95
Time varying extension of Sick-Sicker model (Figure 1)		
- Treatment effect modifier at baseline	`v.x`	*Uniform* (0.95, 1.05)
- Utility decrement of treated sick individuals with every additional year of being sick (S1 and S2)	`ru.S1S2`	0.03
- Proportional increase of the mortality rate with every additional year of being sick/sicker (S1 and S2)	`ru.S1S2`	0.02

The probabilities to die when sick (`p.S1D`) and sicker (`p.S2D`) are calculated by converting `p.HD` to a rate, $-\log(1 - \texttt{p.HD})$, multiply it by the rate ratios `rr.S1` and `rr.S2`, respectively, and then converting them back to a probability. In the extended model, not only the relative risks but also the variable `rp.S1S2` and the duration of being sick are used to adjust the rates. See the R code below for more details on the calculations.

In the next section we provide the R code and some documentation for the most important steps of the algorithm. We first illustrate the direct microsimulation implementation of the Sick-Sicker model and then provide the modifications necessary to extend the model to the scenario where "memory" is incorporated in the model. A full working version of the R codes used for each decision model can be found in the Appendices and on GitHub.

Microsimulation implementation of the Sick-Sicker model

The R code is broken down into three sections: the model input section, the model's main section, and the model output section. In the model input section, input parameters are loaded into the R environment. We assume that the health state of the individuals is known at the start of the simulation, and stored in the variable `v.M_1`. In our example, all individuals enter the model as healthy (i.e., at health state (H)).

First, we specify the number of individuals, their initial state, the time horizon, the health states, the discount rates and the intervention strategies.

```
n.i   <- 100000                          # number of simulated individuals
n.t   <- 30                              # time horizon, 30 cycles
v.n   <- c("H","S1","S2","D")            # model states: Healthy(H), Sick(S1),
Sicker(S2), Dead(D)
n.s   <- length(v.n)                     # the number of health states
v.M_1 <- rep("H", n.i)                   # everyone begins in the healthy state
d.c   <- d.e <- 0.03                     # equal discount of costs and QALYs by 3%
v.Trt <- c("No Treatment", "Treatment") # store the strategy names
```

The next lines of R code specify the transition probabilities, cost and utility input for individuals in the treatment and the no-treatment group (Table 1).

```
# Transition probabilities (per cycle)
p.HD    <- 0.005                  # probability to die when healthy
p.HS1   <- 0.15                   # probability to become sick when healthy
p.S1H   <- 0.5                    # probability to become healthy when sick
p.S1S2  <- 0.105                  # probability to become sicker when sick
rr.S1   <- 3                      # rate ratio of death in sick vs healthy
rr.S2   <- 10                     # rate ratio of death in sicker vs healthy
r.HD    <- -log(1 - p.HD)         # rate of death in healthy
r.S1D   <- rr.S1 * r.HD           # rate of death in sick
r.S2D   <- rr.S2 * r.HD           # rate of death in sicker
p.S1D   <- 1 - exp(- r.S1D)       # probability to die in sick
p.S2D   <- 1 - exp(- r.S2D)       # probability to die in sicker

# Cost and utility inputs
c.H     <- 2000                   # cost of remaining one cycle healthy
c.S1    <- 4000                   # cost of remaining one cycle sick
c.S2    <- 15000                  # cost of remaining one cycle sicker
c.Trt   <- 12000                  # cost of treatment (per cycle)

u.H     <- 1                      # utility when healthy
u.S1    <- 0.75                   # utility when sick
u.S2    <- 0.5                    # utility when sicker
u.Trt   <- 0.95                   # utility when being treated
```

In the model's main section, we define four R functions that constitute the core of the microsimulation model. In the following paragraphs we will be describing each component of these four functions.

The main function is the `MicroSim()` function that operationalizes the algorithm steps described above in R, adjusted to represent the Sick-Sicker model. The `MicroSim()` function includes arguments that control the input parameters (e.g., names of health states) and the components of the model itself (e.g., the seed number). It also provides the option to the modeler to record the transitions between states at every cycle in a TS matrix of size $ni \times (nt + 1)$ and the calculation of a microsimulation trace TR $((nt + 1) \times ns)$ which captures the proportion of individuals occupying each state at every cycle. Some arguments have a default setting, which means that in case the user does not specify them they will be assigned a default value (e.g., the arguments `TR.out` and `TS.out` that control whether a microsimulation trace and a transition array will be calculated are by default set to TRUE and the default seed number is 1). The `Trt` argument is a scalar with a Boolean value (by default set to FALSE), used to run the Sick-Sicker model for a group of treated or non-treated individuals. In other situations, modelers can adjust the code to make use of a vector with Boolean values to model mixed groups of treated and non-treated individuals. The following paragraphs describe each component of the `MicroSim()` function. The R code lines below describe the arguments and functions used in the `MicroSim()` function.

```
MicroSim <- function(v.M_1, n.i, n.t, v.n, d.c, d.e, TR.out = TRUE, TS.out =
TRUE, Trt = FALSE, seed = 1) {
# Arguments
  # v.M_1:        vector of initial states for individuals
  # n.i:          number of individuals
  # n.t:          total number of cycles to run the model
  # v.n:          vector of health state names
  # d.c:          discount rate for costs
  # d.e:          discount rate for health outcomes (QALYS)
  # TR.out:       should the output include a Microsimulation trace? (default is
                  TRUE)
  # TS.out:       should the output include a transition array between states?
                  (default is TRUE)
  # Trt:          are the n.i individuals receiving treatment? (scalar with a
                  Boolean value, default is FALSE)
  # seed:         starting seed number for random number generator (default is 1)
# Makes use of
  # Probs:        function for the estimation of transition probabilities
  # Costs:        function for the estimation of cost state values
  # Effs:         function for the estimation of state specific health outcomes
                  (QALYs)
```

In addition to the `MicroSim()` function, we construct three functions that need to be specified and executed prior to execution of the `MicroSim()` function. These are the `Probs()`, `Costs()` and `Effs()` functions which are used to update the probabilities, costs and health outcome values at each cycle, respectively. We will provide more details on these functions later in this tutorial.

Within the `MicroSim()` function two vectors with the discount weights are generated based on the two discount rates for costs and health outcomes, respectively `d.c` and `d.e`, specified in the argument section of the function. The first cycle is not discounted.

```
v.dwc <- 1 / (1 + d.c) ^ (0:n.t)    # calculate the cost discount weight based
                                    on the discount rate d.c
v.dwe <- 1 / (1 + d.e) ^ (0:n.t)    # calculate the QALY discount weight based
                                    on the discount rate d.e
```

In R, variables used to store information across iterations need to be initialized and their dimensions need to be declared outside of the iterative process. Hence, we first initialize the variables that will capture the matrices M, C, and E.

```
m.M <- m.C <- m.E <- matrix(nrow = n.i, ncol = n.t + 1,
                            dimnames = list(paste("ind", 1:n.i, sep = " "),
                                            paste("cycle", 0:n.t, sep = " ")))
```

Next, we specify the health states the individuals occupy at the start of the simulation.

```
m.M[, 1] <- v.M_1   # indicate the initial health state
```

Given M_{i0} we can calculate C_{i0} and E_{i0} for every individual i and store this information in `m.C` and `m.E`.

```
for (i in 1:n.i) {
    set.seed(seed + i)                      # set the seed for every
                                            individual for the random number
                                            generator
    m.C[i, 1] <- Costs(m.M[i, 1], Trt)      # estimate costs per individual
                                            of the initial health state
                                            conditional on treatment
    m.E[i, 1] <- Effs (m.M[i, 1], Trt)      # estimate QALYs per individual
                                            of the initial health state
                                            conditional on treatment
```


At the beginning of all the subsequent cycles t, the transition probabilities `v.p` given the states occupied at the beginning of the cycle are specified. The states M_{it} occupied by any individual i are sampled from a categorical distribution with frequency based on `v.p`. The next step is to calculate C_{it} and E_{it} for each individual i during each cycle t. The costs and QALYs only depend on the current state of the individual and on whether they are receiving treatment or not. Note that `m.C` and `m.E` are initialized, while `v.p` is not, as we are not interested in storing it for every iteration and it will therefore be overwritten at every cycle.

```
for (t in 1:n.t) {
v.p <- Probs(m.M[i, t])     # calculate the transition probabilities at cycle t
      m.M[i, t + 1] <- sample(v.n, prob = v.p, size = 1) # sample the next
                               health state and store that state in matrix m.M
      m.C[i, t + 1] <- Costs(m.M[i, t + 1], Trt)    # estimate costs per
individual during cycle t + 1 conditional on treatment
      m.E[i, t + 1] <- Effs (m.M[i, t + 1], Trt)    # estimate QALYs per
individual during cycle t + 1 conditional on treatment
```

The following lines of code display the progress of the simulation.

```
} # close the loop for the time points
    if(i/100 == round (i/100,0)) { # display the progress of the simulation
      cat('\r', paste(i/n.i * 100, "% done", sep = " "))
    }
  } # close the loop for the individuals
```

Once the iterative process is completed, the model's output is captured in the initialized variables. To calculate the discounted total costs and QALYs per individual across all cycles, we take the inner products of C and E with the discount weight vectors. The average total costs and QALYs within this patient population are estimated by averaging across the individual total costs and QALYs.

```
tc <- m.C %*% v.dwc                # total discounted cost per individual
  te <- m.E %*% v.dwe              # total discounted QALYs per individual

  tc_hat <- mean(tc)               # average discounted cost
  te_hat <- mean(te)               # average discounted QALYs
```

The function `MicroSim()` also offers the modeler the option to generate a microsimulation trace matrix TR and a matrix of transitions TS (i.e. argument set to TRUE).

```
if (TS.out == TRUE) {              # create a  matrix of transitions across
                                   states
    TS <- paste(m.M, cbind(m.M[, -1], "D"), sep = "->") # transitions from one
                                   state to the other
    TS <- matrix(TS, nrow = n.i)
    rownames(TS) <- paste("Ind",   1:n.i, sep = " ") # name the rows
    colnames(TS) <- paste("Cycle", 0:n.t, sep = " ") # name the columns
  } else {
    TS <- NULL
  }
if (TR.out == TRUE) {
    TR <- t(apply(m.M, 2, function(x) table(factor(x, levels = v.n, ordered =
TRUE))))
    TR <- TR / n.i                 # create a distribution trace
    colnames(TR) <- v.n            # name the rows
    rownames(TR) <- paste("cycle", 0:n.t, sep = " ") # name the columns
  } else {
    TR <- NULL
  }
```

Finally, all the function's output is stored and returned as a "list" object.

```
results <- list(m.M = m.M, m.C = m.C, m.E = m.E, tc = tc, te = te, tc_hat =
tc_hat, te_hat = te_hat, TS = TS, TR = TR)   # store the results from
                                             the simulation in a list
  return(results)                            # return the results
}                                            # end of the MicroSim function
```

Key components in any microsimulation structure are the functions used to update the transition probabilities and to estimate the costs and health outcomes for every cycle. Below we present the R code used to describe these functions for the Sick-Sicker model.

The `Probs()` function is used to update transition probabilities based on the health state occupied at cycle t as represented in Figure 1. The `M_it` argument in the context of the Sick-Sicker model represents a single character, representing the health state for individual i at cycle t, extracted from the `m.M` matrix.

```
Probs <- function(M_it) {
  # M_it:    health state occupied by individual i at cycle t (character
variable)

  v.p.it <- rep(NA, n.s)       # create vector of state transition probabilities
  names(v.p.it) <- v.n         # name the vector

  # update the v.p.it with the appropriate probabilities
# transition probabilities when healthy
  v.p.it[M_it == "H"]  <- c(1 - p.HS1 - p.HD, p.HS1, 0, p.HD)
# transition probabilities when sick
  v.p.it[M_it == "S1"] <- c(p.S1H, 1- p.S1H - p.S1S2 - p.S1D, p.S1S2, p.S1D)
# transition probabilities when sicker
  v.p.it[M_it == "S2"] <- c(0, 0, 1 - p.S2D, p.S2D)
# transition probabilities when dead
  v.p.it[M_it == "D"]  <- c(0, 0, 0, 1)
# return the transition probabilities or produce an error
  ifelse(sum(v.p.it) == 1, return(v.p.it),
  print("Probabilities do not sum to 1"))
}
```

In this simple example where there is no time variation in the transition probabilities, one could specify a transition probability matrix and rely on linear algebra to simplify the function `Probs()`. At the end of the function, there is a check that all probabilities sum to one. If this condition is not met, an error message appears.

In this example, the function `Costs()` depends on the current health state and treatment status of the individual. Individuals occupying states S1 and S2 have higher costs if treated compared to untreated individuals. If `Trt` is set to TRUE, the function will estimate the costs for the treatment strategy; otherwise, the costs for the no treatment group are estimated.

```
Costs <- function (M_it, Trt = FALSE) {
  # M_it: health state occupied by individual i at cycle t (character variable)
  # Trt:  is the individual being treated? (default is FALSE)
  c.it <- 0                                  # by default the cost for everyone
                                             is zero
  c.it[M_it == "H"]  <- c.H                  # update the cost if healthy
  c.it[M_it == "S1"] <- c.S1 + c.Trt * Trt   # update the cost if sick
                                             conditional on treatment
  c.it[M_it == "S2"] <- c.S2 + c.Trt * Trt   # update the cost if sicker
                                             conditional on treatment
  return(c.it)                               # return the costs
}
```

The function `Effs()` is used for the estimation of the health outcomes, QALYs in our example, during each cycle based on the current health state. If `Trt` is set to TRUE, the function will estimate the QALYs for the treatment strategy, this means that all individuals in the sample are treated when sick or sicker, otherwise the QALYs for the no treatment group are estimated. The variable `cl` indicates the cycle length.

```
Effs <- function (M_it, Trt = FALSE, cl = 1) {
  # M_it: health state occupied by individual i at cycle t (character variable)
  # Trt:  is the individual treated? (default is FALSE)
  # cl:   cycle length (default is 1)
  u.it <- 0                     # by default the utility for everyone is zero
  u.it[M_it == "H"]  <- u.H     # update the utility if healthy
  u.it[M_it == "S1"] <- Trt * u.Trt + (1 - Trt) * u.S1  # update the utility if
  sick conditional on treatment
  u.it[M_it == "S2"] <- u.S2    # update the utility if sicker
  QALYs <-  u.it * cl           # calculate the QALYs during cycle t
  return(QALYs)                 # return the QALYs
  }
```

Results of the simple Sick-Sicker microsimulation

After all the functions and parameters are defined, the `MicroSim()` function is executed and the outcomes are stored in the lists named `sim_no_trt` and `sim_trt` for the no-treatment and treatment strategies, respectively.


```
# Run the simulation for both no treatment and treatment
sim_no_trt <- MicroSim(v.M_1, n.i, n.t, v.n, d.c, d.e, Trt = FALSE) # run for
no treatment
sim_trt    <- MicroSim(v.M_1, n.i, n.t, v.n, d.c, d.e, Trt = TRUE)  # run for
treatment
```

The outcomes stored in these lists are useful to generate graphical representations of individuals' trajectories between health states over all cycles. Figure 2 shows the trajectories of three individuals of the Sick-Sicker model. The variation between the individual trajectories is the result of the stochastic nature of the microsimulation model. For example, an individual can stay healthy during the first four cycles, can become sick during cycle 5 and become sicker during cycle 6. From cycle 6 until 16 the individual remains in the sicker state and during cycle 17 the individuals dies (see top trajectory Figure 2).

The cost and QALY outcomes for these three individuals across the time horizon are represented in a graphical form in Figure 3.

Since this model was developed to answer an economic evaluation question, we need to calculate the incremental costs (ΔC) and QALYs (ΔE) and the incremental costs effectiveness ratio (ICER). The code below performs these calculations and stores the values in a table.

```
# store the mean costs (and the MCSE) of each strategy in a new variable v.C
(vector costs)
v.C  <- c(sim_no_trt$tc_hat, sim_trt$tc_hat)
sd.C <- c(sd(sim_no_trt$tc), sd(sim_trt$tc))/sqrt(n.i)
# store the mean QALYs(and the MCSE) of each strategy in v.E (vector health
outcomes)
v.E  <- c(sim_no_trt$te_hat, sim_trt$te_hat)
sd.E <- c(sd(sim_no_trt$te), sd(sim_trt$te))/sqrt(n.i)

delta.C <- v.C[2] - v.C[1]           # calculate incremental costs
delta.E <- v.E[2] - v.E[1]           # calculate incremental QALYs
sd.delta.C <- sd(sim_trt$tc - sim_no_trt$tc)/sqrt(n.i)  # Monte Carlo Squared
Error (MCSE) of incremental costs
sd.delta.E <- sd(sim_trt$te - sim_no_trt$te)/sqrt(n.i)  # Monte Carlo Squared
Error (MCSE) of incremental QALYs
ICER <- delta.C / delta.E           # calculate the ICER
results <- c(delta.C, delta.E, ICER)# store the values in a new variable
```

```r
# Create full incremental cost-effectiveness analysis table
table_micro <- data.frame(
  c(round(v.C, 0), ""),             # costs per arm
  c(round(sd.C, 0), ""),            # MCSE for costs
  c(round(v.E, 3), ""),             # health outcomes per arm
  c(round(sd.E, 3), ""),            # MCSE for health outcomes
  c("", round(delta.C, 0)),         # incremental costs
  c("", round(sd.delta.C, 0)),      # MCSE for incremental costs
  c("", round(delta.E, 3)),         # incremental QALYs
  c("", round(sd.delta.E, 3)),      # MCSE for health outcomes (QALYs) gained
  c("", round(ICER, 0)),            # ICER
)
rownames(table_micro) <- c(v.Trt, "* are MCSE values") # name the rows
colnames(table_micro) <- c("Costs", "*",  "QALYs", "*", "Incremental Costs",
"*", "QALYs Gained", "*", "ICER")   # name the columns
table_micro                         # print the table
```

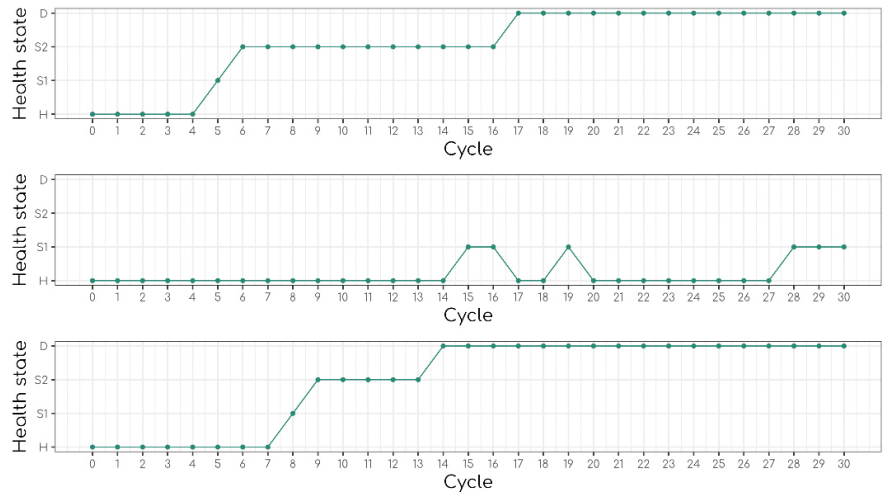
The **table_micro** variable includes the costs and QALYs per strategy, the incremental costs and QALYs, and the ICER of the treatment strategy. Table 2 presents the results of the cost-effectiveness analysis using the microsimulation model for a sample size of 10,000 and 100,000 individuals. For comparison purposes, the results based on the deterministic cohort model are also presented (R code in Supplementary Appendix C). We can observe that the microsimulation model and the cohort model produce almost identical results after a large number of simulations.

Table 2

	Total		Incremental		ICER
Strategies	Costs ($)	QALYs	Costs ($)	QALYs	($/QALY)
Microsimulation model (n = 10,000 and seed = 1)					
No-treatment	75,790 (577)	15.86 (0.049)	-	-	-
Treatment	141,211 (1080)	16.42 (0.051)	65,420 (517)	0.561 (0.004)	116,609
Microsimulation model (n = 100,000 and seed = 1)					
No-treatment	75,996 (183)	15.82 (0.016)	-	-	-
Treatment	141,644 (343)	16.38 (0.016)	65,648 (164)	0.561 (0.001)	117,087
Deterministic cohort model					
No-treatment	75,976	15.83	-	-	-
Treatment	141,623	16.40	65,647	0.562	116,901

Cost-effectiveness analysis results of the simple microsimulation model with two different population sizes (10,000 and 100,000) compared to results from a deterministic cohort model of the Sick-Sicker model. All microsimulation results were generated by setting the random seed to 1 at the beginning of the simulation. Monte Carlo Standard Error in brackets.

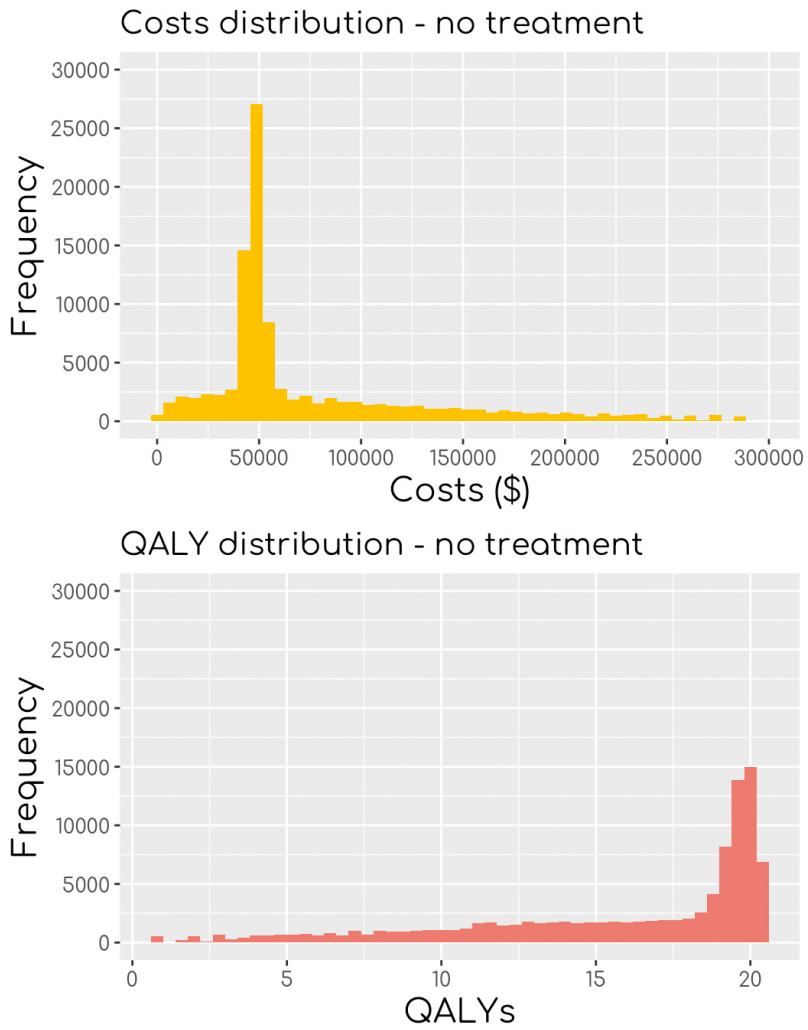
Figure 2 + 3



Trajectories between health states for three individuals in the Sick-Sicker microsimulation model demonstrating in which health state the individual occupied during each cycle of the simulation. In addition, this figure demonstrates the heterogeneity between individuals. Health state 1: Healthy (H), 2: Sick (S1), 3: Sicker (S2) and 4: Dead (D).

Graphical representation of the state costs (dark green squares, left y-axis) and QALY (light green dots, right y-axis) associated with individual trajectories of the first three individuals in the simple microsimulation model during all cycles.

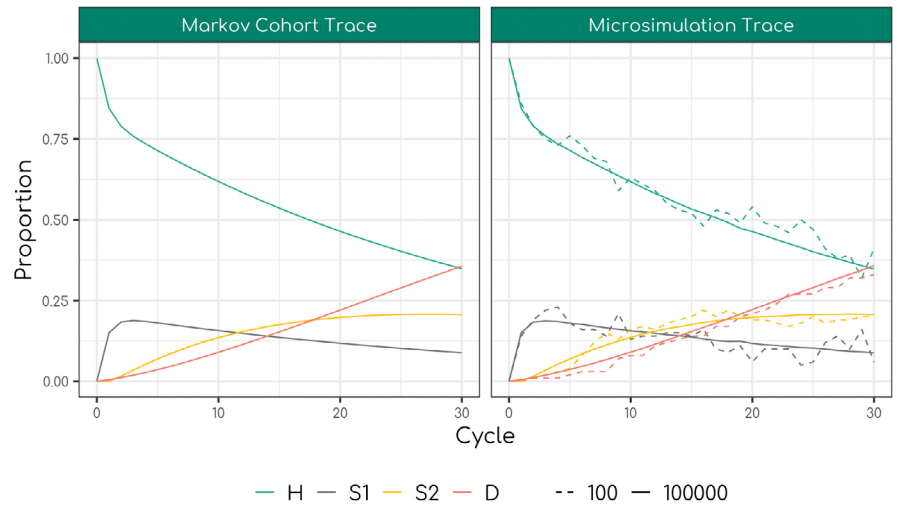
Figure 4



Histograms of the individual costs (top) and individual QALY (bottom) outcomes for the no-treatment strategy for the simple microsimulation model (n=100,000).

Since microsimulation models generate outcomes for each individual, it is possible to observe the distribution of the outcomes (costs and QALYs) by creating histograms (Figure 4).

Figure 5



Markov cohort trace (left) and the microsimulation trace of the simple microsimulation (right) for different numbers of individuals (dashed line: n=100; solid line: n=100,000). The y-axis represents the proportion of individuals in each health state. Since all individuals start healthy, the green line (H) starts at 1. Over time, the individuals transit from H (green line going down) towards other health states (other lines increases).

Comparing microsimulation to cohort model outcomes

Outcomes from microsimulation models where no "memory" and no heterogeneity at baseline across individuals is assumed, should asymptotically converge to those from a deterministic cohort model as the number of individuals simulated in the microsimulation model increases.[11,28,29] The number of hypothetical individuals that need to be generated is related to the magnitude of the MCSE. In Figure 5, we present the proportion of individuals occupying each health state at every cycle of both the microsimulation model and the cohort model. As illustrated in Figure 5, the microsimulation trace converges to the Markov cohort trace as the number of simulated individuals increases.

In addition, the Markov and microsimulation cost-effectiveness outcomes traces are graphically presented to illustrate the convergence of the microsimulation model to the outcomes derived using a cohort representation of the model (Figure 6).

Methods - Adding memory to the Sick-Sicker model

In this section we highlight two of the advantages of using a microsimulation implementation of a cohort model, which are to incorporate i) 'memory' into the disease dynamics and ii) variation in the baseline characteristics for every individual. To illustrate this, we extend the Sick-Sicker microsimulation model to include memory effects and patient heterogeneity at baseline. Specifically, we assume that mortality rate increases the longer a patient spends in on of the sick states (Figure 1 and Table 1) and that effectiveness of treatment is dependent on the duration of stay in the sick states and on baseline characteristics. Below we outline the modifications and additions that are necessary to incorporate these changes to the model. Supplementary Appendix B includes the complete R code used to build the "extended" microsimulation model.

We initially assume that individuals in our cohort have a baseline characteristic that acts as an effect modifier to the treatment effect. This baseline characteristic `x` is assumed to follow a uniform distribution within the population of interest.

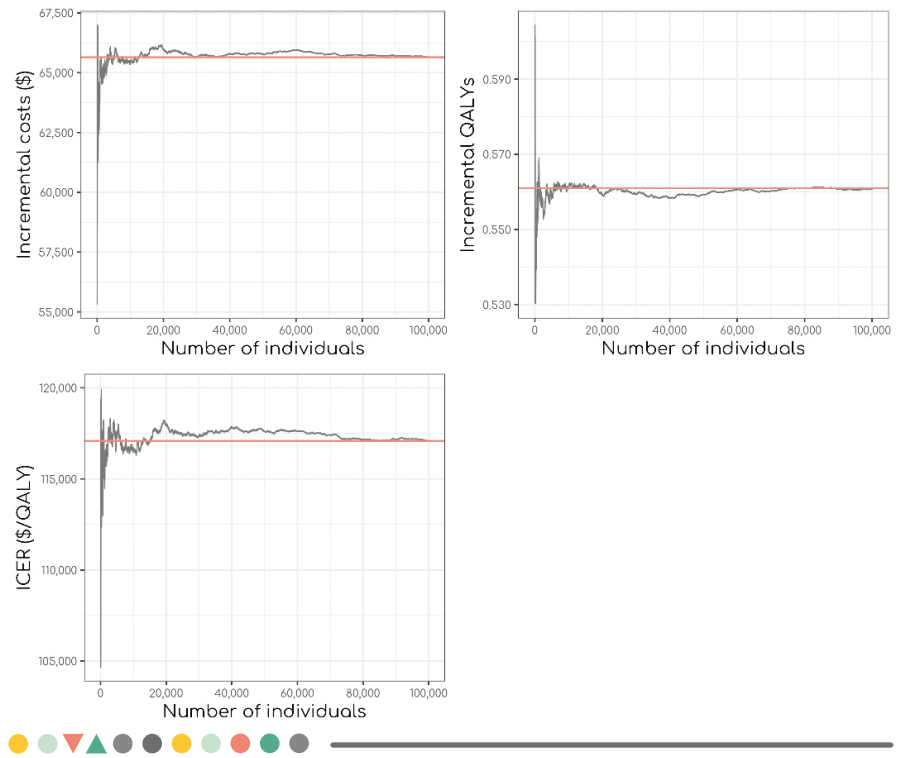
```
v.x <- runif(1, 0.95, 1.05) # vector capturing individuals' effect modifier at baseline
```

To be able to include this extra variable, the `MicSim()` function in the extended model has an extra argument `X`. This argument is assigned the name of the vector or matrix capturing any individual characteristics.

Additionally, we introduce two new variables: the annual increase in mortality rate (`rp.S1S2`) and the annual decrease in utility for treated sick patients (`ru.S1S2`) with every additional cycle spent in the sick states (S1 or S2).

```
rp.S1S2 <- 0.2 # increase of the mortality rate with every additional year being sick/sicker
ru.S1S2 <- 0.03 # decrease in utility of treated sick individuals with every additional year being sick/sicker
```

Figure 6



Cost-effectiveness analysis results from the simple microsimulation model with increasing number of individuals (up to n=100,000). The x-axis represent the number of individuals in the microsimulation model, the y-axis are the values for the incremental costs ($), incremental QALYs and ICER ($/QALY). Horizontal red line: cohort model results. Left top: Convergence of incremental costs, right top: convergence of QALYs left bottom: convergence of the ICER.

Subsequently, we introduce the help variable **dur** that stores the number of consecutive cycles the individual remains in either S1 or S2. Since every individual starts in the healthy state the dur variable takes the value 0 at cycle 0. With every additional cycle that an individual spend in S1 or S2, a value of 1 is added to **dur**. When the individual recovers (transitions back to state H), the **dur** variable is set again to zero. This means that all previous cycles spent in one of the sick health states does not influence the current transition probabilities. The process is repeated every time an individual transition to S1.

```
dur <- 0                            # the individual starts without history
m.C[i, 1] <- Costs(m.M[i, 1], Trt)  # estimate the cost per individual of the
initial health state
m.E[i, 1] <- Effs(m.M[i, 1], dur, Trt, X = X[i])  # estimate the health outcome
per individual at the initial health state conditional on treatment, duration
of being sick/sicker and individual characteristics

for (t in 1:n.t) {
      v.p      <- Probs(m.M[i, t], dur)  # calculate the transition
probabilities at cycle t conditional on the duration of being sick/sicker
      m.M[i, t + 1] <- sample(v.n, prob = v.p, size = 1)  # sample the new
health state and store that state in matrix m.M
      m.C[i, t + 1] <- Costs(m.M[i, t + 1], Trt)     # estimate the cost per
individual during cycle t + 1 conditional on treatment
      m.E[i, t + 1] <- Effs(m.M[i, t + 1] , dur, Trt, X = X[i])   # estimate
the health outcome per individual during cycle t + 1 conditional on treatment,
duration of being sick/sicker and individual characteristics

      if (m.M[i, t + 1] == "S1" | m.M[i, t + 1] == "S2"){         # expression
to identify sick/sicker individuals
        dur <- dur + 1              # updated the duration of being sick/sicker
      } else {
        dur <- 0}                   # reset duration variable
```

We introduced minor modifications on the **Probs()** function to allow for memory with regards to the probability of transitioning to death from S1 and S2. The modified function is presented below.


```
Probs <- function(M_it, dur) {
  # M_it: health state occupied by individual i at cycle t (character
variable)
  # dur:  the duration of being sick (sick/sicker)
  v.p.it <- rep(NA, n.s)           # create vector of state transition
probabilities
  names(v.p.it) <- v.n             # name the vector
  # update probabilities of death after first converting them to rates and
applying the rate ratio
  r.S1D <- -log(1 - p.S1D)
  r.S2D <- -log(1 - p.S2D)
  p.S1D <- 1 - exp(-r.S1D * (1 + dur * rp.S1S2)) # calculate p.S1D
conditional on duration of being sick/sicker
  r.S2D <- 1 - exp(-r.S2D * (1 + dur * rp.S1S2)) # calculate p.S2D
conditional on duration of being sick/sicker

  # update the v.p.it with the appropriate probabilities
# transition probabilities when healthy
  v.p.it[M_it == "H"] <- c(1 - p.HS1 - p.HD, p.HS1, 0, p.HD)       #
transition probabilities when sick
  v.p.it[M_it == "S1"] <- c(p.S1H, 1- p.S1H - p.S1S2 - p.S1D, p.S1S2, p.S1D)
# transition probabilities when sicker
  v.p.it[M_it == "S2"] <- c(0, 0, 1 - p.S2D, p.S2D)
# transition probabilities when dead
  v.p.it[M_it ==  "D"] <- c(0, 0, 0, 1)    }
# return the transition probabilities or produce an error
  ifelse(sum(v.p.it) == 1 ), return(v.p.it),
  print("Probabilities do not sum to 1"))
}
```

The next step is to estimate the cost and QALYs during each cycle. The new **Effs()** function includes the effect modifier **v.x,** the variable **dur** and **ru.S1S2** as input parameters.

```
Effs <- function (M_it, dur, Trt = FALSE, cl = 1, X = NULL) {
  # M_it: health state occupied by individual i at cycle t (character
variable)
  # dur:  the duration of being sick/sicker
  # Trt:  is the individual being treated? (default is FALSE)
  # cl:   the cycle length (default = 1)
  # X:    the vector or matrix of personal characteristics (optional)
  u.it <- 0                    # by default the utility for everyone is zero
  u.it[M_it == "H"] <- u.H     # update the utility if healthy
  # update the utility is sick conditional on treatment and duration of being
sick/sicker
  u.it[M_it == "S1"] <- X * Trt * (u.Trt - dur * ru.S1S2) + (1 - Trt) * u.S1
  [M_it == "S2"]    <-  u.S2        # update the utility if sicker
  QALYs             <- u.it * cl    # calculate the QALYs at cycle t
  return(QALYs)                     # return the QALYs
  }
```

No modification on the costs are needed, hence the function `Costs()` remained the same as the simple microsimulation model. The only thing left is to run the model with all the adjustments.

```
# Run the simulation for both no treatment and treatment
# run for no treatment
sim_no_trt <- MicroSim(v.M_1, n.i, n.t, v.n, X = v.x,  d.c, d.e, Trt = FALSE)
# run for treatment
sim_trt    <- MicroSim(v.M_1, n.i, n.t, v.n, X = v.x, d.c, d.e, Trt = TRUE)
```

Results of the extended Sick-Sicker model

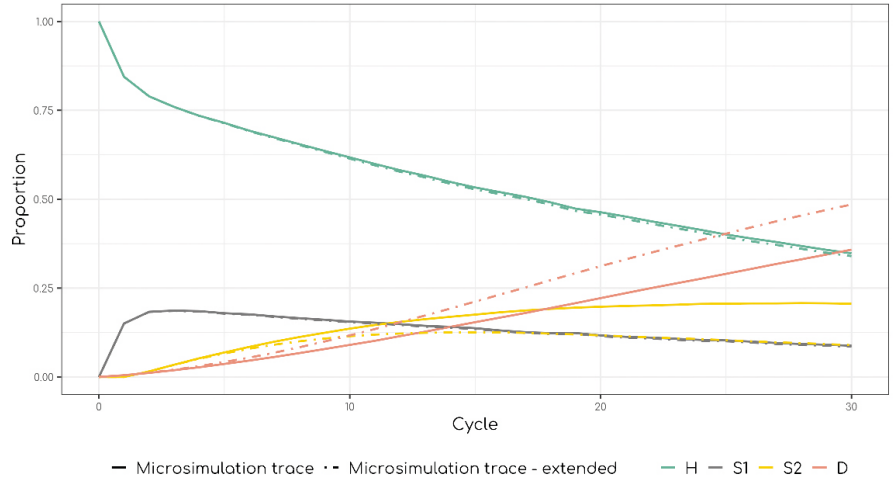
The cost-effectiveness analysis based on the extended microsimulation model is presented in Table 3. The expected costs, QALYs and ICER of this simulation are lower in both the no-treatment and treatment strategy compared with the simple microsimulation (and cohort model) (Table 2 and Table 3).

Table 3

Cost-effectiveness analysis results for the extended microsimulation model for the Sick-Sicker model with a population size of 100,000 individuals. The microsimulation results were generated by setting the random seed to 1 at the beginning of the simulation. Monte Carlo Standard Error in brackets.

	Total		Incremental		ICER
Strategies	Costs ($)	QALYs	Costs ($)	QALYs	ICER ($/QALY
Extended microsimulation model (n = 100,000 and seed = 1)					
No-treatment	62,667 (120)	15.28 (0.017)	-	-	-
Treatment	117,455 (231)	15.79 (0.017)	54,787 (117)	0.507 (0.001)	107,986

Figure 7



Trace of the simple microsimulation (solid line) and the extended microsimulation (dashed line) during all cycles. The y-axis represents the proportion of individuals in each health state. Since all individuals start healthy the solid line starts at 1. Over time, the individuals transit from healthy (H, green line going down) towards other health states; sick (S1, gray line), sicker (S2, yellow line) and dead (D, red line) (other lines increases). The solid and dashed line of the states H (green) and S1 (gray) overlap.

The results are consistent with the modifications that were applied to the model. With an increasing probability of death for each additional year spent in the sick health states, patient life-expectancy is reduced, resulting in fewer QALYs as well as lower healthcare costs. The gains in QALYs due to treatment are also reduced compared to the simple model due to the decreasing effectiveness of treatment over time in the extended model. The microsimulation traces presented in Figure 7 confirm the increased mortality rate. Finally, the uncertainty around the health outcomes, QALYs, has increased due to the introduction of the treatment effect modifier.

Vectorized implementation of the microsimulation model

A microsimulation process was presented above as an iterative process at an individual level over the model's time horizon. Because of the causal dependence of future events on past event histories in the simulation, temporal iterations in the model cannot be easily vectorized. However, the repetitive process of a microsimulation across the individuals can be intuitively represented in a vectorized form. In this case, at each time iteration, all simulated individuals are transitioned simultaneously through the model. Vectorization can improve computational performance, but often results in decision models with high-dimensionality components. For example, what was a vector of transition probabilities for an individual on a given point in time will become a matrix of probabilities, with rows the number of individuals and columns the number of states. Such higher dimensionality sometimes increases complexity of the model structure. A frequently used tool in vectorization is linear algebra as it can provide a convenient way of performing operations in a vector or matrix scale. Although linear algebra can facilitate quick calculations, it can at the same time generate more complex coding and larger matrices that will require more computer memory. Hence, it is often the case with vectorization that modelers must achieve a balance between conceptual complexity, memory, and computational efficiency.

Vectorization can be achieved in the context of our microsimulation example by extending the dimension of the health state sampling process. The function `sample()`, that is used in this tutorial, can support a vectorized solution only if the state transition probabilities are the same across the individuals. However, in a microsimulation context, individuals' risk often depends on a number of individual specific characteristics, which restrict us from using the `sample()` function on a vectorized context. To overcome this, and take advantage of vectorization solutions, we developed the `samplev()` function by modifying a random number generating function for multinomial variables from the `Hmisc` package.[32] The modification allows for a health-state specific vectorization solution based on vectors of random numbers generated from a U(0,1) distribution. The `samplev()` function randomly draws the state name at $t + 1$ based on the probability of each individual occupying each state at $t + 1$.

To make use of the `samplev()` function, the `Probs()` function must also be modified to return a matrix of transition probabilities for the whole cohort rather than a vector of probabilities for a single individual at a time. The output of the modified `Probs()` function is therefore a matrix of size $ni \times ns$. Since the `samplev()` function samples the name of the future states of individuals conditional on their probability of experiencing these states, state names are required as column names. In Appendix D, the code with all necessary modifications is presented.

The new `samplev()` function achieves a 97% reduction in the time needed to run the analysis (Table 4). The results in Table 4 are a time comparison, between the iterative standard approach for the simple Sick-Sicker model using the `sample()` function (Appendix A), and the vectorized approach for the simple Sick-Sicker model using the `samplev()` function (Appendix D). The time shown is the time (average of 3 runs) it takes to run the model for both the treatment and no treatment arm using the `sample()` or `samplev()` function.

Table 4

Time comparison between the iterative (standard) approach using `sample()` and the vectorized approach using `samplev()` to run the simple Sick-Sicker model.

	Time to run (in seconds)	
Sample size	`sample()`	`samplev()`
1,000	5.42	0.16
10,000	38.41	1.21
mt	378.76	11.71
M	4538.80	128.79

Computer: MacBook Pro, macOS Sierra version 10.12.6 Processor 2.3 GHz Intel Core i5, 2 cores, 4GB RAM, 1333 MHz DDR3.

Discussion

This tutorial provides guidance on the implementation of a microsimulation model using a programming language such as R. We outline the conceptual steps involved and provide an algorithm to operationalize these steps using a programming language. In addition, the tutorial illustrates an implementation of microsimulation models in R and presentation of results in graphical form. This tutorial focuses on the implementation rather than the conceptualization of a microsimulation model. For interested readers, Roberts et al. present recommendations and best practices regarding the process of decision model conceptualization.[33] In this paper we focused on state-transition models with discrete time intervals. Another class of models consider time as continuous and simulate discrete events. Researchers interested in the implementation of discrete event simulation models might benefit by making themselves familiar with the `simmer` package. The core of the `simmer` package is written in C++ with automatic monitoring capabilities, which makes it a fast and robust framework to develop DES models and continuous time Markov models.[34]

There are limited examples in the literature where researchers relied on R to build a microsimulation model for health decision-making. However, most of these examples have appeared recently, which is consistent with the observed trend of a recent increase in the use of R in health decision modeling.[19] Manca et al. published an early example of a microsimulation model built in R[35] and Choi et al. more recently published a hypertension, stroke and myocardial infarction microsimulation model.[36] For two other microsimulation models, MILC and CANTRANCE, the authors created R packages, which include all components of the model.[37,38] Thus, R can be used not only to develop the model, but also to provide an open-source platform for distributing the model to other users.

Building a microsimulation model in a programming language, such as R, has several advantages. Models are written in script form that facilitates readability and reproducibility. The fact that R is open-source and freely available increases transparency and reproducibility of the analysis. In addition, due to R's architecture, efficient computational performance can be achieved. The output obtained from the model in R is not limited to the output generated by the proprietary software; it is possible to extract every specific outcome estimate (e.g., incidence, prevalence and average time spent in a certain state). However, decisions on the generated output should be limited to only what is ultimately necessary for the analysis to minimize computational and memory storage demands. Finally, statistical analyses performed to provide input to the decision model could be embedded directly within the decision model.

R may not always be the most appropriate platform of choice, depending on the nature of the microsimulation model and the skills of the researchers. R is also not necessarily an ideal language for computationally intensive processes. Other programming languages such as C/C++, Python or Fortran may be able to achieve significantly better performance. Combining the strengths of different languages is sometimes possible. For example, the package `Rcpp` makes it possible to interface C++ code in R functions and packages.[39]

The flexibility of R implies that there can be multiple approaches to the operationalization of a microsimulation model. This tutorial provides one such approach, but it is not necessarily the most efficient. The implementation of the Sick-Sicker model was developed with a focus on clarity rather than computational efficiency. For example, one could employ matrix operations more extensively rather than relying on iterative or logical procedures. Avoiding iterative loops through vectorization in R has been shown to increase efficiency by multifold[40], as confirmed by our results (Table 4). When iterative loops are inevitable, use of parallel processing can drastically improve computational times for iterative procedures. Alternatively, a microsimulation model in R can be developed using object-oriented programming, using S4 or References Classes. Under this approach, simulated individuals, together with their characteristics, are encoded as a class and followed over time.

In such a modeling approach interaction between individuals can be more straightforwardly incorporated. Implementation of an objected oriented microsimulation in R can be facilitated through the use of the `simecol` package.[41]

We encourage modelers to develop and distribute implementations that are more efficient than the one presented in this tutorial. We are also in the process of developing an additional tutorial on advanced methods around microsimulation using R. This upcoming tutorial will have a strong focus on computational efficiency. However, the modeler should also be aware of the tradeoff between the time savings associated with an efficient algorithm and the time investment required to achieve such efficiency. One way to reduce computational time in microsimulation models is the use of parallel processing techniques. Most modern computers are equipped with multi-core processors, which increase significantly their computational power. However, by default R uses only a single core, thereby limiting its computational capacity. With the use of appropriate packages (e.g., `parallel`) R can achieve parallel processing capabilities, and thereby drastically reduce computational times.[42] One caveat of parallel processing is that the interaction of R with each of the core processors is time consuming, which modelers should be aware of when employing parallel processing techniques. Overall, it is important that the additional complexity and flexibility introduced using a programming language be justified by the complexity needed to answer the research question.

There is always a certain degree of uncertainty in the parameter values used in health decision models. Such parameter uncertainty can be incorporated in a decision model via probabilistic sensitivity analysis (PSA) (second-order Monte Carlo simulation).[7,43] The microsimulation approach presented above is only limited to the random variation of the individuals (first-order uncertainty) and ignores any parameter uncertainty. Efficient programming is particularly important in models that incorporate both individual variation and parameter uncertainty. Models where PSA is incorporated, will require particularly long computational times. Using efficient methods when conducting PSA in R is a topic of relevance to any form of decision analysis. Therefore, we started to prepare a separate tutorial on incorporating sensitivity analysis in decision modeling using R.

In summary, R is increasingly used for building simulation models in health decision sciences, however, information on how to perform these simulations is lacking. This tutorial provides a step-by-step guide to implementing a microsimulation model in R with the aim of supporting health decision scientists who are new to high-level programming languages to develop models in a more flexible, open-source and transparent manner, and encouraging increased transparency and reproducibility in health decision sciences. In subsequent tutorials, and as part of the Decision Analysis in R for Technologies in Health (DARTH) group efforts, we will expand on model optimization, calibration, and value of information analysis among other topics. Future code updates will be placed on our GitHub https://github.com/DARTH-git/Microsimulation-tutorial.

Supplemental material

Supplemental material for this article is available on https://pubmed.ncbi.nlm.nih.gov/29587047/ and the GitHub repository mentioned in the manuscript.

The coding convention in this chapter is not in line with our recommended good coding practices as presented in Chapter 2 of this thesis. We recommend readers to apply the coding convention from Chapter 2. For the newest code, please contact DARTH.

References

1. Friedman SG. Current management of the patient with internal carotid artery occlusion. *Eur J Vasc Surg*. 1989; 3(2):97–101.
2. Kattan M. *Encyclopedia of Medical Decision Making*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. Epub ahead of print 2009. DOI: 10.4135/9781412971980.
3. Kreke JE, Schaefer AJ, Roberts MS. Simulation and critical care modeling. *Curr Opin Crit Care*. 2004; 10(5):395–398.
4. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012; 32(5):678–689.
5. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health*. 2012; 15(6):796–803.
6. Statistics Canada. Microsimulation approaches, http://www.statcan.gc.ca/eng/microsimulation/modgen/new/chap2/chap2 (2016, accessed July 13, 2016).
7. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. Decision Making in Health and Medicine. Integrating Evidence and Values. 2nd ed. *Cambridge University Press*, 2014.
8. Lay-Yee R, Cotterell G. *The role of microsimulation in the development of public policy*. 2015. Epub ahead of print 2015. DOI: 10.1007/978-3-319-12784-2_14.
9. 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. *Med Decis Making*. 2012; 32(5):690–700.
10. Zucchelli E, Jones AM, Rice N. The evaluation of health policies through microsimulation methods. *Heal Econom Data Gr Work Pap*. 2010; 5:2–20.
11. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006; 15(12):1295–1310.
12. Davis S, Stevenson M, Tappenden P, Wailoo AJ. *Cost-effectiveness modelling using patient-level simulation*, http://www.nicedsu.org.uk/TSD15_Patient-level_simulation.pdf (2014).
13. Wolfson MC. POHEM--a framework for understanding and modelling the health of human populations. *World Health Stat* Q. 1994; 47(3–4):157–176.
14. Hennessy DA, Flanagan WM, Tanuseputro P, Bennett C, Tuna M, Kopec J, Wolfson MC, Manuel DG. The Population Health Model (POHEM): an overview of rationale, methods and applications. *Popul Health Metr*. 2015; 13(1):24.
15. Statistics Canada. Health models: Population health model, http://www.statcan.gc.ca/eng/microsimulation/health/health (2016, accessed May 24, 2016).
16. Goldman DP, Shekelle PG, Bhattacharya J, Hurd M, Joyce GF, Lakdawalla DN, Matsui DH, Newberry SJ, Panis CWA, Shang B. *Health status and medical treatment of the future elderly: Final Report*, http://www.rand.org/pubs/technical_reports/TR169.html (2004).
17. Muhlberger N, Kurzthaler C, Iskandar R, Krahn MD, Bremner KE, Oberaigner W, Klocker H, Horninger W, Conrads-Frank A, Sroczynski G, Siebert U. The ONCOTYROL Prostate Cancer Outcome and Policy Model: Effect of Prevalence Assumptions on the Benefit-Harm Balance of Screening. *Med Decis Making*. 2015; 35(August):758–772.
18. Hollman C, Paulden M, Pechlivanoglou P, McCabe C. A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. *Pharmacoeconomics*. 2017; 35(8):817–830.
19. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making*. 2017; :272989X16686559.
20. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: http//www.R-project.org/., 2015.
21. The Popularity of Data Science Software | r4stats.com, http://r4stats.com.
22. The 2016 Top Programming Languages - IEEE Spectrum.
23. Baio G. *Bayesian Methods in Health Economics*. 1st ed. London: Chapman and Hall/CRC, 2012.
24. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-Effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making*. 2016; 37(4):340–352.
25. Weeks M, Mitton L, Sutherland H. Microsimulation Modelling for Policy Analysis Challenges and Innovations. In: *Cambridge University Press*. 2000, pp. 1–11.
26. Soares MO, Canto e Castro L. Simulation or cohort models? Continuous time simulation and discretized Markov models to estimate cost-effectiveness. 2010; :1–20.
27. Sawilowsky SS, Fahoome GC. Statistics via *Monte Carlo Simulation with Fortran*. Rochester Hills, MI: JMASM, 2003.
28. Kurtz TG. Limit Theorems for Sequences of Jump Markov Processes Approximating Ordinary Differential Processes. *J Appl Probab*. 1971; 8(2):344–356.
29. Whitt W. *Stochastic-Process limits: An introduction to stochastic-process limits and their application to queues*. New York, NY: Springer, http://www.columbia.edu/~ww2040/preflongno.pdf (2002).
30. Tezuka S. *Uniform Random Numbers: Theory and Practice*. Boston, MA: Springer US. Epub ahead of print 1995. DOI: 10.1007/978-1-4615-2317-8.

31. Enns EA, Cipriano LE, Simons CT, Kong CY. Identifying Best-Fitting Inputs in Health-Economic Model Calibration: A Pareto Frontier Approach. *Med Decis Making*. 2015; 35(2):170–182.
32. Harrell FE. Hmisc: Harrell Miscellaneous. 2017; :R package version 4.0-3.
33. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, ISPOR-SMDM Modeling Good Research Practices Task Force. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making*. 2012; 32(5):678–689.
34. Ucar I, Smeets B. simmer: Discrete-Event Simulation for R. 2017; :R package version 3.6.3.
35. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: The importance of controlling for baseline utility. *Health Econ*. 2005; 14(5):487–496.
36. Choi SE, Brandeau ML, Basu S. Expansion of the National Salt Reduction Initiative: A Mathematical Model of Benefits and Risks of Population-Level Sodium Reduction. *Med Decis Making*. 2015; 36(1):1–14.
37. Chrysanthopoulou SA. MILC: MIcrosimulation Lung Cancer (MILC) model.
38. Birnbaum JK, Ademuyiwa FO, Carlson JJ, Mallinger L, Mason MW, Etzioni R. Comparative Effectiveness of Biomarkers to Target Cancer Treatment: Modeling Implications for Survival and Costs. *Med Decis Making*. 2015; :1–10.
39. Eddelbuettel D, François R. Rcpp : Seamless R and C++ Integration. J Stat Softw. 2011; 40(8):1–18.
40. Gillespie C, Lovelace R. *Efficient R Programming: A Practical Guide to Smarter Programming*. 1st ed. O'Reilly Media, 2016.
41. Petzoldt T, Rinke K. simecol : An Object-Oriented Framework for Ecological Modeling in *R*. *J Stat Softw*. 2007; 22(9):1–31.
42. R Core Team. *Package 'parallel.'* Requires R version 2.14.0. Package URL: https://stat.ethz.ch/R-manual/R-devel/library/parallel/doc/parallel.pdf, 2015.
43. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD, ISPOR-SMDM Modeling Good Research Practices Task Force. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making*. 2012; 32(5):722–732.

PART II

RESOURCE PRIORITIZATION IN HEALTHCARE

08

MINIMIZING POPULATION HEALTH LOSS IN TIMES OF SCARCE SURGICAL CAPACITY DURING THE COVID-19 CRISIS AND BEYOND: A MODELING STUDY

Benjamin Gravesteijn*
Eline Krijkamp*
Jan Busschbach
Geert Geleijnse
Isabel Retel Helmrich
Sophie Bruinsma
Céline van Lint
Ernest van Veen
Ewout Steyerberg
Kees Verhoef
Jan van Saase
Hester Lingsma
Rob Baatenburg de Jong
and Value Based Operation Room Triage team collaborators**

*Both authors contributed equally

** Value Based Operation Room Triage team collaborators: Chris Bangma, Ivo Beetz, Patrick Bindels, Alexandra Brandt-Kerkhof, Danielle van Diepen, Clemens Dirven, Tjebbe Galema, Jeanette Goudzwaard, Mieke Hazes, Sjoerd Lagarde, Harmke Polinder-Bos, Eva Maria Roes, Hanneke Takkenberg, Mark van Vledder

Value in Health 2020

MINIMIZING POPULATION HEALTH LOSS IN TIMES OF SCARCE SURGICAL CAPACITY DURING THE COVID-19 CRISIS AND BEYOND: A MODELING STUDY

Abstract

COVID-19 has put unprecedented pressure on healthcare systems worldwide, leading to a reduction of the available healthcare capacity. Our objective was to develop a decision model to estimate the impact of postponing semi-elective surgeries on health, to support prioritization of care from a utilitarian perspective.

A cohort state-transition model was developed and applied to 43 semi-elective non-pediatric surgeries commonly performed in academic hospitals. Scenarios of delaying surgery from two weeks were compared with delaying up to one year, and no surgery at all. Model parameters were based on registries, scientific literature, and the World Health Organization global burden of disease study. For each surgery, the model estimated as the average expected Disability-Adjusted Life-Years (DALYs) per month.

"Crisis communication is so hard because you are kind of building the boat as you are sailing."

Carlos del Rio

Given the best available evidence, the two surgeries associated with most DALY's due to surgery delay were bypass surgery for Fontaine III/IV peripheral arterial disease (0.14 DALY/month, 95%-CI: 0.08-0.22) and total nefrectomy for renal cancer (0.09 DALY/month, 95%-CI: 0.07-0.11). The two surgeries with least DALY's were placing a shunt for dialysis (0.01, 95%-CI: 0.004-0.01) and bullectomy for COPD (0.01, 95%-CI: 0.01-0.01).

Expected health loss due to surgical delay can be objectively calculated with our decision model based on best available evidence, which can guide prioritization of surgeries to minimize population health loss in times of scarcity. This model results should yet be placed in the context of different ethical perspectives and combined with capacity management tools to facilitate large-scale implementation.

Introduction

COVID-19 has put unprecedented pressure on healthcare systems worldwide. The healthcare demand of this pandemic supersedes available healthcare capacity, far beyond the demand that was imposed by the 2017 influenza pandemic.[1,2] The pressure on the available healthcare capacity impacts the continuity of regular care. Amongst others because (1) wards and operating theaters are converted to COVID-19 care facilities,[3] (2) physicians are deployed to care for COVID-19 patients,[4,5] and (3) the fear of contagion with the SARS-CoV-2 virus may leave susceptible patients reluctant to seek care[4,5], as was seen in similar health crises like the SARS epidemic in 2003.[6]

Delay in surgical care may dramatically impact healthcare quality and accessibility. In the first weeks of the COVID-19 crisis in the Netherlands, 75-90% fewer surgeries were performed compared to previous years.[7] The delay in cancer surgery already has made a large impact in the life expectancy of oncological patients.[8] Moreover, it may be impossible to treat the whole accumulating group of patients in the near future, as estimated for orthopedic and cardiothoracic surgery in the US.[9,10] Because of these problems, hospitals are facing a dilemma: Which patients should be prioritized?

As stated by Emanuel et al., "*The question is not whether to set priorities, but how to do so ethically and consistently, rather than basing decisions on individual institutions' approaches or a clinician's intuition in the heat of the moment*".[2] In practice, individual surgical patients are most often triaged by experts from the respective surgical fields.[11] Unfortunately the level of agreement on prioritization between experts is low.[12] Additionally, prioritization across disciplines is complicated by the high degree of specialization in modern medicine. Most importantly, this approach does not systematically optimize population health. While the perspective of maximizing population health, a utilitarian ethical perspective which strives to achieve the greatest good for the greatest number,[13] has been described to be most defendable in times of scarcity.[2,14–18]

Therefore, to guide prioritization of semi-elective surgeries[i] across disciplines from a utilitarian perspective, our study aims to develop a decision model to estimate the impact of postponing surgery on health.

Methods

Overview

This study focused on semi-elective surgeries, because urgent procedures always have priority over other procedures, and elective procedure can, by definition, be delayed. The most frequently performed semi-elective surgeries in our institute were selected. Data about these surgeries were collected and used in a broadly applicable computer-based model to estimate the effect of surgical delay on life expectancy and health related quality of life (QoL).

[i] A semi-elective surgery is defined as a surgery that should ideally be performed within three days up to three weeks.

Patients and setting

The evaluated surgeries in this study comprise of non-pediatric and non-obstetric, semi-elective surgeries in Erasmus University Medical Center, an academic tertiary referring hospital in the Netherlands. From the electronic patient registry (HiX, ChipSoft), the number of surgeries, surgery time, length of stay at an intensive care unit (ICU), and length of stay at a non-ICU of all non-urgent surgeries were retrieved from July 2017 to December 2019. Next, two senior clinicians selected the semi-elective surgeries from this list. No objective criterion could be used, because the definition of semi-elective is subject to numerous aspects including environment, expert opinion, and alternative therapies. Finally, the Value Based Operation Room Triage team collaborators approved the selection. This team of collaborators was a diverse expert panel of 18 healthcare professionals from the Erasmus University Medical Center, including surgeons (e.g., cardiothoracic surgeons, neurosurgeons, and gynecologic surgeons) as well as generalists (e.g., internists, geriatricians and GPs) (Appendix C). Ultimately, 43 semi-elective surgeries were selected that were performed more than 80 times during the time window we used. Where relevant, mild and severe cases undergoing the surgery were distinguished based on clinical insight of our collaborators.

Model input parameters

The model required 7 input parameters: (1) survival rates pre-surgery, (2) survival rate post-surgery, (3) QoL pre-surgery, (4) QoL post-surgery, (5) mean age of patients undergoing the surgery, (6) time until no effect of treatment can be expected on survival or (7) time until no effect of treatment can be expected on QoL (see Appendix A). The class of collected evidence was defined as class I (Randomized Controlled Trials (RCT) or systematic reviews of RCTs), class IIa (Prospective observational studies, before-after studies), class IIb (Retrospective observational studies, expert panels for the disutility weights, national registries), and class III (expert opinion).

Survival input

The survival rates post-surgery were obtained from national registries for oncological[19] and cardiothoracic[20] surgeries. For the remaining surgeries, data was obtained from scientific literature. The survival data pre-surgery for all surgeries were based on data from published studies. If either survival with or without treatment was lacking, the reported treatment effect (preferably from an RCT) was used to calculate the missing survival parameter. The disease specific mortality was added to the national overall age-specific mortality from the Central Bureau of Statistics in the Netherlands.[21] The mean age of the patients was obtained from published studies. All survival data had to be converted to mortality risk per week and ultimately converted to probabilities to be used in the model (formulas presented in Appendix C).[22]

Quality of life input

The disability weights from the Global Burden of Disease (GBD) Study 2016 from the World health organization (WHO) were used to value the QoL of our health states.[23] The GBD Study reports disability weights for nonfatal health conditions. These weights represent the magnitude of health loss associated with the conditions, where 0 represents no loss (full health) and 1 all lost (death). When these weights are multiplied with the duration lived in this conditions, one has calculated the weighted 'years lived with disability' (YLD).[24] The YLD summed with the years of life lost to premature death (YLLs) give the disability adjusted life years (DALY).[25] A 'full DALY' can be thought of as losing one year in full health.

Where possible, we based the disability weights of health conditions directly on the GBD study data. The remaining conditions were estimated using methods described by Stouthard et al. with the Value Based Operation Room Triage team collaborators.[26] We used a visual analogue scale (VAS) calibrated with GBD 2016 QoL weights. Like in the study from Stouthard, we framed our VAS with 1 being the best imaginable health state and zero the worst/dead. Therefore, we used the complement (1-x) of the disability weight from the GBD study to

make our calibrated VAS. Stouthard et al. describe how experts can then place (map) the remaining health conditions on the VAS scale. Our protocol was slightly different form the protocol of Stouthard, in the way that we did not make use of the EQ-5D to classify all health conditions at hand. The health conditions were one by one valued by the experts from the Value based Operation Room Triage team using the following procedure; first, the health condition was shortly introduced by an expert with the most clinical experience with this condition. The other experts were allowed to ask questions and discuss the QoL aspects of the condition. Subsequently, all experts wrote down their own QoL estimation of the health condition. Then, two to three other experts were invited to express their estimated QoL value for the health condition. Ultimately, the expert registered their own final values. In this way, the experts could use a maximum of information and opinions, but still express their own estimation. In addition, we could estimate the variance, the 95% confidence interval (95% CI), of the QoL values. The mean and 95% CI of the mapped QoL scores were used in the model. We used two session of three hours to collect QoL values. The preoperative and postoperative health state of 3 surgeries (one with a mild and severe subgroup) were estimated in both sessions, which effectively were 8 estimates of QoL. This allowed us to obtain an indication of the test-retest reliability (based on a t-test) of the valuations. For the model, the first estimates obtained in the first session were used. More details about the methods used to collect all QoL data can be found in Appendix C.

Time until no effect surgery

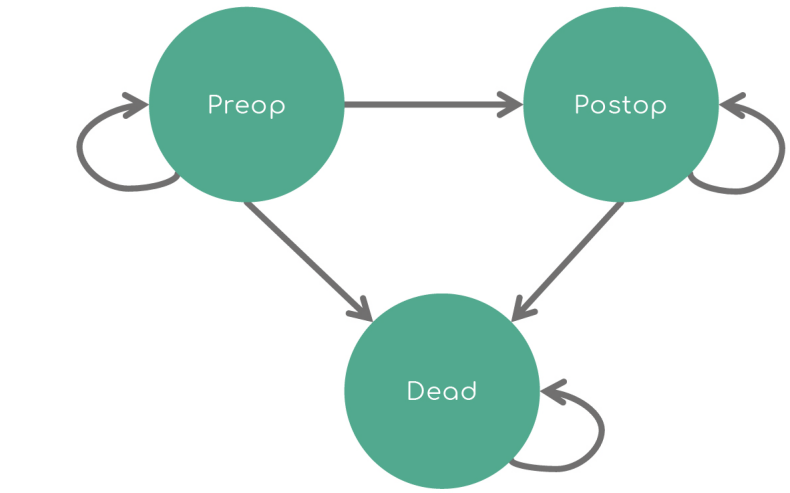
Since postponing surgery can have consequences on the effectiveness of the surgery, we included a model parameter that reflected the time until no effect can be expected of treatment on survival. In practice, this means that when this time has passed, we assumed that the surgery did no longer have an effect on the survival of the patient anymore. This time is often important in oncological surgeries, where after a specific time a tumor becomes inoperable or metastasizes. The effectiveness of non-oncological surgery could be time-dependent as well, for example repairing an abdominal aneurysm of the aorta. The data for this parameter was obtained from the scientific literature (Appendix A). For most surgeries, only data about the minimal delay not associated with worse survival could be obtained from the scientific literature. For those surgeries, we assumed the upper limit of this parameter to be a year (the maximum delay we evaluated), and the mean of the lower and upper limit as average. The same was done for the time until no effect can be expected on QoL.

Markov model

A three-state cohort state-transition model was developed. This model simulates a hypothetical cohort of patients over a defined period in fixed time intervals, called cycles, to estimate the average time individuals spend in the various health conditions, called health states.[22,27] Individuals could transition between a preoperative state, a postoperative state, and a dead state (Figure 1). Based on the time spent in these states, health benefits, like life years (LYs) or years lived with disability (YLD) are calculated.[22,28,29] Since health benefits now are enjoyed more than in the distant future, it is recommended to perform discounting.[30,31] A discount rate of 0.015 per year for health benefits was used, as this is common practice in the Netherlands.[31]

The entire cohort started in the preoperative state, and was followed their remaining lifespan, until they were 100 years old, using weekly cycles. The transition from the preoperative state to the postoperative state was set to a specific week, depending on the scenario. Scenarios of surgical delay of two weeks were compared with surgical delay up to a year using intervals of ten weeks. In addition, the scenario where patients never received treatment was evaluated: this was modelled by following patients their entire remaining lifespan in the preoperative health state. In all scenarios, the transitions from the pre- and postoperative states to the dead state were based on survival data. A description of the model assumptions can be found in Appendix C.

Figure 1



State-transition diagram of the cohort model. The model is a state-transition cohort model with 3 health states, a preoperative health states (preop), a postoperative state (postop), and dead. All patients start in the preop health state. This is the health state where patient eligible for surgery start in our simulation. We follow these patients over time using fixed time intervals of 1 week; these fixed time intervals are called cycles. Every cycle, patients can transition to one of the other health states or they can remain in the health states they currently are. From the preop health state they either die (transition to dead health state) or continue to wait for their surgical procedure (stay in the preop health state, the arrow points back into the health state). At the time of surgery, which is determined by the selected model scenario of surgical delay, all individuals still alive in the preop health state transition to the postop health state. The cohort is followed their remaining lifetime, defined as up to 100 years of age. While they are followed, they can die (transition from the postop state to dead state) or stay alive in the postop health state (transition back to the postop state). Finally, patients in the dead state remain dead, so every cycle they stay in the dead state.

Health effects of surgery

In order to be able to prioritize across disciplines, we chose to minimize the years of life lost (YLL) and the DALY. Priority is given to the patients where surgery delay is associated with more DALYs per unit of time. The YLL and DALYs as a result of delays in surgery were evaluated. YLL disregard QoL (QoL = 100%), while DALYs incorporate QoL and are therefore preferred. The expected health outcomes without surgery were compared to the expected health outcomes with surgery at 2 weeks to determine the overall health associated with not performing surgery. The expected health outcomes with surgery at 2 weeks were compared to the expected health outcomes at 52 weeks to determine the health lost per 50 weeks. This measure of urgency was converted to health lost per month delay and was used to rank the surgeries, where a high DALY/month or high YLL/month indicates an urgent surgery.

Analysis

Probabilistic sensitivity analysis (PSA) was used to incorporate parameter uncertainty in the model outcome (see Appendix A for the parameter distributions and Appendix C for more details about the used PSA method). Rankings based on health benefits or health loss per unit of time were compared using Spearman's rank correlation coefficient.

The model was built with R software[32] and adapted from previously published code.[33,34] The full model code is available on GitHub: https://github.com/bgravesteijn/Utilitarian-distribution-of-OR-capacity-during-COVID-19. The model results are described in the next section and can also be viewed in an online accessible tool where users can interactively select the surgeries of interest via the following link: https://tinyurl.com/y2yzudgw.

Table 1

n	Age 43	Quality of life: Preop 43	Quality of life: Postop 43	Survival: Preop 43	Survival: Postop 43	Time no eff QoL 6	Time no eff Survival 23	Treatment effect 22
Type of evidence (%)								
Before-after study	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (4.5)
Expert opinion	2 (4.7)	0 (0.0)	0 (0.0)	8 (18.6)	2 (4.7)	5 (83.3)	4 (17.4)	4 (18.2)
Expert panel*	0 (0.0)	29 (67.4)	29 (67.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
WHO GBD study	0 (0.0)	14 (32.6)	14 (32.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
National registry	21 (48.8)	0 (0.0)	0 (0.0)	12 (27.9)	31 (72.1)	0 (0.0)	9 (39.1)	6 (27.3)
Observational, Prospective	5 (11.6)	0 (0.0)	0 (0.0)	4 (9.3)	3 (7.0)	0 (0.0)	3 (13.0)	1 (4.5)
Observational, Retrospective	10 (23.3)	0 (0.0)	0 (0.0)	9 (20.9)	4 (9.3)	0 (0.0)	7 (30.4)	3 (13.6)
RCT	5 (11.6)	0 (0.0)	0 (0.0)	10 (23.3)	3 (7.0)	1 (16.7)	0 (0.0)	7 (31.8)
Class of evidence (%)								
I	5 (11.6)	0 (0.0)	0 (0.0)	10 (23.3)	3 (7.0)	1 (16.7)	0 (0.0)	7 (31.8)
IIa	5 (11.6)	0 (0.0)	0 (0.0)	4 (9.3)	3 (7.0)	0 (0.0)	3 (13.0)	2 (9.1)
IIb	31 (72.1)	43 (100.0)	43 (100.0)	21 (48.8)	35 (81.4)	0 (0.0)	16 (69.6)	9 (40.9)
III	2 (4.7)	0 (0.0)	0 (0.0)	8 (18.6)	2 (4.7)	5 (83.3)	4 (17.4)	4 (18.2)

Class and type of evidence underlying the model parameter inputs. Class definitions: I = Randomized Controlled Trials (RCT) or systematic reviews of RCTs, IIa = Prospective observational studies, before-after studies, IIb = Retrospective observational studies, expert panels for the utilities, national registries, class III = expert opinion. Time no eff: Time until no effect on QoL/Survival expected, QoL = Quality of Life, Preop = preoperative, Postop = Postoperative, WHO = World Health Organisation, GBD = Global burden of Disease.

**Expert panel refers to the Value Based Operation Room Triage team collaborators, see appendix C for details of this panel.*

Results

Data collection

In total, 12 cardiothoracic surgeries were evaluated, along with 23 oncological surgeries, 2 transplantations (liver and living donor kidney), 5 vascular surgeries, and 1 other type of surgery (creation of a shunt to facilitate hemodialysis). These 43 evaluated surgeries comprised of 69% of the total number of semi-elective surgeries in our hospital.

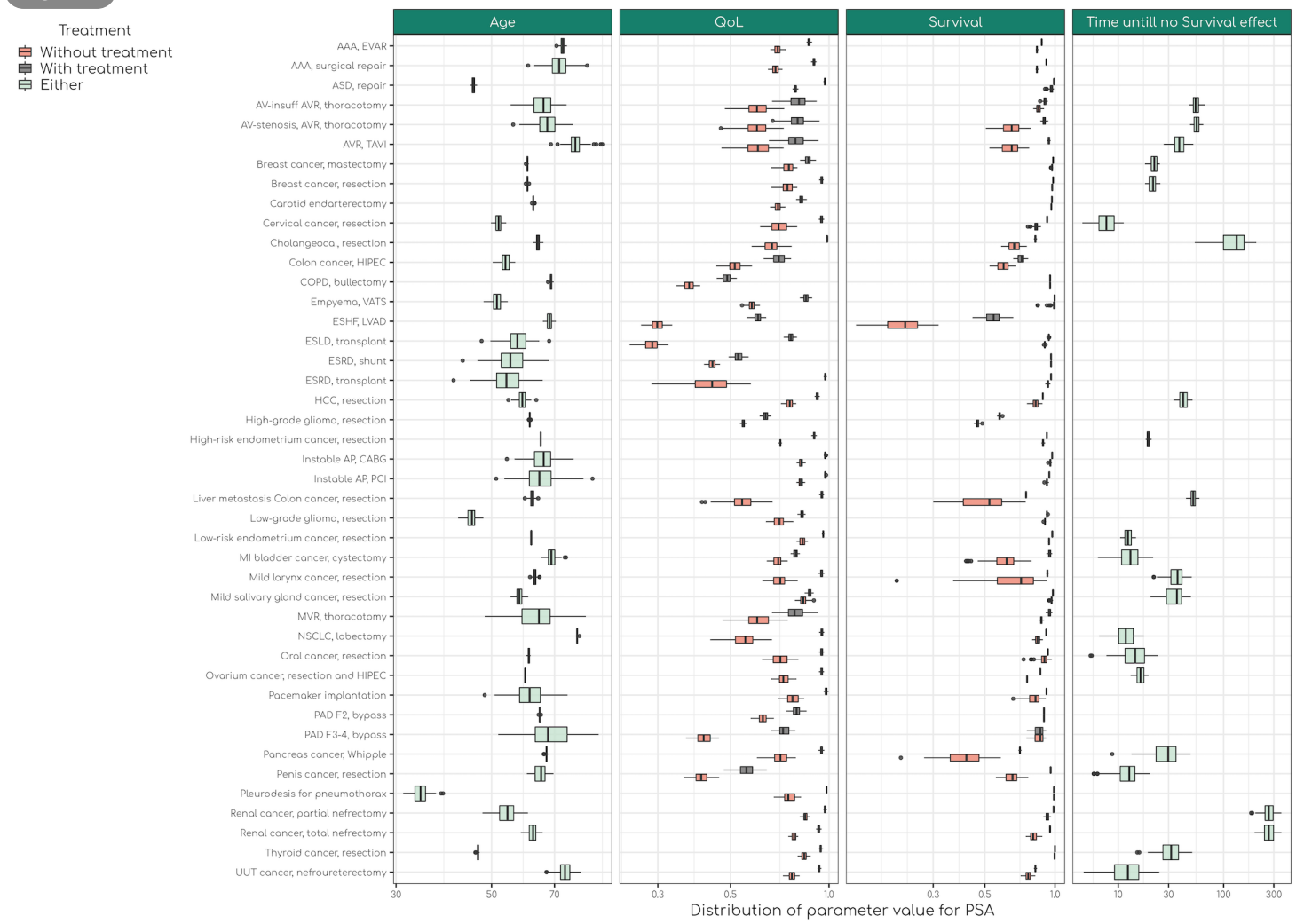
Survival with treatment was mostly based on national registries (31/43, Table 1). Survival without treatment was mostly based on data from (inter)national registries (12/43 surgeries, 6 calculated through the treatment effect), but also frequently from RCT's (10/43 surgeries, 7 calculated through the treatment effect), and observational studies (9/43 surgeries, 3 calculated through the treatment effect).

For 14/43 surgeries, QoL was available through the WHO Global Burden of Disease study.[35] For the remaining 29 surgeries, the QoL of the pre- and postoperative health state was estimated by the expert panel as described in the methods section. Test-retest validation analysis showed that the gain in QoL due to surgery was consistent in the two separate expert panel sessions (standardized mean difference 0.025, 95% CI: -0.11 – 0.16, Appendix C: Table 1 and Figure 1).

For 6/43 surgeries, a "time-to-no-effect-on-QoL" within one year, our maximum period of delaying surgery, was found in the literature. For 23 surgeries, a "time-to-no-effect-of-treatment-on-survival" was assumed based on qualitative assessment of the literature. Most of these surgeries were oncological surgeries (20/23). The estimates for the time until surgery becomes ineffective was mostly based on class IIb evidence (retrospective and prospective observational studies, Table 1).

Overall, input parameters varied widely between surgeries (Figure 2). Appendix A presents all input parameters, their sources[19,20,43–52,35,53–62,36,63–72,37,73–82,38,83–89,39–42], and the corresponding model output.

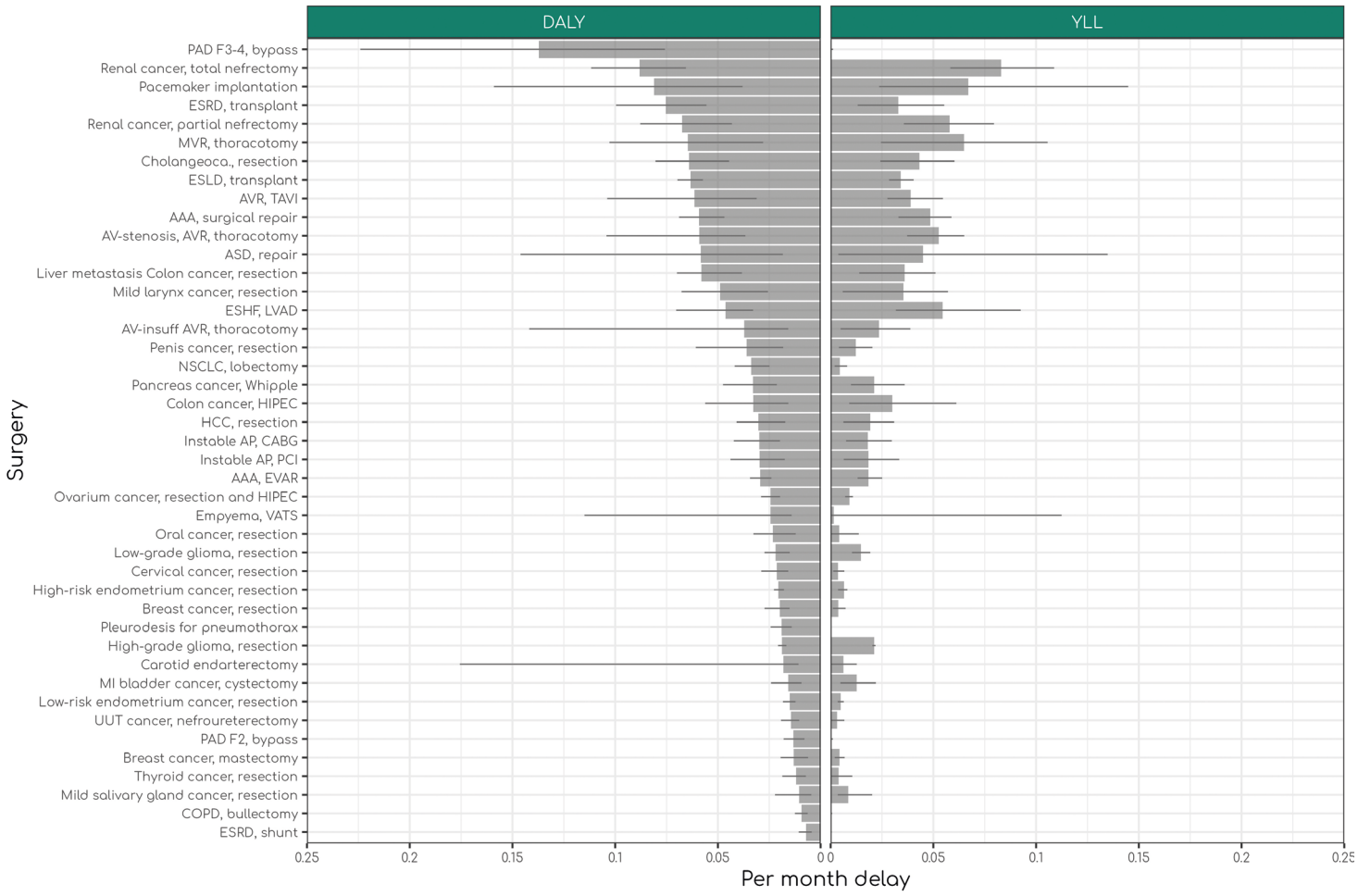
Figure 2



This figure shows the distribution of the parameter values as used during the probabilistic sensitivity analysis (PSA). For each PSA iteration (100 iterations in total), a value for each parameter was sampled from the original source input as described in Appendix A in Supplemental Materials. distribution of the final values used in the model is shown here. The y-axis shows the names of the surgical procedures. In the column called survival the x-axis represents the weekly probability of surviving. In the column Time until no Survival effect the x-as represents the days until treatment is not effective. For a full list of input parameters per disease and source, see Appendix A in Supplemental Materials.

Abbreviations Figure titles: *QoL_no_tx, Quality of Life without treatment; QoL_tx, quality of life with treatment; Surv_no_tx, 1-year survival probability without treatment; Surv_tx, 1-year survival probability with treatment; Time_noeff_surv, days until no treatment is effective.* ***Abbreviations surgery:*** *AAA indicates aneurysm of the abdominal aorta; AP, angina pectoris; ASD, atrial septum defect; AV, aortic valve; AVR, aortic valve replacement; -ca., cancer; CABG, coronary artery bypass graft; COPD, chronic obstructive pulmonary disease; ESHF, end-stage heart failure; ESLD, end-stage liver disease; ESRD, endstage renal disease; EVAR, endovascular aortic repair; HIPEC, hyperthermic intraperitoneal chemotherapy; HCC, hepatocellular cancer; LVAD, left ventricle assist device; MI, muscle invasive; MVR, mitral valve replacement; NSCLC, non-small cell lung cancer; PAD F2, peripheral arterial disease Fontaine classification 2; PAD F3-4, peripheral arterial disease Fontaine classification 3-4; PCI, percutaneous coronary intervention; TAVI, transaortic valve implantation; UUT, upper urinary track; VATS, video-assisted thoracoscopic surgery.*

Figure 3



The average DALYs and YLLs per month of delay for the investigated surgical procedures based on the simulation of surgery delay of 52 weeks. The estimates (gray bars) and 95% confidence intervals (black lines) are shown. The actual data are presented in Appendix B in Supplemental Materials.

Abbreviations Figure titles: *DALY, Disability adjusted life years; YLL, Years of life lost.* ***Abbreviations surgery:*** *AAA, aneurysm of the abdominal aorta; AP, angina pectoris; ASD, atrial septum defect; AV, aortic valve; AVR, aortic valve replacement; -ca., cancer; CABG, coronary artery bypass graft; COPD, chronic obstructive pulmonary disease; ESHF, end-stage heart failure; ESLD, end-stage liver disease; ESRD, end-stage renal disease; EVAR, endovascular aortic repair; HIPEC, hyperthermic intraperitoneal chemotherapy; HCC, hepatocellular carcinoma; LVAD, left ventricle assist device; MI, muscle invasive; MVR, mitral valve replacement; NSCLC, non-small cell lung carcinoma; PAD F2, peripheral arterial disease Fontaine classification 2; PAD F3-4, peripheral arterial disease Fontaine classification 3-4; PCI, percutaneous coronary intervention; TAVI, transaortic valve implantation; UUT, upper urinary track; VATS, video-assisted thoracoscopic surgery.*

Urgency

The delay of most surgeries resulted in a linear increase in DALYs per delay, except surgeries where a time until no effect of treatment on survival was included in the model (Figure 1, Appendix B).

The DALYs associated with delay of the surgeries ranged from 0.01 DALY/month (95% CI: 0.00-0.01) for placing a shunt for dialysis, to 0.14 DALY/month (0.08-0.22) for a bypass surgery for Fontaine III/IV peripheral arterial disease (Figure 3, and Appendix B Table 1). If the latter would be postponed by a month, patients lose approximately 51 days (0.14×365) spent in perfect health.

After bypass surgery for Fontaine III/IV peripheral arterial disease, the surgery associated with most DALYs in case of delay was total nefrectomy for renal cancer (0.09 DALY/month, 95% CI: 0.07-0.11). Following placing a shunt for patients with end-stage renal disease, the surgery associated with the least DALYs was bullectomy for COPD (0.01 DALY/month, 95% CI: 0.01-0.01). Surgeries that were associated with a higher expected DALY if not performed, were also associated with more DALYs per month delay: The Spearman correlation coefficient between the ranking of health benefit, in terms of DALYs, and urgency, in terms of DALY/month, was 0.31 (p=0.047).

The DALYs were strongly correlated with the YLLs by not performing surgery: The Spearman rank correlation coefficient between the ranking of surgeries based on YLL and DALY/month was 0.69 ($p< 0.001$).

Discussion

Our proposed decision model is an attempt to systematically guide prioritization of surgeries from a utilitarian perspective. The decision model provides the expected health loss due to surgery delay, which can be interpreted as a measure of urgency. Our approach operationalizes ethical values that are the most appropriate in times of scarcity.[2] Available evidence suggests that semi-elective surgeries can be ranked based on their urgency using a simple decision model. For survival after surgery, most evidence was based on national registries, while treatment effects were mostly derived from RCTs. The time until no effect of treatment on survival or QoL was most often derived from retrospective observational studies and expert opinion, respectively class IIb/III evidence.

Among the 43 surgeries analyzed, bypass surgery for Fontaine III/IV peripheral arterial disease and total nefrectomy for renal cancer appeared to be the most urgent surgeries, since delay was associated with most DALYs. Less urgent surgeries were placement of a shunt for dialysis and bullectomy for COPD.

Interestingly, the ranking of health loss due to delay is primarily driven by the YLL associated by not performing surgery: Surgeries that are associated with substantial YLL when not performed (e.g., mitral valve replacement), also result in more DALY per month delay than surgeries that are associated with no YLL when not performed (e.g., creation of a shunt for hemodialysis). The larger the total health benefit associated with surgery, the more health can potentially be lost by postponing surgery.

To make optimal use of operation room (OR) capacity, our metric for urgency could be used to fill OR schedules. Hospital capacity, however, is a dynamic multidimensional concept, which includes for example staff, number of beds, number of operation theatres, and medical equipment. All these factors need all be sufficiently present, and bottlenecks in one of these factors can vary from week to week. Therefore, future research aims to model the dynamics of capacity, and use the predictions to optimally distribute capacity over most urgent procedures.

Although our modeling approach rationalizes and objectively quantifies urgency from a utilitarian perspective, it needs to be complemented by other perspectives to be used effectively in practice. First, a financial perspective might also be explored. This perspective might be less relevant in a crisis such as the COVID-19 pandemic, where the bottleneck mainly seems hospital capacity instead of costs. Moreover, in a high-income country such as the Netherlands, the bottleneck costs might be less relevant than in a low- or middle-income country. Even in our setting, the fairness of directing all resources to care for COVID-19 patients is discussed as the crisis evolves. If this approach would be applied to the context of regular care, or in low- and middle-income countries, this perspective might be of increasing importance. A more traditional cost-effectiveness approach can be used to provide guidance on decision making in these settings. Second, other perspectives include the availability of alternative treatment strategies, e.g., in cancer treatment, (chemo-)radiation or systematic therapy alone instead of surgery, and other ethical standpoints, e.g., rule of rescue.[13] By exploring all these perspectives, it can be established whether our approach is applicable to all surgical procedures.

There are practical advantages of comparing “average patients” on urgency, despite the fact that there is no such thing as an “average patient”: It prevents our approach from systematically discriminating against a specific group of patients. Our approach would only discriminate if specific socioeconomic groups would suffer more frequently from diseases that are less urgent. It is known that lower socioeconomic groups are more prone to develop diseases that have clear association with unhealthy behavior, such as lung cancer.[90] However, these diseases do not systematically rank low in our approach. Comparing the average patients across specialties on urgency may seem not seem to be a personalized approach, but it can be tailored to an individual’s context by providing input for shared decision making. We believe that next to a quantitative estimation of urgency from a utilitarian perspective, individual patient’s preferences, social contexts, and operability should also be included in the decision-making process.

Since all models are a simplification of reality, our model has several limitations. First, the survival data used were not all derived from high-quality evidence. Although survival with treatment might be validly estimated from national registries, the survival without treatment is harder to be unbiasedly estimated. The surgeries that were evaluated are often part of standard clinical practice. Therefore, data might be biased (e.g., selection bias in the survival without treatment because patients opt for palliative care), or not available (it would be unethical now to perform RCTs evaluating surgery versus no surgery). Instead, best available evidence was used, which in part included evidence from more historical RCTs. As such, data might be biased, and as a result, so might the estimates from our model be. Because of this limitation, our approach is to aggregate transparently and systematically the best currently available evidence using a model. However, we are convinced we used best available evidence.

Second, no extra harm due to surgery was assumed. The current model does not simulate adverse events, like major bleedings or death due to surgery. It can be argued that the estimates from comparison studies incorporate these harms of surgery, therefore the impact of this limitation on survival might be irrelevant. However, the potential reduction of QoL due to these adverse events was not incorporated, nor the QoL reduction of a temporary period of recovery after surgery. Because of these assumptions, the overall DALYs associated with not performing surgery should not be interpreted as an absolute estimate. They are the maximum possible DALYs that can result from not performing the surgery. However, these assumptions were considered reasonable to achieve the main goal of this study: when surgery without delay is compared to surgery with delay, the harm in both scenarios is similar and therefore cancel out.

Third, because the health loss in 50 weeks was converted to loss per month, a linear approximation was effectively used to quantify urgency by delaying surgery up to a year. However, some surgeries did show a slightly curved trend in the period up to 32 weeks delay (Appendix B). The data needed to validly model this decay in DALYs per

unit of time for all surgeries likely does not exist: most of the estimates of time to no effect on survival were based on observational studies, which are likely biased. A more detailed approximation would be possible using a more individualized model which also models the natural growth of tumors, or aneurysms, and validly models the development of metastasis. It was not feasible to develop this for all evaluated surgeries. Instead, we opted for a more pragmatic approach.

Fourth, QoL weights were derived from expert-opinion. In this approach the patient is not involved. Instead, experts interpret the health states and give weights, thereby our approach takes a societal perspective. Besides being a relevant perspective, an advantage of our approach is that it is a more distanced evaluation of health. Patient reported evaluation of health might be less relevant in the prioritization of care due to distortion by coping mechanisms. There are also multiple methodological, ethical, and contextual disadvantages of using DALYs, but it should be noted that most of those discussions are more about utilitarian principles.[91]

Fifth, the potential impact on QoL of delaying a semi-elective surgery was not included. This impact might differ across surgeries. Whereas literature on waiting lists for transplants indicates that especially the physical functioning part of QoL declines over time[92,93], longer waiting time for elective procedures such as repairing an inguinal hernia mostly impacts the emotional wellbeing part of QoL.[94] Moreover, it might be hypothesized that surgeries performed after already a long disease history (e.g., kidney transplant) might have less "waiting time disutility" than recently diagnosed diseases (e.g., breast cancer).

Part of the input parameters were based on national registry data, but a substantial amount of the input originated from various international sources. Therefore, with some modifications, the model can easily be adapted to different contexts. Therefore, this study can be considered the first step towards a triaging strategy which optimizes surgical benefit in times of scarcity in surgical capacity, such as during the COVID-19 pandemic. To improve validity, it is however essential to periodically review the literature and update the model with higher quality evidence, much like a living systematic review.[95] If accepted, a wider range of surgeries should be considered, implementation strategies should be explored and evaluated, and the model should be applied to a variety of settings.

Conclusion

By transparently aggregating the best available evidence, our decision model supports prioritization of surgical care in times of scarce surgical capacity (e.g., during pandemics) from a utilitarian perspective. Our approach quantifies the expected health loss due to delay for semi-elective surgeries often performed in an academic hospital in the Netherlands. This approach can help to minimize health losses when trying to overcome delay in surgeries across disciplines. This approach is more transparent, more evidence-based, and more consistent than the alternative strategy of triaging based on expert opinion.

It should be noted that evidence from well-controlled comparison studies is often lacking. Instead, adjusted estimates from observational studies are often the best available evidence for benefit of surgery and the effects of delay on survival. Therefore, model inputs should be periodically updated with newer, higher quality evidence.

Finally, our approach should be placed in the context of other ethical perspectives and combined with capacity management tools. If accepted, we believe this modeling strategy should be implemented on a large scale, in order to minimize health loss of the accumulating group of patients awaiting surgery.

Supplemental material

Appendix A, B and C of this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.12.010.

Data and model code are provided in the GitHub repository (https://github.com/bgravesteijn/Utilitarian-distribution-of-OR-capacity-during-COVID-19) and the work also comes with online tool where users can interactively select the surgical procedures of interest via the following link: https://tinyurl.com/y2yzudgw.

Acknowledgements

We are grateful for Lisa Caulley for her suggestions to revise the final manuscript. We are grateful for H. Karreman and C. Van der Velden-van der Graaf for the work they have done for the quality of life data collection. Moreover, we want to thank Ruben Goedhart, Esther van Spronsen and Linda van der Sluijs – van der Beek for extracting the data from the electronic patient registry. Finally, we also want to thank the anonymous reviewers of Value in Health for their valuable suggestions.

References

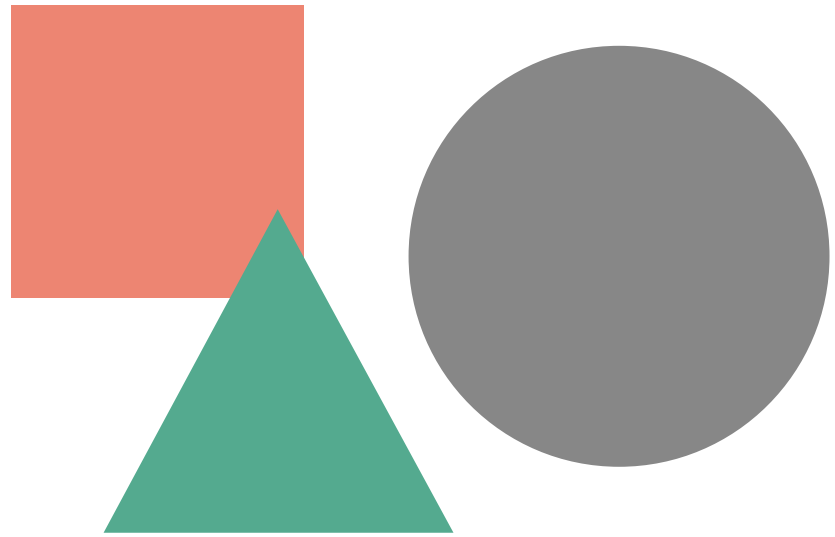
1. Office of the Assistant Secretary for Preparedness H. *Pandemic Influenza Plan - Update IV* (December 2017). 2017.
2. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, Zhang C, Boyle C, Smith M, Phillips JP. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *N Engl J Med*. 2020; 382:1–7.
3. D'Agostino A, Demartini B, Cavallotti S, Gambini O. Mental health services in Italy during the COVID-19 outbreak. *The Lancet Psychiatry* 2020; 7(5):385–387.
4. Lazzerini M, Barbi E, Apicella A, Marchetti F, Cardinale F, Trobia G. Delayed access or provision of care in Italy resulting from fear of COVID-19. *The Lancet Child and Adolescent Health* 2020; 4(5):e10–e11.
5. Harahsheh AS, Dahdah N, Newburger JW, Portman MA, Piram M, Tulloh R, Mccrindle BW, De Ferranti SD, Cimaz R, Truong DT, Burns JC. Missed or Delayed Diagnosis of Kawasaki Disease During the 2019 Novel Coronavirus Disease (COVID-19) Pandemic. *J Pediatr Pandemic J Pediatr*. Epub ahead of print 2020. DOI: 10.1016/j.jpeds.2020.04.052.
6. Chang H-J, Huang N, Lee C-H, Hsu Y-J, Hsieh C-J, Chou Y-J. The Impact of the SARS Epidemic on the Utilization of Medical Services: SARS and the Fear of SARS. *Am J Public Health*. 2004; 94(4):562–564.
7. NZA. *Analyse van de gevolgen van de coronacrisis voor de reguliere zorg,* https://zorgdomein.com/media/documents/NZa-analyse_van_de_gevolgen_van_de_coronacrisis_voor_de_reguliere_zorg_-....pdf (2020, accessed May 17, 2020).
8. Sud A, Jones ME, Broggio J, Loveday C, Torr B, Garrett A, Nicol DL, Jhanji S, Boyce SA, Gronthoud F, Ward P, Handy JM, Yousaf N, Larkin J, Suh Y-E, Scott S, Pharoah PDP, Swanton C, Abbosh C, et al. Collateral damage: the impact on outcomes from cancer surgery of the COVID-19 pandemic. *Ann Oncol*. 2020; 31(8):1065–1074.
9. Powell SN, Mullen T, Young L, Morgan C, Heald D, Powell ET. Experiences from the SARS-CoV-2 Pandemic. *J Bone Jt Surg*. 2020; 102(13):1123–1125.
10. Salenger R, Etchill EW, Ad N, Matthew T, Alejo D, Whitman G, Lawton JS, Lau CL, Gammie CF, Gammie JS. The Surge after the Surge: Cardiac Surgery post-COVID-19. *Ann Thorac Surg*. Epub ahead of print May 3, 2020. DOI: 10.1016/j.athoracsur.2020.04.018.
11. Qadan M, Hong TS, Tanabe KK, Ryan DP, Lillemoe KD. A Multidisciplinary Team Approach for Triage of Elective Cancer Surgery at the Massachusetts General Hospital During the Novel Coronavirus COVID-19 Outbreak. *Ann Surg*. 2020; 272:1.
12. MacCormick AD, Parry BR. Judgment analysis of surgeons' prioritization of patients for elective general surgery. *Med Decis Making*. 2006; 26(3):255–64.
13. Garner RT, Rosen B. *Moral Philosophy: A Systematic Introduction to Normative Ethics and Meta-Ethics*. New York: Macmillan, 1967.

14. Vergano M, Bertolini G, Giannini A, Gristina G, Livigni S, Mistraletti G, Petrini F. *Clinical Ethics Recommendations for the Allocation of Intensive Care Treatments in exceptional, resource-limited circumstances,* http://www.siaarti.it/SiteAssets/News/COVID19 - documenti SIAARTI/SIAARTI - Covid-19 - Clinical Ethics Reccomendations.pdf (2020, accessed May 17, 2020).
15. Daugherty Biddison L, Berkowitz KA, Courtney B, De Jong MJ, Devereaux A V., Kissoon N, Roxland BE, Sprung CL, Dichter JR, Christian MD, Powell T. Ethical considerations: Care of the critically ill and injured during pandemics and disasters: CHEST consensus statement. Chest. 2014; 146(4 Suppl):e145S-e155S.
16. Bayer R. *Ethical Considerations for Decision Making Regarding Allocation of Mechanical Ventilators during a Severe Influenza Pandemic or Other Public Health Emergency*. 2011.
17. York State Department of Health N. *VENTILATOR ALLOCATION GUIDELINES New York State Task Force on Life and the Law New York State Department of Health*. 2015.
18. Toner E, Waldhorn R. Responding to pandemic influenza - The ethical framework for policy and planning | Information | Health Service Journal, https://www.hsj.co.uk/swine-flu/responding-to-pandemic-influenza-the-ethical-framework-for-policy-and-planning/5005219.article (2020, accessed May 17, 2020).
19. Kankersoorten - IKNL, https://iknl.nl/kankersoorten (accessed May 19, 2020).
20. NHR, https://nederlandsehartregistratie.nl/ (accessed May 19, 2020).
21. CBS. Sterftekansen naar leeftijd, geslacht, opleidingsniveau, https://www.cbs.nl/nl-nl/maatwerk/2017/23/sterftekansen-naar-leeftijd-geslacht-opleidingsniveau (accessed May 19, 2020).
22. Hunink M, Mc E, Glasziou P, Elstein A. *Decision Making in Health and Medicine: Integrating Evidence and Values*. 2nd ed. Cambridge: Cambridge University Press, http://www.cambridge.org (2003, accessed May 19, 2020).
23. Global Burden of Disease Collaborative Network. *Global Burden of Disease Study 2016 (GBD 2016) Disability Weights*. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
24. *General Guidance for DALYs calculation*, https://montagu.vaccineimpact.org/contribution/resources/c978e5a1acf6a502679c92200e78ef61.pdf (accessed May 19, 2020).
25. Disability-adjusted life years (DALYs), https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158 (accessed May 19, 2020).
26. Stouthard EA, Essink-Bot M-L, Bonsel GJ. Disability weights for diseases A modified protocol and results for a Western European region. *Eur J Public Health*. 2000; 10(1):24–30.
27. 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.
28. Klarman H, Rosenthal GD. Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. Med Care. 1968; 6.1:48–54.
29. Sonnenberg FA, Beck JR. Markov Models in Medical Decision Making. *Med Decis Making*. 1993; 13(4):322–338.
30. Torgerson DJ, Raftery J. Economic notes. Discounting. *BMJ*. 1999; 319(7214):914–5.
31. Zorginstituut Nederland. *Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg*. 2016.
32. R Core Team. R: A language and Environment for Statistical Computing, https://www.r-project.org (2013).
33. Alarid-Escudero F, Krijkamp EM, Enns EA, Hunink MGM, Pechlivanoglou P, Jalal H. Cohort state-transition models in R: From conceptualization to implementation, http://arxiv.org/abs/2001.07824 (2020, accessed May 19, 2020).
34. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao SYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics*. 2019; 37(11):1329–1339.
35. Salomon JA, Haagsma JA, Davis A, Maertens De Noordhout C, Polinder S, Havelaar AH, Cassini A, Devleesschauwer B, Kretzschmar M, Speybroeck N, Murray CJL, Vos T. *Disability weights for the Global Burden of Disease 2013 study*, www.thelancet.com/lancetgh (2015, accessed May 14, 2020).
36. Wang J, Yan C, Fu A. A randomized clinical trial of comprehensive education and care program compared to basic care for reducing anxiety and depression and improving quality of life and survival in patients with hepatocellular carcinoma who underwent surgery. *Medicine (Baltimore)*. 2019; 98(44):e17552.
37. Verwaal VJ, Bruin S, Boot H, Van Slooten G, Van Tinteren H. 8-Year follow-up of randomized trial: Cytoreduction and hyperthermic intraperitoneal chemotherapy versus systemic chemotherapy in patients with peritoneal carcinomatosis of colorectal cancer. *Ann Surg Oncol*. 2008; 15(9):2426–2432.
38. Konstantinides S, Geibel A, Olschewski M, Görnandt L, Roskamm H, Spillner G, Just H, Kasper W. A comparison of surgical and medical therapy for atrial septal defect in adults. *N Engl J Med*. 1995; 333(8):469–473.
39. Fein DA, Mendenhall WM, Parsons JT, McCarty PJ, Stringer SP, Million RR, Cassisi NJ. Carcinoma of the oral tongue: A comparison of results and complications of treatment with radiotherapy and/or surgery. *Head Neck*. 1994; 16(4):358–365.
40. Jakola AS, Myrmel KS, Kloster R, Torp SH, Lindal S, Unsgård G, Solheim O. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *JAMA - J Am Med Assoc*. 2012; 308(18):1881–1888.

41. Haruna A, Muro S, Nakano Y, Ohara T, Hoshino Y, Ogawa E, Hirai T, Niimi A, Nishimura K, Chin K, Mishima M. CT scan findings of emphysema predict mortality in COPD. *Chest*. 2010; 138(3):635–640.
42. Ruys AT, Heuts SG, Rauws EA, Busch ORC, Gouma DJ, Van Gulik TM. Delay in surgical treatment of patients with hilar cholangiocarcinoma: Does time impact outcomes? *HPB*. 2014; 16(5):469–474.
43. Shin DW, Cho J, Kim SY, Guallar E, Hwang SS, Cho B, Oh JH, Jung KW, Seo HG, Park JH. Delay to curative surgery greater than 12 weeks is associated with increased mortality in patients with colorectal and breast cancer but not lung or thyroid cancer. *Ann Surg Oncol*. 2013; 20(8):2468–2476.
44. 4Chen EY, Mayo SC, Sutton T, Kearney MR, Kardosh A, Vaccaro GM, Billingsley KG, Lopez CD. Effect of Time to Surgery of Colorectal Liver Metastases on Survival. *J Gastrointest Cancer.*; Epub. Epub ahead of print 2020. DOI: 10.1007/s12029-020-00372-5.
45. Yusuf S, Zucker D, Passamani E, Peduzzi P, Takaro T, Fisher LD, Kennedy JW, Davis K, Killip T, Norris R, Morris C, Mathur V, Varnauskas E, Chalmers TC. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet*. 1994; 344(8922):563–570.
46. Noorbakhsh A, Tang JA, Marcus LP, McCutcheon B, Gonda DD, Schallhorn CS, Talamini MA, Chang DC, Carter BS, Chen CC. Gross-total resection outcomes in an elderly population with glioblastoma: A SEER-based analysis. Clinical article. *J Neurosurg*. 2014; 120(1):31–39.
47. Nakano R, Ohira M, Kobayashi T, Ide K, Tahara H, Kuroda S, Shimizu S, Kimura T, Nagata Y, Aikata H, Chayama K, Ohdan H. Hepatectomy versus stereotactic body radiotherapy for primary early hepatocellular carcinoma: A propensity-matched analysis in a single institution. *Surg (United States)*. 2018; 164(2):219–226.
48. Lee JN, Kwon SY, Choi GS, Kim HT, Kim TH, Kwon TG, Kim BW. Impact of surgical wait time on oncologic outcomes in upper urinary tract urothelial carcinoma. *J Surg Oncol*. 2014; 110(4):468–475.
49. Lim C, Bhangui P, Salloum C, Gómez-Gavara C, Lahat E, Luciani A, Compagnon P, Calderaro J, Feray C, Azoulay D. Impact of time to surgery in the outcome of patients with liver resection for BCLC 0-A stage hepatocellular carcinoma. *J Hepatol*. 2018; 68(1):100–108.
50. Moss AJ, Jackson Hall W, Cannom DS, Daubert JP, Higgins SL, Klein H, Levine JH, Saksena S, Waldo AL, Wilber D, Brown MW, Heo M. Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *N Engl J Med*. 1996; 335(26):1933–1940.
51. Scott SWM, Batchelder AJ, Kirkbride D, Naylor AR, Thompson JP. Late Survival in Nonoperated Patients with Infrarenal Abdominal Aortic Aneurysm. *Eur J Vasc Endovasc Surg*. 2016; 52(4):444–449.
52. Nyboe C, Karunanithi Z, Nielsen-Kudsk JE, Hjortdal VE. Long-term mortality in patients with atrial septal defect: a nationwide cohort-study. DOI: 10.1093/eurheartj/ehx633.
53. Brewster DC, Jones JE, Chung TK, Lamuraglia GM, Kwolek CJ, Watkins MT, Hodgman TM, Cambria RP. Long-term outcomes after endovascular abdominal aortic aneurysm repair: The First Decade. *Annals of Surgery* 2006; 244(3):426–436.
54. Brunner M, Olschewski M, Geibeli A, Bode C, Zehender M. Long-term survival after pacemaker implantation: Prognostic importance of gender and baseline patient characteristics. *Eur Heart J*. 2004; 25(1):88–95.
55. Rose EA, Gelijns AC, Moskowitz AJ, Heitjan DF, Stevenson LW, Dembitsky W, Long JW, Ascheim DD, Tierney AR, Levitan RG, Watson JT, Ronan NS, Shapiro PA, Lazar RM, Miller LW, Gupta L, Desvigne-Nickens P, Meier P, Howard Frazier O, et al. Long-term use of a left ventricular assist device for end-stage heart failure. *N Engl J Med*. 2001; 345(20):1435–1443.
56. Mazzone E, Preisser F, Nazzani S, Tian Z, Fossati N, Gandaglia G, Gallina A, Soulieres D, Tilki D, Montorsi F, Shariat SF, Saad F, Briganti A, Karakiewicz PI. More Extensive Lymph Node Dissection Improves Survival Benefit of Radical Cystectomy in Metastatic Urothelial Carcinoma of the Bladder. *Clin Genitourin Cancer*. 2019; 17(2):105-113.e2.
57. Kann BH, Verma V, Stahl JM, Ross R, Dosoretz AP, Shafman TD, Gross CP, Park HS, Yu JB, Decker RH. Multi-institutional analysis of stereotactic body radiation therapy for operable early-stage non-small cell lung carcinoma. *Radiother Oncol*. 2019; 134:44–49.
58. Huang CE, Yang YH, Chen WC, Huang KT, Chen PC, Tsai YH, Lin WY. Nephroureterectomy increase 5 year survival in patients on dialysis with upper urinary tract urothelial carcinoma. *Oncotarget*. 2017; 8(45):79876–79883.
59. Shalowitz DI, Epstein AJ, Ko EM, Giuntoli RL. Non-surgical management of ovarian cancer: Prevalence and implications. *Gynecol Oncol*. 2016; 142(1):30–37.
60. Pedregal-Mallo D, Sánchez Canteli M, López F, Álvarez-Marcos C, Llorente JL, Rodrigo JP. Oncological and functional outcomes of transoral laser surgery for laryngeal carcinoma. *Eur Arch Oto-Rhino-Laryngology*. 2018; 275(8):2071–2077.
61. Kim WR, Lake JR, Smith JM, Schladt DP, Skeans MA, Harper AM, Wainright JL, Snyder JJ, Israni AK, Kasiske BL. OPTN/SRTR 2016 Annual Data Report: Liver. *Am J Transplant*. 2018; 18:172–253.
62. Muluk SC, Muluk VS, Kelley ME, Whittle JC, Tierney JA, Webster MW, Makaroun MS. Outcome events in patients with claudication: A 15-year study in 2777 patients. *J Vasc Surg*. 2001; 33(2):251–258.
63. Holtzman A, Morris CG, Amdur RJ, Dziegielewski PT, Boyce B, Mendenhall WM. Outcomes after primary or adjuvant radiotherapy for salivary gland carcinoma. *Acta Oncol (Madr)*. 2017; 56(3):484–489.

64. Murphy MM, Simons JP, Hill JS, McDade TP, Ng SC, Whalen GF, Shah SA, Harrison LH, Tseng JF. Pancreatic resection: A key component to reducing racial disparities in pancreatic adenocarcinoma. *Cancer*. 2009; 115(17):3979–3990.
65. Mikkola R, Kelahaara J, Heikkinen J, Lahtinen J, Biancari F. Poor late survival after surgical treatment of pleural empyema. *World J Surg*. 2010; 34(2):266–271.
66. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet*. 2003; 361(9351):13–20.
67. Piehler JM, Crichlow RW. Primary Carcinoma of the Gallbladder. *Arch Surg*. 1977; 112(1):26–30.
68. Warner L, Chudasama J, Kelly CG, Loughran S, Mckenzie K, Wight R, Dey P. Radiotherapy versus open surgery versus endolaryngeal surgery (with or without laser) for early laryngeal squamous cell cancer. *Cochrane Database of Systematic Reviews*; 2(CD002027). Epub ahead of print December 12, 2014. DOI: 10.1002/14651858.CD002027.pub2.
69. European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: Final results of the MRC European Carotid Surgery Trial (ECST). *Lancet*. 1998; 351(9113):1379–1387.
70. Soran A, Ozmen V, Ozbas S, Karanlik H, Muslumanoglu M, Igci A, Canturk Z, Utkan Z, Ozaslan C, Evrensel T, Uras C, Aksaz E, Soyder A, Ugurlu U, Col C, Cabioglu N, Bozkurt B, Uzunkoy A, Koksal N, et al. Randomized Trial Comparing Resection of Primary Tumor with No Surgery in Stage IV Breast Cancer at Presentation: Protocol MF07-01. *Ann Surg Oncol*. 2018; 25(11):3141–3149.
71. Ginsberg RJ, Rubinstein L V. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. *Ann Thorac Surg*. 1995; 60(3):615–623.
72. Redden MD, Chin TY, van Driel ML. Surgical versus non-surgical management for pleural empyema. *Cochrane Database of Systematic Reviews*; 2017(3). Epub ahead of print March 17, 2017. DOI: 10.1002/14651858.CD010651.pub2.
73. Sørensen VR, Heaf J, Wehberg S, Sørensen SS. Survival Benefit in Renal Transplantation Despite High Comorbidity. *Transplantation*. 2016; 100(10):2160–2167.
74. Shalowitz DI, Epstein AJ, Buckingham L, Ko EM, Giuntoli RL. Survival implications of time to surgical treatment of endometrial cancers. *Am J Obstet Gynecol*. 2017; 216(3):268.e1-268.e18.
75. van Harten M, de Ridder M, Hamming-Vrieze O, Smeele L, Balm A, van den Brekel M. The association of treatment delay and prognosis in head and neck squamous cell carcinoma (HNSCC) patients in a Dutch comprehensive cancer center. *Oral Oncol*. 2014; 50(4):282–290.
76. Stewart JM, Tone AA, Jiang H, Bernardini MQ, Ferguson S, Laframboise S, Murphy KJ, Rosen B, May T. The optimal time for surgery in women with serous ovarian cancer. *Can J Surg*. 2016; 59(4):223–232.
77. Davies L, Welch G. Thyroid cancer survival in the United States: Observational data from 1973 to 2005. *Arch Otolaryngol - Head Neck Surg*. 2010; 136(5):440–444.
78. Bleicher RJ, Ruth K, Sigurdson ER, Beck JR, Ross E, Wong YN, Patel SA, Boraas M, Chang EI, Topham NS, Egleston BL. Time to surgery and breast cancer survival in the United States. *JAMA Oncol*. 2016; 2(3):330–339.
79. Morse E, Fujiwara RJT, Judson B, Mehra S. Treatment Times in Salivary Gland Cancer: National Patterns and Association with Survival. *Otolaryngol - Head Neck Surg* (United States). 2018; 159(2):283–292.
80. USRDS, https://www.usrds.org/2015/view/ (accessed May 19, 2020).
81. Kirkegård J, Mortensen FV, Hansen CP, Mortensen MB, Sall M, Fristrup C. Waiting time to surgery and pancreatic cancer survival: A nationwide population-based cohort study. *Eur J Surg Oncol*. 2019; 45(10):1901–1905.
82. Chung JH, Lee SH, Kim KT, Jung JS, Son HS, Sun K. Optimal Timing of Thoracoscopic Drainage and Decortication for Empyema. Ann Thorac Surg. 2014; 97(1):224–229.
83. Organ Procurement and Transplantation Network, https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/ (accessed May 19, 2020).
84. Howard DPJ, Banerjee A, Fairhead JF, Hands L, Silver LE, Rothwell PM. Population-Based Study of Incidence, Risk Factors, Outcome, and Prognosis of Ischemic Peripheral Arterial Events: Implications for Prevention. *Circulation*. 2015; 132(19):1805–1815.
85. Bonow RO, Greenland P. Population-wide trends in aortic stenosis incidence and outcomes. *Circulation* 2015; 131(11):969–971.
86. Badalato GM, Gaya JM, Hruby G, Patel T, Kates M, Sadeghi N, Benson MC, McKiernan JM. Immediate radical cystectomy vs conservative management for high grade cT1 bladder cancer: Is there a survival difference? *BJU Int*. 2012; 110(10):1471–1477.
87. Lee CT, Madii R, Daignault S, Dunn RL, Zhang Y, Montie JE, Wood DP. Cystectomy delay more than 3 months from initial bladder cancer diagnosis results in decreased disease specific and overall survival. *J Urol*. 2006; 175(4):1262–1267.
88. Tan WS, Trinh QD, Hayn MH, Marchese M, Lipsitz SR, Nabi J, Kilbridge KL, Vale JA, Khoubehi B, Kibel AS, Sun M, Chang SL, Sammon JD. Delayed nephrectomy has comparable long-term overall survival to immediate nephrectomy for cT1a renal cell carcinoma: A population-based analysis. *Urol Oncol Semin Orig Investig*. 2020; 38(3):74.e13-74.e20.
89. Janssen MWW, Linxweiler J, Terwey S, Rugge S, Ohlmann CH, Becker F, Thomas C, Neisius A, Thüroff JW, Siemer S, Stöckle M, Roos FC. Survival outcomes in patients with large (7cm) clear cell renal cell carcinomas treated with nephron-sparing surgery versus radical nephrectomy: Results of a multicenter cohort with long-term follow-up. *PLoS One*.; 13(5). Epub ahead of print May 1, 2018. DOI: 10.1371/journal.pone.0196427.

90. Ae LXC, Reichman ME, Ae BAM, Hankey BF, Gopal AE, Singh K, Yi AE, Lin D, Marc AE, Goodman T, Charles AE, Lynch F, Stephen AE, Schwartz M, Chen VW, Leslie AE, Ae B, Gomez SL, Graff JJ, et al. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer Causes Control.* 2009; 20(4):417–435.
91. Pettitt D, Raza S, Naughton B, Roscoe A, Ramakrishnan A, Ali A, Davies B, Dopson S, Hollander G, Smith J, Brindley D. The Limitations of QALY: A Literature Review. *J Stem Cell Res Ther.*; 6(4). Epub ahead of print 2016. DOI: 10.4172/2157-7633.1000334.
92. Feltrim MIZ, Coelho AAC, Scatimburgo MM, Pereira GM, Pego-Fernandes P. Quality of Life Assessment in Two Consecutive Years of Patients in a Waiting List for Lung Transplantation. *Transplant Proc*. 2014; 46:3060–3063.
93. Auneau-Enjalbert L, Hardouin J-B, Blanchin M, Giral · Magali, Morelon E, Cassuto E, Meurette · Aurélie, Véronique Sébille ·. Comparison of longitudinal quality of life outcomes in preemptive and dialyzed patients on waiting list for kidney transplantation. *Qual Life Res*. 2020; 29:959–970.
94. Oudhoff JP, Timmermans D, Knol DL, Bijnen AB, Van Der Wal G. Waiting for elective general surgery: impact on health related quality of life and psychosocial consequences. Epub ahead of print 2007. DOI: 10.1186/1471-2458-7-164.
95. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, Mcdonald S, Salanti G, Meerpohl J, Maclehose H, Hilton J, Tovey D, Shemilt I, Thomas J. Living systematic review: 1. Introductiondthe why, what, when, and how on behalf of the Living Systematic Review Network. DOI: 10.1016/j.jclinepi.2017.08.010.



09

PRIORITISATION AND DESIGN OF CLINICAL TRIALS

Anna Heath
MG Myriam Hunink
Eline Krijkamp
Petros Pechlivanoglou

*All authors contributed equally to the methodology, drafting, and revising the paper. Therefore, they are named in alphabetic order.

European Journal of Epidemiology 2021

This chapter is written in UK spelling, while the rest of this thesis uses US spelling.

"Uncertainty is an uncomfortable position. But certainty an absurd one."

Voltaire

PRIORITISATION AND DESIGN OF CLINICAL TRIALS

Abstract

Clinical trials require participation of numerous patients, enormous research resources and substantial public funding. Time-consuming trials lead to delayed implementation of beneficial interventions and to reduced benefit to patients. This manuscript discusses two methods for the allocation of research resources and reviews a framework for prioritisation and design of clinical trials. The traditional error-driven approach of clinical trial design controls for type I and II errors. However, controlling for those statistical errors has limited relevance to policy makers. Therefore, this error-driven approach can be inefficient, waste research resources and lead to research with limited impact on daily practice. The novel value-driven approach assesses the currently available evidence and focuses on designing clinical trials that directly inform policy and treatment decisions. Estimating the net value of collecting further information, prior to undertaking a trial, informs a decision maker whether a clinical or health policy decision can be made with current information or if collection of extra evidence is justified. Additionally, estimating the net value of new information guides study design, data collection choices, and sample size estimation. The value-driven approach ensures the efficient use of research resources, reduces unnecessary burden to trial participants, and accelerates implementation of beneficial healthcare interventions.

Introduction

Unnecessary or poorly designed clinical trials waste research resources[1] and delay implementation of effective interventions. A well-conducted trial can also waste resources if the collected information is irrelevant to patients, physicians, or healthcare policy makers. Importantly, wasted research resources and implementation delays negatively affect patients' well-being and the efficiency of healthcare systems.[1] Therefore, before clinical trials are performed, we should assess and prioritise them based on their potential value and impact.[2,3]

In healthcare, research priorities can be set using several methods, e.g., using burden of disease or a qualitative assessment of the potential research impact.[4] These methods can identify some important research areas, but they do not use formal methodology to assess whether research is justified and how to design the best clinical trial. Therefore, they are unlikely to make the most efficient use of available resources.[4–6]

Furthermore, trials are usually designed to control the type I and type II errors when making conclusions about the primary outcome of the trial using a statistical hypothesis test.[7] These primary outcomes are usually selected using expert consensus[5], rather than assessing the relevance of that outcome to clinical and policy decision making. Furthermore, the trial sample size, a key element of trial efficiency[8], is computed to control error rates.[9] This can lead to large sample sizes that cause feasibility issues, require excessive time and money, put an unnecessary burden on patients, and delay implementation of effective interventions.[10,11] This paper defines this approach as the *error-driven* approach to clinical trial design.

Critiques of the *error-driven* approach highlight a risk of publishing misleading research findings[12] and a propensity to interpret research findings incorrectly.[12,13] These errors mean that research efforts and resources are wasted and clinical and policy decisions are misinformed.[13] Therefore, to improve research impact and the use of research resources, we must approach clinical trial prioritisation and design differently.

This paper describes an alternative approach to research prioritisation and trial design that evaluates the *value* that research can provide to clinical and policy decision makers. This approach can help research groups prioritise and design future clinical trials that make the best use of limited research resources and ensure that trial evidence supports decisions around the use and reimbursement of interventions.

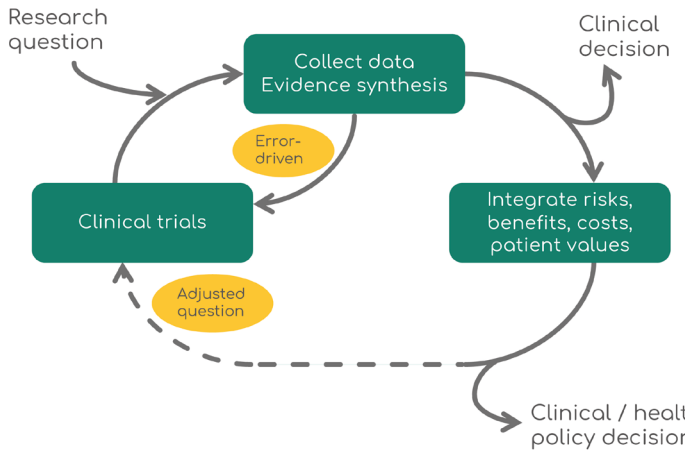
Iterative research cycles: errors versus value

Healthcare research is an iterative process, where treatment effect estimates are contested or confirmed in successive studies. Throughout this process, researchers rarely claim certainty about a particular result and usually call for more research. Findings from trials add to the current evidence but uncertainty around the effectiveness or efficiency of interventions is rarely eliminated. However, clinicians and policy makers must make decisions, even in the face of uncertainty. Thus, we must determine if the remaining uncertainty justifies further research and how to design research that reduces uncertainty efficiently and effectively. Research should be an iterative cycle of designing studies, analysing the collected evidence in the context of what is known already, refining the research questions and designing future studies until we can make justifiable decisions about improving clinical practice.

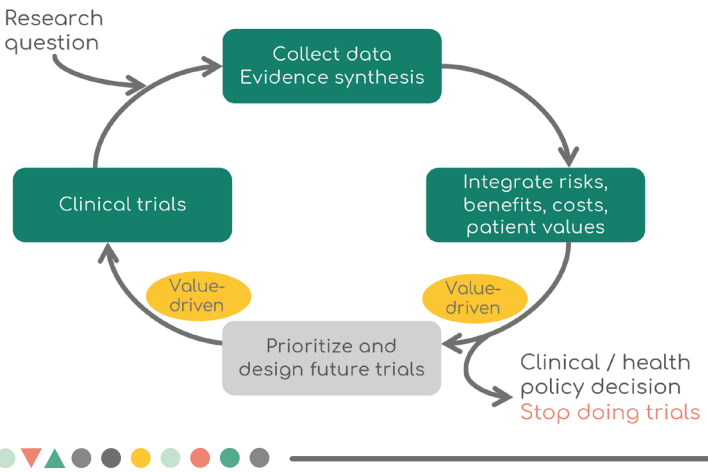
Error-driven approach

Currently, the majority of clinical trials are designed using the error-driven approach.[14] All research starts with a research question (Figure 1A). Using this research question, the design process proceeds with a systematic review of existing evidence from clinical trials, where feasible accompanied by a meta-analysis. Sometimes the results from the systematic review are combined with information on risks and benefits, patient values and costs to support clinical or health policy decision making.[15]

Figure 1



Iterative research cycles. (A): The current research cycle based on controlling type I and II errors. This classical method for developing and designing clinical trials is called the 'error-driven' approach. We consider that this approach has both a long and short iterative design process. The short route is in the top left-hand portion of the Figure and only iterates between the Evidence Synthesis and the Clinical trials boxes. The longer process includes all three key boxes while the dashed line represents the disconnect between how the information from the trials is used in policy making and the subsequent design of the next clinical trial.



(B): A novel iterative research cycle that is driven by determining the value of different research strategies and pursuing research with the maximum value. This approach is called the 'value-driven' approach. Here the connection between policy making and the next clinical trial is determined using 'value of information' methods that prioritise and guide the design of future trials.

These processes may be sufficient to guide decision making but lack of statistical significance for a treatment benefit in the synthesised evidence, coupled with evidence or an a-priori belief of a true benefit, often leads to a new clinical trial.

In the *error-driven* approach, trials are commonly designed by selecting a key outcome, known as the primary outcome, and fixing the trial sample size so that a statistically significant difference can be seen for this outcome when the trial results are analysed in isolation. Following this isolated analysis, the trial data may be added to the previous systematic review and meta-analysis and the research cycle starts anew.

In general, the primary outcome is selected using expert consensus[5], often considering feasibility, e.g., progression free survival is used in oncology trials to allow for shorter follow-up times. The sample size calculation is often based on currently available evidence about the baseline behavior of the primary outcome and expert specification of the "minimally important clinical difference".[16] Using these two values, the sample size is set to control the type I error typically below a nominal level of 5% and the type II error below 10 or 20%.[9] Thus, the error rates, clinical judgement and an informal incorporation of existing information of treatment benefit are the key drivers of clinical trial design in the *error-driven* approach.

Box 1

Functions of "net value" used in health decision sciences and health technology assessment:

$$\text{Net Health Benefit (NHB)} = HB - \frac{1}{WTP} * C$$

$$\text{Net Monetary Benefit (NMB)} = HB * WTP - C$$

Where:

- HB is the health benefit, ideally integrating life expectancy and quality of life (e.g., QALYs)
- C is the total costs, including healthcare and non-healthcare costs
- WTP: society's willingness-to-pay in monetary units for one unit of health

Example: A treatment has an estimated health benefit of 12 QALYs (HB = 12), a cost of $200,000 (C = 200,000) and society's willingness-to-pay is set to $50,000/QALY (WTP = 50,000). Then:

$$\text{NHB} = 12 - \frac{1}{\$50,000} * \$200,000 = 8 \text{ QALYs}$$

$$\text{NMB} = 12 * \$50,000 - \$200,000 = \$400,000$$

Value-driven approach

The *value-driven* approach asserts that research requires substantial investment of time, resources and money and should only be undertaken when it generates value. The value of research can be calculated using two key concepts: an estimate of the value of healthcare interventions and a suite of methods known as Value of Information (VOI) methods.[17] (For an explanation of the abbreviations used see Table 1).

Table 1

Table of abbreviations and definition used throughout the manuscript in alphabetic order and the associated units of measurement commonly used.

Abbreviation	Full name	Units of measurement commonly used*
ENBS	Expected net benefit of sampling	Monetary units
EVPI	Expected value of perfect information	Monetary units
EVPPI	Population expected value of partial perfect information	Monetary units
EVSI	Expected value of sample information	Monetary units
Forgone benefit	Foregone benefit, or potential lost value, refers to the benefit that could have been gained if a more "optimal" decision had been made	Monetary units
HB	Health benefit	Health units, e.g., life years or QALYS
NBM	Net monetary benefit	Monetary units
NHB	Net health benefit	Health units, e.g., life years or QALYS
popEVPI	Population expected value of perfect information	Monetary units
popEVSI	Population Expected value of sample information	Monetary units
PSA	Probabilistic sensitivity analysis	Not applicable
QALY	Quality adjusted life years	Is an unit of measurement that combines quality and the quantify of life years
RCT	Randomized controlled trial	Not applicable
VOI	Value of information	Not applicable
WTP	Willingness-to-pay	Monetary units per unit of health

**All value of information outcomes can alternatively be expressed in health units but this is less commonly done because it makes comparison with the costs of research more complicated.*

VOI methods are applicable irrespective of the method used to value the healthcare intervention. Nonetheless, we usually take a health policy making perspective and value interventions using a composite of health benefit and costs.[17] Health benefit is either measured using "hard" outcomes, such as the number of life years saved, or, more commonly, by combining quantity and quality of life into a measure known as quality-adjusted life years (QALYs).[18,19] Costs can include directly related healthcare costs alongside wider societal costs, such as productivity or leisure time loss [18]. Health benefits and costs are then combined into one of two composite outcomes: the net monetary benefit or net health benefit. Net health benefit (NHB) measures the number of health units saved by the interventions, while the net monetary benefit (NMB) is evaluated in monetary units, e.g., $.[20] Both measures require an estimate of society's willingness to pay (WTP) for one unit of health[21], which can be thought of as an "exchange rate" between health benefits and costs (Box 1).

Once the value of each intervention has been calculated, we can find the best intervention by considering which has the highest potential value. However, the available evidence on benefits and costs is uncertain, resulting in uncertainty about the intervention that maximizes value. Thus, VOI methods estimate the value of future research as the chance of making the wrong decision about the best intervention with the current level of evidence, multiplied by the benefit of changing the decision in settings where we would be wrong. Thus, the *value-driven* approach is based on understanding that a decision about the best intervention *must* be made following a trial and controlling the consequences and probability of incorrect decision making.

Figure 1B presents the *value-driven* approach, starting with a research question, data collection and evidence synthesis. The *value-driven* approach then estimates the value of each intervention using information and data on health outcomes and costs. This process uses methods from statistics, health economics and decision science to characterise the impact of uncertainty on the estimates of value. The current best intervention for use in clinical practice is the intervention that is expected to have the highest value.[22] VOI methods then formally assess whether the current evidence is sufficient to determine the best intervention.[23]

To achieve this, we estimate whether the cost of undertaking additional research exceeds the value of the research (bottom Figure 1B).[3] We can also compute the value of alternative trial designs to prioritise the trial protocol with the greatest net value.[24] The *value-driven* approach then assumes that evidence collected in the trial will be analysed and interpreted alongside the current evidence to improve decision making following the trial. The value of each intervention can be estimated using the updated evidence and VOI methods can determine if further research is required. Thus, the *value-driven* approach is a full iterative research process.

Steps of the value-driven approach

The steps of the *value-driven* approach are summarised in Figure 2 and Table 2. While these steps may seem cumbersome, recent methodological advances and software can facilitate the process.[25–27] We provide a clarifying example of the *value-driven* approach in Box 2.

Firstly, the *value-driven* approach determines the clinical or public health decision making problem that is relevant to the research question. Next, we summarise the available evidence using systematic reviews and meta-analyses[28], integrate evidence on benefits and costs using a decision model[15] and calculate the *expected value* of each intervention.

Following this, we use distributions around the input parameters to model uncertainty in the current evidence and propagate this uncertainty through the decision model using "probabilistic sensitivity analysis" (PSA).[15] PSA determines the effect of input parameter uncertainty on the expected NMB or NHB. Using the PSA results, VOI methods can determine the chance and the consequences of making the wrong decision about the best treatment, i.e., the value of future

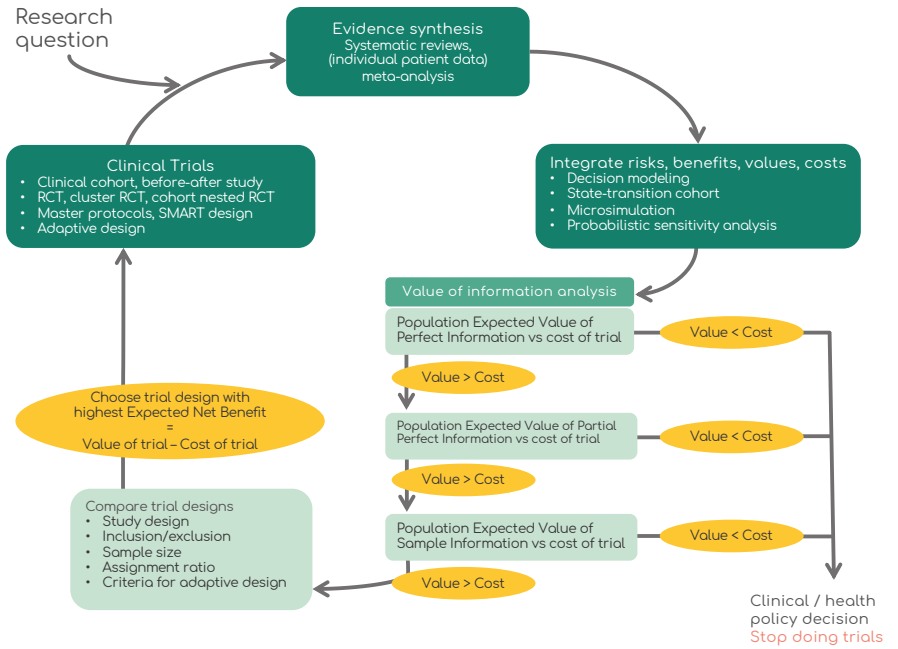
research (Figure 2 and Table 2). Strictly speaking, VOI is the "expected cost of the uncertainty" where "cost" is expressed in foregone health benefit or monetary units. Foregone benefit, or potential lost value, refers to the benefit that could have been gained if a more "optimal" decision had been made.

A VOI analysis begins by calculating the value of eliminating all sources of parameter uncertainty, known as the expected value of perfect information (EVPI). EVPI is the upper limit on the value that can be generated from a future study collecting evidence about the model parameters. If the EVPI is low, then no future research should be proposed.[29]

As a clinical trial is unlikely to estimate all parameters that are relevant to the decision, a VOI analysis proceeds by considering which outcomes should be included in the future study by identifying the parameters that would generate the most value if we were to gather more information about them. This is assessed by computing the value of eliminating uncertainty in a smaller group of parameters using the expected value of partial perfect information (EVPPI).[30] EVPPI is computed for different groups of parameters and those with the highest value should be considered as study outcomes. This information about the study outcomes of interest, helps to select the most appropriate study design. For example, a VOI analysis can help us to answer the question: "should we undertake a longitudinal cohort study to determine incidence or a randomised controlled trial (RCT) to determine treatment effect?".

The final VOI design phase uses the expected value of sample information (EVSI) to determine whether a specific trial design would give value. To undertake this final analysis, EVSI must be scaled up by the number of patients who could benefit from the research results to compute the population EVSI (popEVSI). If the popEVSI exceeds the cost of the proposed trial, the trial has value. To facilitate this analysis, we compute the expected net benefit of sampling (ENBS), defined as the difference between the popEVSI and the study cost.[3] If ENBS is less than zero, then this trial is not valuable, and the current best intervention should be used in clinical practice. Finally, ENBS can be computed for different study protocols by changing the study type, inclusion/exclusion criteria and sample size, to determine the most valuable trial.[24] Given that there is a cost associated with enrolling participants, the *value-driven* approach enrols participants until the value of their data is smaller than the cost of enrolling them in the trial.[24]

Figure 2



Details of the 'value-driven' approach. Table 2 gives explanations of each of the steps.

Table 2

Steps of the value-driven approach to prioritise and design future clinical trials. The research question focuses on a decision problem in clinical medicine or public health. Each step uses specific methods. Interpretation of the results of each step affects whether it is useful to proceed to the next step.

Specific question	Methods	Interpretation
What is already known?	Evidence synthesis, systematic reviews, (individual level) meta-analysis	Review and calculate summary measures of the best-available evidence.
What is best for our patient population?	Decision modeling, state-transition cohort model, microsimulation	Expected value of each strategy is the value that we can expect on average given the current best-available evidence on risks, benefits, costs and patient values.
How certain are we that the decision is the best choice?	Probabilistic sensitivity analysis (PSA)	Propagate the uncertainty in the input parameters through the model to estimate the uncertainty in the outputs.
What is the value of eliminating all uncertainty around all parameters, per patient?	Expected value of perfect information (EVPI)	The expected value for the hypothetical scenario where further research eliminates all decision uncertainty, meaning that we would know the exact values of all parameters.
What is the value of eliminating all uncertainty around all parameters, considering all patients that could benefit?	Population expected value of perfect information (popEVPI)	popEVPI equals the EVPI per patient multiplied by the number of patients that can benefit from the new information. The popEVPI is the maximum obtainable value of performing more research. It needs to exceed the cost of performing new research in order to proceed to the next step. This is a necessary but not sufficient condition to do more research.
Which parameters are driving the uncertainty around the decision?	Expected value of partial perfect information (EVPPI)	EVPPI calculates the EVPI of a limited subset of parameters. The EVPPI guides the design of a future study by focusing the data collection on those parameters that are the most valuable to collect.
What is the value of reducing uncertainty rather than eliminating it, per patient?	Expected value of sample information (EVSI)	EVSI is the expected value of reduction in uncertainty by collection of a limited subset of parameters in a study with limited sample size, which depends on the sample size of the envisioned trial.
What is the value of reducing uncertainty, considering all patients that could benefit?	Population expected value of sample information (popEVSI)	popEVSI is the EVSI per patient multiplied by the number of patients that can benefit from the new information.
Determine the "cost" of doing more research	Resources for new study + foregone benefit	Cost of a new study includes the resources required to perform the study, the foregone benefit due to delayed implementation of a potentially beneficial intervention, and the foregone benefit due to allocation to a suboptimal strategy in an RCT.
Does the information provided by a new trial justify its cost?	Expected net benefit of sampling (ENBS): popEVSI of the trial – Cost of the trial	ENBS is the expected net benefit of reducing uncertainty by collection of data depending on sample size of the trial. ENBS needs to exceed zero in order to justify performing the trial.
Which trial design is optimal?	Compare the ENBS of different trial protocols	Choose the trial design that maximises the ENBS. Choices are the study design, inclusion/exclusion criteria, sample size, assignment ratio, criteria for adaptive design.

How the value-driven approach addresses challenges in trial design

This section outlines how the *value-driven* approach can offer solutions to challenges researchers have to deal with when developing and performing clinical trials.

Efficiency

Clinical trials are expensive and time-consuming, mainly related to the required number of trial participants. The *value-driven* approach assumes that trial data is analysed within the totality of evidence relevant to the policy decision. This can reduce the sample size, cost and patient burden.

The *value-driven* approach also improves trial design efficiency when multiple interventions are available. In the *value-driven* approach, decisions about which interventions to include are made by considering which interventions are key to determining the best intervention upon completion of the research process. The *error-driven* approach uses expert consensus to determine the trial interventions, which may exclude valuable interventions from the trial.

Finally, the time required for research and reimbursement decisions delays implementation, which can have health consequences as effective treatments are slow to reach patients (see Box 2). The *value-driven* approach addresses the foregone benefits of delayed implementation explicitly as a cost of further research. Furthermore, it ensures that information for decision/policy making is available at the start of the trial and can be updated following the trial. Finally, clinical trial design is optimised to support decision making. The downside of this approach is that in-depth analysis is required in the trial planning phase, which requires more time and resources.

Box 2

Example of value-driven approach in trial design

Willan and Kowgier compared value of information methods to traditional power calculations.[33]

Example: A randomized clinical trial (RCT) funded by the Canadian Institute of Health Research (CIHR) investigating early vs late external cephalic version (ECV) for pregnant women presenting with a fetus in breech position.

Error-driven Approach:
Primary outcome: Non-Caesarean delivery.
Sample Size Calculation: The investigators of the trial used evidence from a pilot study (n = 116 in both arms, where the proportion of non-Caesarean deliveries in the early ECV arm was 35.3% compared to 28.4% in the late ECV arm).[34] The minimally clinically important difference was determined to be an 8-percentage-point increased probability of a non-Caesarean delivery in the early ECV arm. The type-II error rate of the trial was set to 0.20 with a two-sided type-I error of 0.05. Thus, the trial had an 80% probability of correctly rejecting the null hypothesis if the treatments differed by eight percentage points or more and a 5% probability of incorrectly rejecting the null hypothesis if there was no difference between treatments. The sample size was calculated for a two-sample test for proportions, including a continuity correction to adjust for binary outcomes.[35]

Sample Size: 730 patients per arm
This large trial was successfully funded by CIHR and completed in 2008.[36]

Value-driven Approach:
Estimating Value: The prior distribution of the incremental net benefit was estimated based on the pilot data (probability difference (41/116 – 33/116)) combined with the assumed societal willingness-to-pay of $1,000 to achieve a non-Caesarean delivery. To estimate the total number of patients affected by the decision, a time horizon of 20 years and annual North American incidence of 100,000 breech presentations was assumed.

Sample Size Calculation: A decision model was developed to estimate the effects of both strategies based on the published pilot data. Next the uncertainty around the decision was simulated and

the expected value that can be gained by reducing uncertainty (EVPPI) was calculated. This expected value of new evidence minus the cost of collecting this information yielded the expected net benefit of further research (ENBS). The sample size that maximizes the ENBS was selected as the optimal sample size.

Sample Size: 345 patients per arm
The *value-driven* approach would have resulted in a 52.7% reduction in trial sample size.

Efficiency:
The required trial budget would be reduced from $2,836,000 to $1,604,000 (43.4% reduction)
The expected net monetary benefit of the trial would increase from $179,000 to $736,383 (around 4 times higher).

Note that the two approaches take different perspectives allowing the value-based approach to require a lower sample size than the 730 patients required to achieve statistical significance in the error-driven approach. The value-based approach does not aim to achieve statistical significance. Instead, it optimizes the trade-off between collecting more information, which is costly, and making an incorrect decision about the best treatment. The value-based approach considers that a decision can be made between two treatments, even if the difference between them on some clinically-relevant outcome is not statistically significant.[22] This allows the *value-driven* approach to potentially increase trial efficiency to such an extent that justifies the added complexity of designing research with this approach.[33]

Note that Willan and Kowgier[33] also consider a two-stage value-based approach which increases the expected net benefit of the trial still further to approximately 8 times higher than that of the error-driven approach.

Generalisability

Clinical trials are often criticised for their lack of generalisability[31] and their use of outcomes with limited relevance to clinicians, patients or policy makers.[32] The *value-driven* approach addresses these issues by ensuring that trial outcomes will support decision-making. The value of trials collecting evidence on alternative outcomes, i.e., short-term surrogate outcomes vs long-term outcomes, can be compared to their required resources. This would determine whether the additional information in the long-term outcomes is worth the increased complexity and cost. The value of a trial is also proportional to the number of people affected by the decision (Table 2). Thus, if the trial has limited generalisability, the value of the trial is limited. This supports the development of trials with wide inclusion criteria. Moreover, reducing the time between trials and their implementation ensures that trial information more closely reflects current practice.

Validity

Clinical trials can have issues with validity as patients switch interventions, are lost to follow-up and do not follow protocol.[1,37] While treatment switching and protocol adherence are issues for the *value-driven* approach too, we can assess the impact of losing patients to follow-up and we can value efforts to reduce loss-to-follow up, i.e., financial incentives for follow-up questionnaires.[38] Furthermore, trial simulations that consider these issues can define how they influence the value of the trial.

Feasibility

The *value-driven* approach directly considers the available budget, the time-taken to undertake the trial and the delay in widespread implementation.[3] Thus, the *value-driven* approach designs trials that are, by definition, feasible conditional on budgetary and time constraints. Conversely, the *error-driven* approach can lead to designs that are infeasible, i.e., requiring infeasible sample sizes in rare diseases.[39]

Personalised and precision medicine

Precision medicine is becoming an important part of healthcare[40,41] but causes methodological issues in trial design.[42] However, by focusing on supporting personalised decision-making, the *value-driven* method can offer alternative trial designs that are feasible and generalisable. Furthermore, novel *value-driven* methods are available to optimise the design of trials in precision medicine.[43]

Emerging technologies

Finally, fast evolving technologies can mean that interventions are outdated before trial completion. The *value-driven* approach considers that trial evidence will be added to the current evidence, facilitating adaptive trials compared to the *error-driven* approach. We can assess the value of new interventions and compute the value of adapting the trial to include them. Thus, the *value-driven* approach includes flexibility that ensures evidence is relevant to decision-making, even in the face of emerging technologies and a changing research landscape.

Discussion

This paper proposes the *value-driven* approach as an alternative to the current *error-driven* approach for clinical research studies, focusing on clinical trial design although these methods are also applicable to observational studies. We now discuss project- and system-level barriers to the widespread implementation of the *value-driven* approach and highlight the potential benefits of the approach.

Time required for trial design

Under the *value-driven* approach, it takes more information and time to design a trial as the value of each intervention must be estimated using decision modeling. This requires, ideally patient-level, evidence on the interventions' costs and benefits as well as their effect sizes. In contrast, the *error-driven* approach focuses on a key primary outcome for interventions that have been selected for the trial using expert consensus. Thus, the *value-driven* approach requires a wider literature search and different modeling and data synthesis methods. However, these analyses are required for policy making and thus, we can reduce the time for trial data analysis by including them at the design stage. The increased time and cost of the research prioritisation and trial design process will require additional funds. However, as the *value-driven* approach optimises the spending of research resources, the savings from efficient and effective trials are expected to recoup the cost of this design phase.

Data access

The decision modeling required in the *value-driven* approach and the final trial analysis should include all the currently available data. This includes data in aggregated form from the literature and patient-level data from previous trials. Accessing these data to design a new trial may be challenging and, if these data are not made available, then the *value-driven* approach could develop designs that result in inefficient use of resources. However, wider data access and the secondary use of trial data are increasingly used to improve the efficiency of healthcare research.[44] In addition, recent efforts by

academic journals on data sharing (e.g., Plos ONE requirements for public data access[45]) can facilitate such data access, and further efforts by other academic journals would facilitate this. Thus, reusing data to improve trial design and to ensure that research effectively targets decision uncertainty should be a key part of this effort.

Expertise required for trial design

The *value-driven* approach requires collaboration with an interdisciplinary group of researchers, including trialists, statisticians, health economists, and decision modelers for the trial design. There is a lack of expertise among trialists and statisticians with VOI methods, in part because the methods have been challenging to implement.[26] However, recent research has focused on facilitating VOI analyses[26,46,47] reducing the barriers to implementation by increasing education and software.[25,48–51] Including researchers familiar with cost-effectiveness and VOI methods in the trial design will increase costs, again, offset by the more seamless and efficient use of resources in the trial and its analysis. Specifically, this collaboration ensures that information required for cost effectiveness analyses can be collected in the trial outcomes.

Adaptive research questions

If funding is available for the design and conduct of a trial, challenges may arise if the VOI analysis indicates that the proposed trial is an inefficient use of research funding, e.g., an alternative study or a smaller sample size may be required. In this case, funding may need to be returned or repurposed. Flexible funding instruments would allow researchers to undertake valuable research, even when it was not originally proposed. To benefit fully from the *value-driven* approach, the current method of research funding where deliverables are pre-specified will require modification.

Status quo in regulatory process

Regulatory authorities worldwide have strict guidelines around the type of evidence that must be submitted to demonstrate treatment efficacy and safety. If a trial is developed as the basis of a submission to these regulatory bodies, innovative trial designs, such as those developed through the *value-driven* approach, may be limited by those guidelines. However, there is an increasing trend for regulatory authorities to become more flexible and acceptive of innovative and efficient trial designs (e.g., umbrella/basket trial designs) given the challenges facing the current regulatory landscape (e.g., personalized medicine, expedited access). Additionally, *value-driven* approaches can suggest expanding the data collection beyond the typical primary efficacy/safety outcomes (e.g., costs or quality of life), thereby strengthening the evidence submitted to a regulatory body.

Current Research Infrastructure

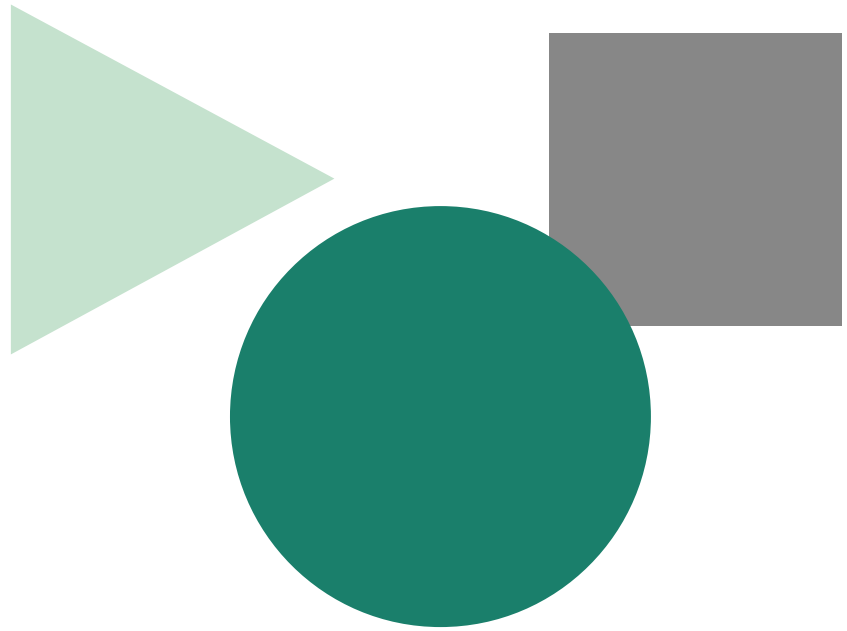
The current research infrastructure and publication culture support the *error-driven* approach, e.g., statistically significant results increase the chances of publication in high impact journals[52], and analysing trial results within a decision model is less well accepted. However, the *error-driven* approach has been heavily criticised[13] and journals are beginning to accept trials analysed using alternative methods.[53]

In conclusion, the *value-driven* approach has advantages over the error-driven approach as research performed based on a *value-driven* trial design will collect data that are valuable to society and thus reduce research waste.[54] The *value-driven* approach can also justify a more streamlined implementation of interventions, which is particularly important when facing an urgent situation affecting a large number of patients. The *value-driven* approach can guide the choice of study type, inclusion/exclusion criteria, sample size, allocation ratio, and criteria for adaptive designs.

References

1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009; 274(4):86–89
2. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996; 5(6):513–524.
3. Conti S, Claxton K. Dimensions of design space: a decision-theoretic approach to optimal research design. *Med Decis Mak*. 2009; 29(6):643–60.
4. Fleurence RL, Torgerson DJ. Setting priorities for research. *Health Policy (New York)*. 2004; 69(1):1–10.
5. Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. *PLoS Med*. 2015; 12(9):1–5.
6. Mooney G, Wiseman V. Burden of disease and priority setting. *Health Econ*. 2000; 9(5):369–372.
7. Flight L, Julious SA. Practical guide to sample size calculations: Superiority trials. *Pharm Stat*. 2016; 15(1):75–79.
8. Sakpal TV. Sample size estimation in clinical trial. *Perspect Clin Res*. 2010; 1(2):67–69.
9. Chow SC, Liu JP. *Design and analysis of clinical trials: concepts and methodologies*. Second edi. John Wiley & Sons., 2008.
10. Zhong B. How to calculate sample size in randomized controlled trial? *J Thorac Dis*. 2009; 1(1):51–54.
11. Bouter LM, Zielhuis GA, Zeegers MPA. *Textbook of Epidemiology*. 1st ed. Houten: Bohn Stafleu van Loghum, 2018.
12. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005; 2(8):e124.
13. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019; 567(7748):305–307.
14. NIH U.S. National Library of Medicine. Clinicaltrials.gov. *Visited May 2020*, Clinicaltrials.org.
15. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. *Decision Making in Health and Medicine. Integrating Evidence and Values*. 2nd ed. Cambridge University Press, 2014.
16. Freiman JA, Chalmers TC, Smith H, Kuebler RR. The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial. *N Engl J Med*. 1978; 299(13):690–694.
17. Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S. A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme. *Health Technol Assess (Rockv)*.; 8(31). Epub ahead of print 2004. DOI: 10.3310/hta8310.
18. Briggs A, Claxton K, Sculper M. *Decision Modelling for Health Economic Evaluation*. Oxford University Press, 2006.
19. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med*. 1989; 28(4):299–308.
20. Stinnett AA, Mullahy J. Net Health Benefits: A New Framework for the Analysis of Uncertainty in Cost-Effectiveness Analysis. *Med Decis Mak*. 1998; 18(2_ suppl):S68–S80.
21. Neumann PJ, Weinstein MC. Legislating against use of cost-effectiveness information. *N Engl J Med*. 2010; 363(16):1495–1497.
22. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999; 18(3):341–364.
23. Claxton K, Sculpher M. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics*. 2006; 24(11):1055–1068.
24. Willan AR, Pinto EM. The value of information and optimal clinical trial design. *Stat Med*. 2005; 24(12):1791–1806.
25. Baio G, Berardi A, Heath A. *Bayesian Cost-Effectiveness Analysis with the R package BCEA*. Cham: New York: Springer 2017.
26. Kunst NR, Wilson E, Alarid-Escudero F, Baio G, Brennan A, Fairley M, Glynn D, Goldhaber-Fiebert JD, Jackson C, Jalal H, Menzies NA, Strong M, Thom H, Heath A. Computing the Expected Value of Sample Information Efficiently: Expertise and Skills Required for Four Model-Based Methods. *Value Health*. 2020; 20(6):734–742.
27. Alarid-Escudero F, Knowlton G, Easterly C, Enns E. Decision Analytic Modeling Package (dampack). R package version 1.0.0. 2021. https://github.com/DARTH-git/dampack
28. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www. train ing. cochr ane.org/handbook.
29. Wilson ECF. A practical guide to value of information analysis. *Pharmacoeconomics. 2015;* 33(2):105–121.
30. Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Mak*. 1998; 18(1):95–109.
31. Sharpe N. Clinical trials and the real world: Selection bias and generalisability of trial results. *Cardiovasc Drugs Ther*. 2002; 16(1):75–77.
32. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, Howells DW, Ioannidis JPA, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014; 383(9912):156–165.
33. Willan A, Kowgier M. Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clin Trials J Soc Clin Trials*. 2008; 5(4):289–300.

34. Hutton EK, Kaufman K, Hodnett E, Amankwah K, Hewson SA, McKay D, Szalai JP, Hannah ME. External cephalic version beginning at 34 weeks' gestation versus 37 weeks' gestation: A randomized multicenter trial. *Am J Obstet Gynecol*. 2003; 189(1):245–254.
35. Fleiss J, Levin B, Cho Paik M. *Statistical Methods for Rates and Proportions*. Third edit. Hoboken, New Jersey: John Wiley & Sons, Inc., Publication, 2003.
36. Hutton E, Hannah M, Ross S, Delisle M-F, Carson G, Windrim R, Ohlsson A, Willan A, Gafni A, Sylvestre G, Natale R, Barrett Y, Pollard J, Dunn M, Turtle P. The Early External Cephalic Version (ECV) 2 Trial: an international multicentre randomised controlled trial of timing of ECV for breech pregnancies. *BJOG An Int J Obstet Gynaecol*. 2011; 118(5):564–577.
37. Henshall C, Latimer NR, Sansom L, Ward RL. Treatment switching in cancer trials: Issues and proposals. *Int J Technol Assess Health Care*. 2016; 32(3):167–174.
38. Heath A, Manolopoulou I, Baio G. Estimating the Expected Value of Sample Information across Different Sample Sizes Using Moment Matching and Nonlinear Regression. *Med Decis Mak*. 2019; 39(4):346–358.
39. Abrahamyan L, Willan AR, Beyene J, Mclimont M, Blanchette V, Feldman BM. Using value-of-information methods when the disease is rare and the treatment is expensive - The example of hemophilia A. *J Gen Intern Med*. 2014; 29(SUPPL. 3):767–773.
40. Collins FS, Varmus H. A New Initiative on Precision Medicine. *N Engl J Med*. 2015; 372(9):793–795.
41. Maughan T. The Promise and the Hype of 'Personalised Medicine.' *New Bioeth*. 2017; 23(1):13–20.
42. Senn S. Mastering variation: Variance components and personalised medicine. *Stat Med*. 2016; 35(7):966–977.
43. Fairley M, Cipriano LE, Goldhaber-Fiebert JD. PS 4-55 Optimal allocation of clinical trial sample size to subpopulations with correlated parameters. 41st Annual meeting of the society for medical decision making, Portland, Oregon, October 20–23, 2019. (2020). Med Decis Mak, 40(1), E1–E379.
44. Cheng HG, Phillips MR. Secondary analysis of existing data: opportunities and implementation. *Shanghai Arch Psychiatry*. 2014; 26(6):371–375.
45. PLOS ONE. Data Availability, https://journals.plos.org/plosone/s/data-availability (accessed November 26, 2020).
46. Heath A, Manolopoulou I, Baio G. A Review of Methods for Analysis of the Expected Value of Information. *Med Decis Mak*. 2017; 37(7):747–758.
47. Heath A, Kunst N, Jackson C, Strong M, Alarid-Escudero F, Goldhaber-Fiebert JD, Baio G, Menzies NA, Jalal H. Calculating the Expected Value of Sample Information in Practice: Considerations from 3 Case Studies. *Med Decis Mak*. 2020; 40(3):314–326.
48. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, Koffijberg H (Erik), Strong M, Sanders Schmidler GD, Rothery C. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020; 23(2):139–150.
49. Rothery C, Strong M, Koffijberg H (Erik), Basu A, Ghabri S, Knies S, Murray JF, Sanders Schmidler GD, Steuten L, Fenwick E. Value of Information Analytical Methods: Report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020; 23(3):277–286.
50. Strong M, Brennan A, Oakley J. How to Calculate Value of Information in Seconds Using 'Savi', the Sheffield Accelerated Value of Information Web App. *Value Health*. 2015; 18(7):A725–A726.
51. Grimm SE, Strong M, Brennan A, Wailoo AJ. The HTA Risk Analysis Chart: Visualising the Need for and Potential Value of Managed Entry Agreements in Health Technology Assessment. *Pharmacoeconomics*. 2017; 35(12):1287–1296.
52. Cristea IA, Ioannidis JPA. P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PLoS One*. 2018; 13(5):1–15.
53. Quintana M, Viele K, Lewis RJ. Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. *JAMA*. 2017; 318(16):1605.
54. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, Salman RA-S, Chan A-W, Glasziou P. Biomedical research: increasing value, reducing waste. *Lancet*. 2014; 383(9912):101–104.



EMERGING THERAPIES FOR COVID-19: THE VALUE OF INFORMATION FROM MORE CLINICAL TRIALS

Stijntje W. Dijk
Eline Krijkamp
Natalia Kunst
Cary P. Gross
John B. Wong
M.G. Myriam Hunink

Value in Health – accepted

“Progress always involves risk. You can’t steal second base and keep your foot on first.”

Frederick B. Wilcox

EMERGING THERAPIES FOR COVID-19: THE VALUE OF INFORMATION FROM MORE CLINICAL TRIALS

Abstract

The COVID-19 pandemic necessitates time-sensitive policy and implementation decisions regarding new therapies in the face of uncertainty. The aim of this study was to quantify consequences of approving therapies or pursuing further research: either immediate approval, use only in research, approval with research (e.g., Emergency Use Authorization), or reject.

Using a cohort state-transition model for hospitalized COVID-19 patients, we estimated quality-adjusted life years (QALYs) and costs associated with the following interventions: Hydroxychloroquine, Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Baricitinib-Remdesivir, Tocilizumab, Lopinavir-Ritonavir, and Interferon beta-1a, and usual care. We used the model outcomes to conduct cost-effectiveness and value of information analyses from a US healthcare perspective and a lifetime horizon.

Assuming a $100,000-per-QALY willingness-to-pay-threshold, only Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Baricitinib-Remdesivir and Tocilizumab were (cost-) effective (incremental net health benefit 0.252, 0.164, 0.545, 0.668 and 0.524 QALYs and

incremental net monetary benefit $25,249, $16,375, $54,526, $66,826 and $52,378). Our value of information analyses suggest that most value can be obtained if these 5 therapies are approved for immediate use rather than requiring additional RCTs (net value $20.6 Billion, $13.4 Billion, $7.4 Billion, $54.6 Billion and $7.1 Billion); Hydroxychloroquine (net value $198 Million) only used in further RCTs if seeking to demonstrate decremental cost-effectiveness, and otherwise rejected; and Interferon beta-1a and Lopinavir-Ritonavir are rejected (i.e., neither approved nor additional RCTs).

Estimating the real-time value of collecting additional evidence during the pandemic can inform policymakers and clinicians about the optimal moment to implement therapies and whether to perform further research.

Introduction

Amidst over 250 million cases worldwide, and 3,000-5,500 daily hospitalizations in the US alone, the COVID-19 pandemic represents the greatest global public health crisis since 1918.[1,2] In the absence of known effective pharmaceutical interventions during early stages of the pandemic, many clinicians prescribed treatments off-label. Since the start of the outbreak, over 3,500 clinical trials investigating potential therapies have been registered[3], and new trials continue to emerge. These trials are all competing for resources and patient enrollment. Decisions on early implementation of promising treatments have been the source of substantial academic and public debate, however objective criteria for research prioritization remain absent.[4,5]

Policymakers and clinician-researchers face a difficult choice; giving Emergency Use Authorization[6] (approval of the drug conditional on conducting more research); approval of the drug for wide-spread clinical implementation; approval of the drug only in research; or rejection of the drug based on limited existing data.[4] A "*study, then treat or reject*" approach would optimize expected benefit by gaining more certainty about treatment effects, whereas a "*treat first, investigate later*" approach seeks to prevent lives being lost due to delayed implementation and denial of potentially beneficial treatments. However, this strategy increases the risk of harm from implementation of a possibly ineffective or deleterious treatment.

At a given point in time, findings from completed clinical trials – representing the current body of evidence - can be modelled to provide an estimate of the potential (health) benefits of further research or implementing findings of existing research.[7] A key tool to quantify this cost-benefit trade-off is Value of Information analysis (VOI). VOI quantifies the value of treatment choices made with the expected evidence from additional research, compared to making the choice based on currently available information.[7–9] VOI is increasingly applied as part of health economic evaluations[10–17], both to aid the determination of the optimal sample size as well as to direct research efforts to where the greatest return can be expected from finite resources.[10] Whereas meta-analyses investigate drug efficacy and effectiveness, and commonly conclude that further research is needed based on lack of statistical significance[7], VOI results consider both current uncertainty relevant to the decision as well as the potential consequences of making decisions with and without further evidence.[16] These VOI analyses can be used to quantify the benefit of further research, in terms of reducing uncertainty around treatment efficacy and avoiding unintended harm that would result from premature use of a therapy that turns out to be ineffective or deleterious. The benefits of further research are then balanced against possible forsaken benefits due to delayed implementation and research costs.[7]

In this study, we apply VOI to express the value of performing either further RCTs with delayed approval decisions (use only in RCTs), emergency use while performing RCTs (approval with research), or immediate approval of treatments for COVID-19 versus rejection without further research. Focusing on drug therapies for hospitalized COVID-19 patients for which large RCTs or meta-analyses have been published to date, we examined Hydroxychloroquine[18], Remdesivir[19,20],

Box 1

Glossary of terms

Willingness to pay (WTP)	A threshold that represents what the decision maker or society is willing to pay for a unit of health outcome. The threshold is expressed in monetary units per health outcome.
Incremental Cost Effectiveness Ratio (ICER)	A ratio demonstrating the trade-offs between costs and benefits, calculated as the ratio of the incremental cost of an intervention to the incremental benefit in health outcomes.
Incremental Net Health Benefit (iNHB)	A summary statistic representing the impact of an intervention on a population's health for a given willingness-to-pay threshold, compared to an alternative intervention, calculated as: incremental health benefit – incremental cost of the intervention / willingness-to-pay threshold.
Incremental Net Monetary Benefit (iNMB)	A summary statistic representing the value of an intervention in monetary terms for a given willingness-to-pay threshold, compared to an alternative intervention, calculated as: incremental health benefit × willingness-to-pay threshold – incremental cost of the intervention.
Incremental Net Benefit (iNB)	A summary statistic representing the impact of an intervention on population outcome compared to an alternative intervention, calculated as either incremental net health benefit or incremental net monetary benefit.
Probabilistic Analysis (PA)	A technique used to propagate uncertainty from model inputs to model outcomes. Also referred to in the literature as Probabilistic Sensitivity Analysis (PSA).
Value of Information analysis (VOI)	The estimation of decision uncertainty and the value of collecting more information on key parameters influencing a decision, expressed in monetary or health terms.
Overall strategy	The combined choice of strategy with respect to both treatment and research. The options for the overall strategy are: - **Only in Research (OIR)**, where the drug is made available to patients participating in further research trials, but no Emergency Use Authorization nor widespread use is granted. - **Approval with Research (AWR)**, where Emergency Use Authorization is granted, the drug is made available to patients, and in addition further research is being conducted. - **Approval**, where the drug is immediately approved and no further research is conducted. - **Reject**, where the drug is rejected and no further research is conducted. This situation is the default strategy where usual care does not change.
Population Expected Value of Partial Perfect Information (EVPPI)	The value of collecting perfect information on selected parameter(s) or subset(s) of parameters in the model, extrapolated to the size of the target population that can benefit from the information (future patients) over a specific time horizon.
Population Expected Value of Sample Information (EVSI)	The value of collecting additional information on selected parameter(s) or subset(s) of parameters in the model with a trial with finite sample size, extrapolated to the size of the target population that can benefit from the information (future patients) over a specific time horizon.
Costs of performing research	Resources required to perform a new trial (fixed cost and variable cost per participant) plus, for study participants, the foregone benefit due to randomized assignment to suboptimal treatment in the trial.
Net benefit due to implementation	Incremental Net Monetary (or Health) Benefit that is gained because implementation of a beneficial therapy is approved, either through Emergency Use Authorization or definitive approval. This net benefit is foregone in current patients if approval and implementation are delayed while more evidence is obtained from further RCTs (OIR strategy).
Net Value of research (new RCT)	Expected value of performing further research, in this analysis a new RCT, (population EVSI) minus the cost of performing the RCT.
Net value of the Overall strategy	The net value of the combined treatment-and-research strategy which equals the net value of performing an RCT if further research is performed plus the net benefit of treatment if approved.

Casirivimab-Imdevimab[21], Dexamethasone[22], Baricitinib-Remdesivir[23], Tocilizumab[24], Lopinavir-Ritonavir[25] and Interferon beta-1a[20] compared to usual care. We considered this as a non-competing choice problem, since each of these drugs may be beneficial in the armamentarium of drug therapies for COVID-19. Our analysis aims to inform both treatment decisions and research prioritization decisions regarding therapies for hospitalized COVID-19 patients and demonstrate how a VOI approach can inform clinical and public health decision-making during a pandemic.

Methods

Decision trade-off

We performed a model-based cost-effectiveness analysis to examine both costs and health outcomes for all treatments considered. Next, we used the VOI framework to determine the net value of research. This estimate quantifies the trade-off between resources required for another RCT versus the added value of the RCT to gain more solid evidence (Figure 1).[26] If the net value of research exceeds zero, performing a new RCT is worthwhile and the new evidence should be incorporated into the decision-making (*upper quadrants*). Ideally, trials are performed until the cost of future research exceeds the expected benefits (*lower quadrants*). This process can be applied to trials that demonstrate potential beneficial treatment effects (*right quadrants*), as well as those demonstrating no beneficial effects (*left quadrants*).

Model description

We simulated the effect of treatment for a cohort of hospitalized COVID-19 patients based on meta-analyses or large multicenter RCT's that reported in-hospital mortality. For this simulation we used a Markov cohort state-transition model with four health states (hospitalized, recovered from hospital ward as highest level of care, recovered from intensive care unit (ICU) and dead (Appendix-Figure 1). As in the RCT's, Hydroxychloroquine[18], Remdesivir[19,20], Casirivimab-Imdevimab[21], Dexamethasone[22], Baricitinib-Remdesivir[23], Tocilizumab[24], Lopinavir-Ritonavir[25] and Interferon beta-1a[20] were compared to the control arm (usual care) rather than to each other. We considered the implementation and research decision as a non-competing choice problem. Each drug has the potential to be a valuable component in the armamentarium against COVID-19. The various drugs have different therapeutic mechanisms, for example, as an antiviral or corticosteroid, and may be useful in sequence, in combination, or in different contexts. Similarly, from a regulatory perspective, approval is based on available safety and efficacy evidence. Since it is a non-competing choice problem, the incremental cost-effectiveness ratio (ICER) was calculated for each drug individually compared to care-as-usual. This choice allowed treatments investigated in earlier trials to become integrated in control arms in later trials. Furthermore, this modeling choice allowed for differences in context and trial populations, such as percentage of patients in the ICU and hospital wards.

All model parameters were based on best available evidence as of November 1, 2021 (Appendix-Table 1). The probabilities to transition between health states are based on a large UK cohort study[27] of hospitalized COVID-19 patients with a mean age of 72 years. For Dexamethasone and Tocilizumab, we assumed that only patients in the ICU received treatment with these drugs. This is in accordance with current treatment recommendations.[22] During hospitalization, patients either remained in their respective recovered states, or died from COVID-19 or other causes.[28] The maximum hospitalization duration was 73 days.[27] Post-hospitalization, patients were followed over their lifetime.

The model was developed in the statistical programming language R based on the DARTH framework.[29–31] We followed the CHEERS[32] and ISPOR[33] reporting and analysis recommendations. We validated the model by performing internal validation using the observational cohort data[27] and replicating the model's cohort transitions in the decision-analytical software Amua.[34]

Figure 1

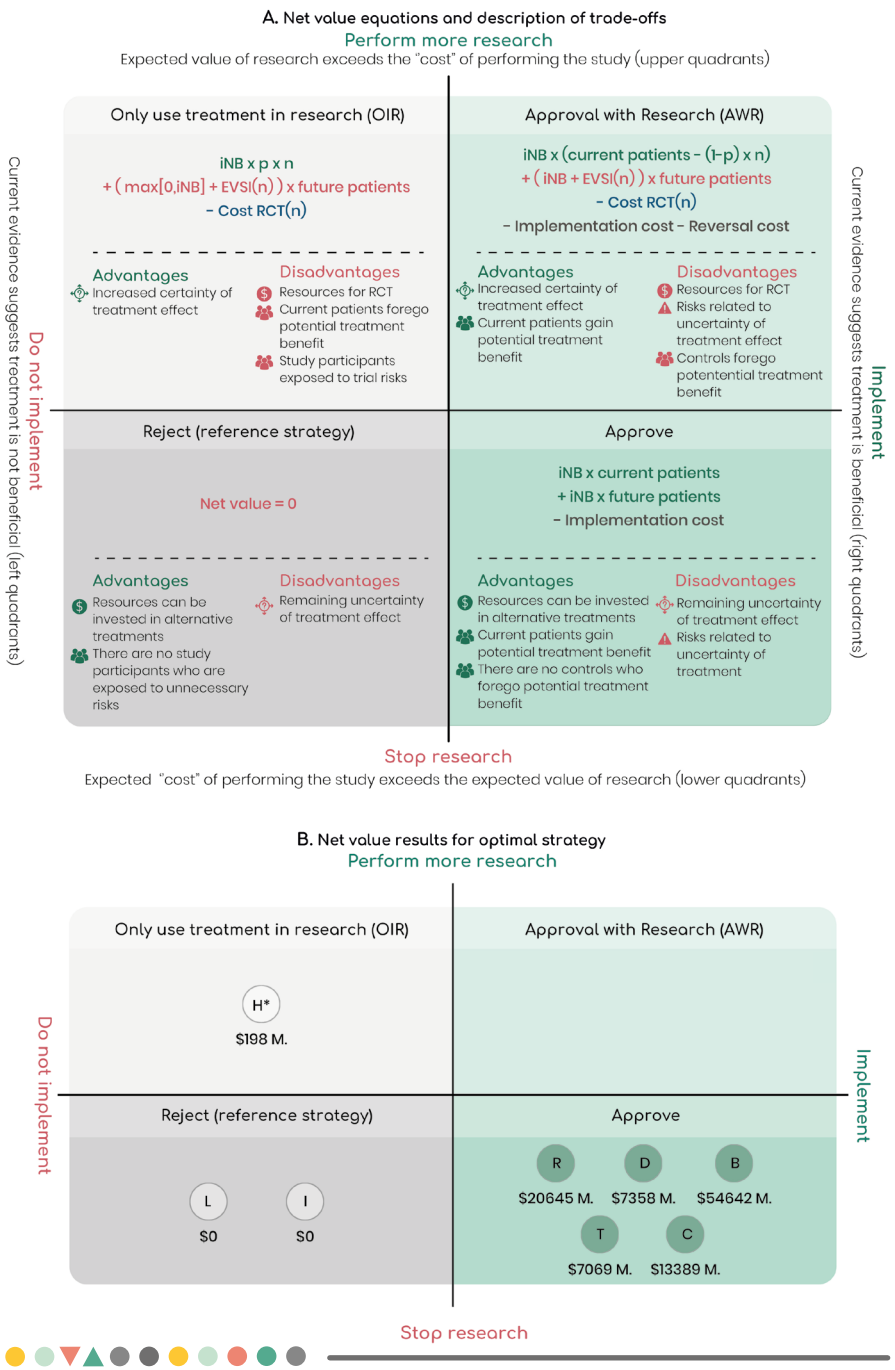


Figure 1: Trade-off between implementation of promising COVID-19 treatments and conducting further research.

*1A: "**Net value equations and description of trade-offs**". Demonstrates the equations used to quantify the net value for the overall strategy options, compared to reject as default strategy. These equations take iNB, EVSI, RCT cost, and number of patients (current and future) into account. iNB may be expressed in monetary units (NMB) or health units (NHB). One could also consider irrecoverable costs for the implementation of a new treatment or the possible reversal of implementation. However, in our analysis implementation and reversal costs are assumed negligible and therefor, shown in gray. The figure additionally shows the advantages and disadvantages of the corresponding implementation and research strategy. These quadrants are based on whether the drug's current evidence suggests benefit versus standard care or placebo (the right quadrants) or not (left quadrants). Within these right and left quadrants, the upper and lower quadrants indicate whether the value of doing additional research to reduce the uncertainty in benefit exceeds the "cost" of performing additional research expressed economically or (as quality-adjusted) life years lost) in the upper quadrants or not in the lower quadrants.*

Abbreviations:
***iNB** = incremental Net (Monetary or Health) Benefit (NB treatment – NB usual care)*
***p** = proportion of patients n in new RCT randomly assignet to trreatment. 1 – p is assigned to control*
***EVSI** is calculated in comparision to the optimal trreatment, ie. It is over and above iNMB gained if iNBM>0*
***EVSI(n)** = EVSI depends on sample size n of the the RCT*
***Cost RCT(n)** = fixed + variable cost of performing the RCT, where variable cost depends on sample size n.*
*The **optimal sample size** is determined by maximizing the fucntion in the quadrant with respect to n.*

*1B: "**Net value results for optimal strategy**". The net value results for the currently existing evidence and its uncertainty for eight drugs are calculated and each drug is placed in the resulting optimal health policy quadrant. Other factors, in particular ethical issues, also need to be considered to decide whether a strategy is desirable. For our study this is particularly true for Hyrdroxychloroquine. H* = Hydroxychloroquine: OIR has the highest net value if further research would demonstrate decremental cost-effectiveness (that is, saving costs but with loss of quality-adjusted life years). The ethics of investigating such decremental cost-effectiveness should be considered.¬ If not justifiable, then hydroxychloroquine would move to the Reject category, where the net value would be 0. **Abbreviations: H** = Hydroxychloroquine*, **L** = Lopinavir-Ritonavir, **I** = Interferon beta-1a, **R** = Remdesivir , **D** = Dexamethasone, **B** = Baricitinib-Remdesivir, **T** = Tocilizumab, **C** = Casirivimab-Imdevimab and **M**. = Million.*

Costs and effects

Health outcomes in the model were expressed as life years (LY) and quality-adjusted life-years (QALYs). Costs were estimated in 2020 US Dollars ($).[35] The analyses were performed from a US healthcare perspective. Costs during hospitalization included daily costs per person in a hospital ward or in the ICU and depended on the estimated mean length of hospital stay based on trial data. Treatment costs were based on the price proposed by the manufacturer or public pharmacy databases. Following hospitalization, patients in the ICU Recovery state accrued a one-time rehabilitation cost based on treatment needs as estimated by the Dutch National Health Authority guidelines[36] and converted to US prices for equivalent services. Recovered patients incurred mean healthcare costs for US citizens according to their age group annually until their death.[37] A review of 59 novel therapeutic drugs informed the costs for additional research, where the fixed cost estimate of trials up to 26 weeks were adjusted pro rata to represent a shorter trial duration of 3 months.[38,39] We applied a 3% annual discount rate for both costs and effects[40,] and a $100,000 willingness-to-pay (WTP) threshold.[41–43]

Uncertainty analyses and value of information

To assess the benefit of conducting further research, VOI analysis quantifies the opportunity cost of suboptimal decisions due to uncertainty. It takes a random sample of the value of each model parameter (Appendix-Table 1 and Supplementary Excel file) and evaluates the resulting outcomes to determine the optimal strategy (treatment or usual care) for that iteration (Probabilistic Analysis, PA). We performed a PA and calculated the expected value of each strategy for each of the 10,000 iterations. Input parameter distributions were lognormal (treatment effects), beta (utilities, transition probabilities), uniform or triangular (costs, or when distributions were not available from data sources).[26] Detailed information on all parameters is included in the Excel supplementary file.

With perfect knowledge of the parameter values chosen from their distributions for each of the 10,000 iterations of the PA, we determined the optimal strategy and its expected value for each iteration. Averaging these yielded the average expected value of the 10,000 decisions made with perfect information. Next, we calculated the difference between the average expected value of the 10,000 decisions made with perfect information and the decision based on the average expected value of each of the strategies, i.e. the decision made with current information. This difference yielded the loss in expected value due to suboptimal decisions as a result of parameter uncertainty, also known as the expected value of perfect information. We expressed the VOI results on a single scale of net monetary benefit (NMB) by converting QALYs to a monetary amount by multiplying these QALYs by a societal willingness-to-pay, e.g., $100,000 per QALY gained, and subtracting the resource costs.

Similarly, we calculated the expected added value of performing an RCT to reduce *only* the uncertainty surrounding treatment-related decrease in mortality as partial perfect information (EVPPI). For drugs with identified potential positive value of further research (EVPPI > 0), we determined the value of collecting additional information on treatment efficacy with a trial of finite sample size (expected value of sample information, EVSI). We performed an EVPPI estimation with a linear-regression meta-model and EVSI with a Gaussian approximation approach as proposed by Jalal and Alarid-Escudero.[44–46] In this approach the opportunity loss from a suboptimal decision is approximated by a linear relation of the parameters of interest.[43] This process was followed by a Gaussian approximation which simplifies the traditional Bayesian approach by computing the posterior mean for each of the parameters of interest (i.e., treatment efficacy in our analysis).[44] The approximation allows for multiple correlated parameters and parameters by different sample sizes, and a wide range of univariate and multivariate non-Gaussian distributions, and is computationally substantially more efficient than traditional Bayesian updating in EVSI.

Next, we calculated the cost of further trials based on fixed and variable cost across sample sizes. As it is impossible to obtain perfect

information, VOI places an upper bound on the cost of additional research aimed at reducing uncertainty.[16] The optimal sample size of a new RCT was calculated as the size at which the net value of the optimal overall strategy is highest. Given reported concerns of insufficient trial enrollment[5], we evaluated both the optimal and a maximum feasible sample size of 2500 patients (reported in Appendix section 1.4). Net benefit obtained with Emergency Use Authorization of treatments while performing further RCTs was determined for the expected number of patients to be hospitalized in the USA while awaiting trial results and their implementation (current patients) over 3 months. The expected value of information was extrapolated to the patient population that could benefit from new trial results, i.e., the number of patients expected to be hospitalized after the trial results are available (future patients). The number of patients was calculated as the sum of the number of daily hospitalizations forecasted by the Institute of Health Metrics and Evaluation as of November 1st 2021 until March 1st 2022.[47]

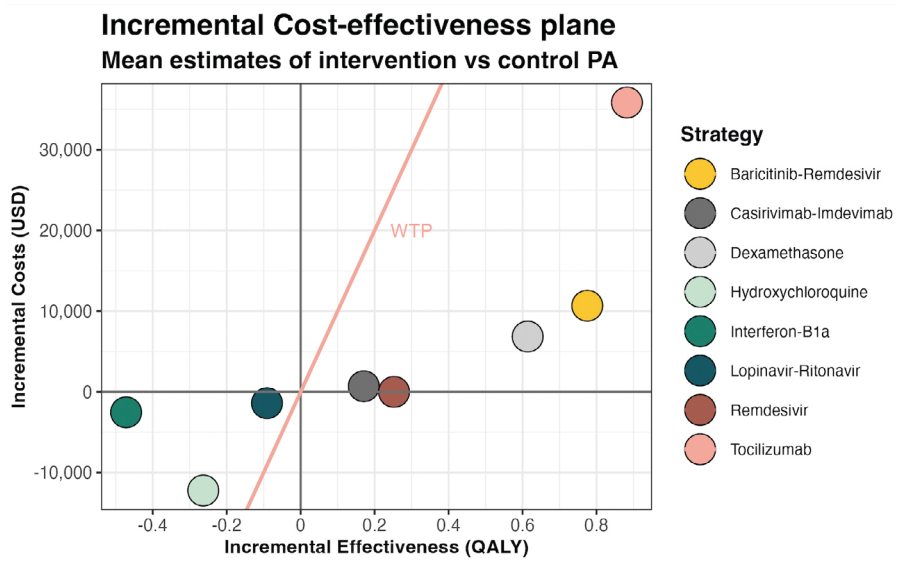
The net value for each strategy was calculated according to the equations in Figure 1A. Potential strategies included were Reject (*left lower quadrant*), Approve (*right lower quadrant*), Approve with Research (AWR, *right upper quadrant*), or use the drug Only in Research settings (OIR, *left upper quadrant*).[4] Rejection without further research (*left lower quadrant*) was considered the reference ("default") strategy (net value = 0). We assumed both the AWR and Approve strategies irrecoverable implementation and reversal costs[48] to be $0 since treatment protocols for COVID-19 are continuously and expeditiously updated, which does not require major investments, and no fixed capital investments are required for these treatments.

In sensitivity analyses we investigated main drivers of the results by testing extreme values and consequences of underlying modeling assumptions. In addition to the analyses in the manuscript assuming a WTP-threshold of $100,000, the Appendix provides the cost-effectiveness acceptability curves for each drug and in addition illustrates how the EVPPI results depend on the WTP threshold.

Ethics approval

Medical ethical review board approval was not required since we performed mathematical modeling and simulation using published data. No data from human participants was collected in this study.

Figure 2



Incremental cost-effectiveness plane for mean estimates per individual resulting from the Probabilistic Analysis (PA). Incremental costs in US dollars (USD) and effects are calculated and shown as the treatment group versus the control group within the respective trial, and not in comparison to the other treatments projected.

Results

Cost-effectiveness analysis

Table 1 summarizes key findings for each treatment strategy. Incremental costs and effectiveness (in QALYs) per individual are shown in Figure 2.

Our PA results indicate a decreased mortality and increased quality of life for Baricitinib-Remdesivir, Dexamethasone, Remdesivir and Tocilizumab, whereas Hydroxychloroquine, Interferon beta-1a and Lopinavir-Ritonavir are associated with increased mortality and decreased quality of life. Over a lifetime, the average QALY gains per patient were 0.775 [Uncertainty Interval -0.192;1.670] with Baricitinib-Remdesivir, 0.614 [0.039;1.258] with Dexamethasone, 0.252 [-0.141;0.670] with Remdesivir, 0.171 [-0.082;0.440] with Casirivimab-Imdevimab, and 0.882 [-0.052;1.937] for Tocilizumab; at an incremental cost of $10,673 [-$3930;$24372] for Baricitinib-Remdesivir, $6856 [-$19,696;$33,723] for Dexamethasone, $695 [-$33,082;$36,541] for Casirivimab-Imdevimab, and $35,849 [$20,447;$52,175] for Tocilizumab, and with marginal cost savings of -$5 [-$33,318;$235,724] for Remdesivir, making Remdesivir dominant, with higher effects and lower costs compared to usual care. Higher costs for Tocilizumab were mainly driven by treatment costs, whereas Baricitinib-Remdesivir, Dexamethasone and Casirivimab-Imdevimab costs were driven by the high healthcare costs during remaining life years in surviving patients. Conversely, lower costs for Hydroxychloroquine (-$12,227 [-$32,725;$7344]), Interferon beta-1a (-$2,538 [-$14,453;$8622]), and Lopinavir-Ritonavir (-$1,404 [-$40,584;$37,062]) were due to decreased survival leading to reduced future healthcare costs.

At a WTP of $100,000/QALY, positive incremental NMB's were found for Baricitinib-Remdesivir ($66,826 [-$15.895;$144,126]), Dexamethasone ($54,526 [-$10,111;$120,427]), Remdesivir ($25,249 [-$23,881;$73,206]), Casirivimab-Imdevimab ($16,375 [-$25,404;$57,641]), and Tocilizumab ($52,378 [-$30,049;$149,555]), consistent with being cost-effective. The remaining strategies Hydroxychloroquine, Interferon beta-1a and Lopinavir-Ritonavir were not cost-effective. The expected values as presented above should be used to identify the optimal strategy.[49] The uncertainty intervals reflect the range of decision uncertainty and the consequence of this uncertainty should be assessed in VOI analysis.

These PA results, uncertainty intervals, cost-effectiveness planes and cost-effectiveness acceptability curves for each treatment strategy are provided in the Appendix-Supplementary graphs and tables (Appendix-Supplementary graphs and tables –1.2/1.3/2.1-2.3/3.1-3.3/4.1/4.3/5.1-5.3/6.1-6.3/ 7.1-7.3/8.1-8.3/9.1-9.3).

Value of information and optimal overall strategy

The population EVPPI was only positive for Hydroxychloroquine ($375 Million), Remdesivir ($127 Million) and Tocilizumab ($1.4 Million) (Table 1, Appendix Supplementary graphs and tables–1.4/1.5/2.4/3.4/4.4/5.4/6.4/7.4/8.4/9.4), suggesting further RCTs to determine treatment effect more precisely may be worthwhile. Conversely, high certainty surrounding the treatment effect of Dexamethasone and Baricitinib-Remdesivir, and the absence of benefit of Lopinavir-Ritonavir and Interferon beta-1a suggest that performing more RCTs will not affect the decision about the use of these treatments in clinical practice (Table 1). This indicates that implementing Dexamethasone ($7.4 Billion), Casirivimab-Imdevimab ($13.4 Billion), and Baricitinib-Remdesivir ($54.6 Billion), and rejecting Lopinavir-Ritonavir ($0) and Interferon beta-1a ($0) without further trials were implementation strategies with the highest overall value.

In order to decide whether further research is warranted for Hydroxychloroquine, Remdesivir, and Tocilizumab, the net value of the relevant strategies (Figure 1B: *upper quadrants*) was calculated for an optimal trial sample size and then compared to Reject or Approve as appropriate. The optimal overall strategy for Hydroxychloroquine, Only in Research (Figure 1B: *left upper quadrants, OIR*), had an estimated net value of $198 Million with an optimal sample size of 4800 patients. For the pre-set maximum feasible sample size of 2500 patients, this net value is reduced to $174 Million. The value of this further research would stem from the investigation of decremental cost-effectiveness, answering the question whether the cost-savings justify the (QA)LYs lost, which has ethical implications.

For Remdesivir and Tocilizumab, the potential benefit of further trials did not outweigh the costs of research, and the highest net value ($21 Billion and $7 Billion, respectively) was obtained with approval (*Figure 1B: right lower quadrant, Approve*). Our findings suggest that none of the investigated treatments should be granted Emergency Use Authorization (*right upper quadrant, AWR*).

Sensitivity analyses

The cost-effectiveness acceptability curves illustrating the cost-effectiveness across WTP-thresholds, and tables displaying the EVPPI for different WTP thresholds are provided in the Appendix for all treatments (Appendix–Supplementary graphs and tables 2.3/2.4/3.3/ 3.4/4.3/4.4/5.3/5.4/6.3/ 6.4/7.3/7.4/8.3/8.4/9.3/9.4).

The WTP thresholds at which the therapy would be cost-effective was $0 for Remdesivir, $10,000 for Casirivimab-Imdevimab, $20,000 for Dexamethasone, $20,000 for Baricitinib-Remdesivir and $40,000 for Tocilizumab. The WTP threshold at which usual care would be cost effective was $20,000 for Lopinavir-Ritonavir, $10,000 for Interferon beta-1a, and $50,000 for Hydroxychloroquine. At lower WTP thresholds, these treatments would be decrementally cost-effective, as they would save costs through reduced long-term healthcare expenditures due to reduced survival.

Table 1

	Hydroxychloroquine	Remdesivir	Casirivimab-Imdevimab	Dexamethasone	Baricitinib-Remdesivir	Tocilizumab	Interferon-B1a	Lopinavir-Ritonavir
Is treatment cost-effective	No**	Yes*	Yes	Yes	Yes	Yes	No**	No**
Incremental costs ($)	-12227	-5	696	6856	10673	35849	-2538	-1404
Incremental QALYs	-0.263	0.252	0.171	0.614	0.775	0.882	-0.472	-0.091
ICER ($/QALY)	46427	n/a	4075	11169	13772	40633	5377	15418
Incremental net monetary benefit ($) (Thousand)	-14	25	16	55	67	52	-42	-8
Incremental net health benefit (QALY)	-0.141	0.252	0.164	0.545	0.668	0.524	-0.447	-0.077
EVPPI (Million)	375	127	0	0	0	1,3	0	0
Current patients (Thousand)	598	598	598	99	598	99	598	598
Future patients (Thousand)	220	220	220	36	220	36	220	220
Optimal strategy	OIR	Approve	Approve	Approve	Approve	Approve	Reject	Reject
Net value (Million)	198	20645	13389	7358	54642	7069	0	0

Table 1: Summary results from our analysis. Results shown are the mean results from the probabilistic analysis, calculated as the treatment arm versus the care-as-usual arm of each trial at a WTP of $100,000/QALY and the results of the value of information analysis. Yes = Treatment is dominant. Yes = Treatment is effective and ICER < WTP. No** = Treatment is cost-saving, but not enough that ICER > WTP (i.e., treatment is not decrementally cost-effective). n/a ICER not applicable because of dominance. OIR = Only in Research. Future/current patients are based on all expected hospitalized patients (Hydroxychloroquine, Remdesivir, Casirivimab-Imdevimab, Baricitinib-Remdesivir, Interferon beta-1a, Lopinavir-Ritonavir) or ICU patients only (Dexamethasone, Tocilizumab).*

Discussion

Summary of findings

Our results illustrate how VOI can inform policy and practice amidst a pandemic when considering whether to approve therapies, permit emergency use authorization, perform additional research or simply reject potential therapies in the treatment of hospitalized patients with COVID-19. As of November 2021, our results indicate that at a WTP of $100,000, treatment with Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Baricitinib-Remdesivir and Tocilizumab leads to positive mean incremental net benefit compared to care as usual, whereas treatment with Hydroxychloroquine, Lopinavir-Ritonavir and Interferon beta-1a does not. Additionally, our results suggest sufficient certainty that decisions about treating patients with Dexamethasone, Casirivimab-Imdevimab, Baricitinib-Remdesivir, Lopinavir-Ritonavir and Interferon beta-1a would not change with further RCTs. Further research could needlessly consume resources, expose trial participants to avoidable risks and preclude them from receiving alternative (potentially) effective treatments. For Remdesivir and Tocilizumab, the net value of further trials did not outweigh the cost of research, making immediate approval their optimal overall strategies. The net value of further trials for Hydroxychloroquine outweighed the cost of research and therefore the highest net value for this drug was found in the OIR strategy. However, this further research would be conducted to investigate decremental cost-effectiveness (saving costs due to reduced survival), the ethical implications of which should be considered (see Box 2).

Policy implications

As further trials unfold, the allocation of drugs in specific strategies should be considered as non-static. FDA Emergency Use Authorization has been granted for Remdesivir, Casirivimab-Imdevimab and Baricitinib-Remdesivir.[50,51] For hospitalized COVID-19 patients, the Infectious Diseases Society of America's (IDSA) guidelines recommend against the use of Hydroxychloroquine and Lopinavir-Ritonavir, and suggest the use of Tocilizumab, Remdesivir and Baricitinib-Remdesivir. Our findings support the IDSA's guidelines for all investigated treatments. A summary of policy implications for each of the included treatments is provided in Box 2.

Although cost-effectiveness is an important tool to inform drug approval policy, there is little consensus on the WTP threshold[74], let alone during a pandemic. A potential point of discussion is whether a WTP threshold should be higher in pandemic times due to the emergency status, or lower due to considerations of affordability when a large number of people are treated. In their assessment of the cost-effectiveness of Remdesivir, the Institute for Clinical and Economic Review applied a $50,000 WTP threshold, stating their belief that this threshold is more likely to be policy-relevant for consideration of treatments in public health emergencies.[75] A potential additional consideration when treatment is cost-effective but raises concerns on affordability is to consider the health opportunity costs of overall budget impact of the approval of new drugs.[76,77] The U.S. has no specified WTP threshold or a single, defined budget for healthcare spending.[74] Historically U.S. based cost-effectiveness studies have considered ICERs at thresholds ranging from $50,000-$300,000 per QALY.[43,74,77] Our assumption of a WTP of $100,000 follows recent criticisms in health economics research that a WTP of $50,000 would be relatively low on the basis of increases in healthcare spending and increased per capita annual income.[41–43,74] Still, in the Appendix, we provide the cost-effectiveness acceptability curves across a wide range of WTP thresholds for all investigated treatments to inform the decision-making process. These results show that at a WTP of $50,000, the conclusions of which drugs were cost-effective (Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Baricitinib-Remdesivir, and Tocilizumab) would remain the same.

Clinical practice

For clinicians, these findings have implications for both prescription practice and participation in further trials. Firstly, our evidence-based approach can identify treatments under investigation that may not only be ineffective, but also harmful, e.g., Hydroxychloroquine is still widely prescribed based on personal beliefs or experience.[78] With rapidly expanding evidence in the pandemic, it is paramount that clinical guidelines and practice are continuously updated to reverse

Box 2

Policy implications and discussion per treatment

Treatment	Overall strategy	Current status	Reflections and other considerations
Hydroxychloroquine	OIR (highest net value) Reject (ethical consideration)	The FDA initially granted it Emergency Use Authorization but later revoked that designation following further scientific data.[52] Nevertheless there are countries which continue to recommend its prescription[53,54] based on early studies that later received heavy criticism.[55] The WHO announced in June 2020 that the Hydroxychloroquine arm of the SOLIDARITY trial was stopped, based on findings from the RECOVERY and SOLIDARITY trials. At least [27] trials investigating Hydroxychloroquine have been prematurely terminated and 12 withdrawn prior to enrolment[3], and some countries have banned the drug in the treatment of COVID-19[56]. As of November 2021, still more than 100 trials were ongoing or planned to investigate (Hydroxy)chloroquine with perhaps over 100,000 participants involving all severity levels (including prophylaxis).[3]	Our findings support the conclusion that Hydroxychloroquine should be rejected. Yet they also suggest that there is expected positive net value in investigating Hydroxychloroquine in trials (OIR). However, the expected value from further trials stems from the drug being close to decrementally cost-effective[57]: it saves costs through reduced long-term healthcare expenditures due to decreased survival. On average, a healthcare cost of $12101 per year would be "saved" by pre-mature death.[37] Performing further trials to demonstrate cost-savings due to reduced survival raises ethical questions, and we do *not* advocate this strategy.
Lopinavir-Ritonavir	Reject	Treatment guidelines have recommended against the use of Lopinavir-Ritonavir[25], after the drug showed no effect in hospitalized patients. This recommendation is in line with the findings in our analysis. According to the NMA initiative, 1 trial was reported terminated, 1 withdrawn and 1 suspended.[3] However, 44 trials continue to recruit patients[3].	Our VOI found the highest net value in the rejection strategy. Conducting further trials may expose patients to unnecessary harms and prevents them from receiving potentially effective treatments.
Interferon beta-1a	Reject	For Interferon beta-1a some improvements of clinical aspects of COVID-19 have been identified[58], however when administered in later stages of infection the drug exacerbates the disease severity due to excessive inflammation and tissue damage.[58,59] 3 trials were registered as terminated, 1 as withdrawn, 1 as suspended, and 54 continue recruiting.[3]	Our VOI found the highest net value in the rejection strategy. Conducting further trials may expose patients to unnecessary harms and prevents them from receiving potentially effective treatments.
Remdesivir	Approve	In early findings of Remdesivir in the ACTT-I trial, policymakers and regulatory agencies concluded that given the urgent need for COVID-19 treatment, the 4-day reduction in recovery time was a satisfactory proxy of the drug's effectiveness. Therefore, the FDA issued Emergency Use Authorization conditional on further research investigating its impact on mortality.[60] There are 29 trials investigating Remdesivir still recruiting patients as of November 2021.[3] In September 2021 a fifth large RCT, the DisCoVeRy trial[61] was published. The results have not yet been integrated into the major meta-analyses that we used in our analysis, but this trial also demonstrated a small but statistically non-significant beneficial effect of Remdesivir on the secondary outcome of mortality. Gyselink et al. have suggested that differences in findings across trials may be partially explained by the different levels of use of systemic steroids.[62]	While we found persistent uncertainty surrounding the effectiveness of Remdesivir, our model found that the benefits of widespread immediate implementation outweighed the net value of further research. This conclusion is in contrast to the WHO's recommendation to not treat with Remdesivir and to continue to recruit patients. Although both our and the WHO panels' conclusions are based on the same meta-analysis[20], our findings differ because our model accounts for a lifetime horizon and the potential life years and QALYs lost, rather than statistical significance of treatment effect during the trial timeframe alone.[8] Further analyses that will include latest trials such as the DisCoVeRy trial[61], analyze specific subgroups, and that build on network meta analyses[63] may help to identify the drivers of the different results for Remdesivir trials over time.

Treatment	Overall strategy	Current status	Reflections and other considerations
Baricitinib-Remdesivir	Approve	Baricitinib is an FDA approved treatment for rheumatoid arthritis. In November 2020 an Emergency Use Authorization was issued for the use of Baricitinib in combination with Remdesivir for hospitalized COVID-19 patients.[64] This guidance is based on the ACTT-2 results.[23] However, the ACTT-2 trial has been criticized for its inability to evaluate the effect of Baricitinib in addition to corticosteroids.64 The NIH panel recommended the use of either Baricitinib or Tocilizumab in combination with Dexamethasone alone or Dexamethasone plus Remdesivir.[64] Their recommendations are based on the results of the COV-BARRIER study published in September 2021.[64] This study compares Baricitinib alone versus usual care (including corticosteroids and antivirals). As of November 2021, 11 trials are registered as recruiting patients.[3]	In our model, Baricitinib-Remdesivir is compared to Remdesivir as usual care, as per the ACTT-2 results.[23] This is made explicit, as all patients from the usual care group received Remdesivir, whereas in other trials some did and some did not. Incremental costs and acquired QALYs should therefore be interpreted as incremental to Remdesivir alone.
Tocilizumab	Approve	The treatment effect found in meta-analyses[65] for Tocilizumab relied heavily on the population that was treated. Tocilizumab's clinical implementation has been mainly based on the results from the REMAP-CAP[24] and RECOVERY[65] trials, and is recommended only for severe patients already receiving Dexamethasone.66 As of November 2021, 28 trials in the WHO register are recruiting patients.[3]	In our simulation, we use the treatment effect evidence from REMAP-CAP which was solely applied to ICU patients. This is in contrast to the RECOVERY[65] meta-analysis, which included patients ranging from non-severe (EMPACTA[67]) to severe, as this reflects clinical practice.

Treatment	Overall strategy	Current status	Reflections and other considerations
Dexamethasone	Approve	Dexamethasone did not require Emergency Use Authorization, as it is a drug currently in use for patients who require respiratory support.[68] In patients with severe infections, the drug can prevent lung injury caused by community-acquired pneumonia by suppressing exuberant systematic inflammation.[70] In non-severe COVID-19, the use of corticosteroids is not recommended.[71] There are 29 trials that continue to recruit patients to investigate Dexamethasone.[3]	Similarly to Tocilizumab, Dexamethasone[22] is only given to the ICU population in our simulation, in line with the treatment recommendations.[69]
Casirivimab-Imdevimab	Approve	As of September 2021, the WHO issued a conditional recommendation for the use of Casirivimab and Imdevimab for only patients with seronegative status.[72] This recommendation is based on the statistically significant reduction of mortality in the seronegative subgroup but not the seropositive subgroup. There are 10 trials recruiting patients to investigate Casirivimab-Imdevimab.	Our current analysis included the overall effect of both seropositive and seronegative patients since our underlying population and transition probabilities represented a mixed serostatus group. Performing subgroup analysis while applying the same transition probabilities to both groups is in our opinion not justified since it would bias the results.[73] Future analyses investigating subgroup-specific effects with varying control group characteristics should take serostatus into account which will require the collection of subgroup-specific prognostic data.

previously approved therapies when appropriate so that clinicians and patients can be informed about potential benefits and harms of treatments. Furthermore, VOI summarizes existing evidence and can inform clinicians about the potential added information value, costs of further research, and ethical implications that come with it. Performing VOI can aid decision-making for policymakers, researchers, clinicians and patients on whether to initiate or continue enrollment in drug trials, stop such trials or simply approve and implement.

Strengths and limitations

To the best of our knowledge, our study is the first to perform a VOI analysis for the treatment of COVID-19 patients. Whereas trials and meta-analyses often conclude further research is needed7, and where current guidelines are based on statistical significance in clinical evidence, VOI results are based on both uncertainty as well as the potential consequences of making decisions with and without further evidence. Our paper expands the potential impact of drug approval by investigating not only drug efficacy and effectiveness, but also the overall net benefits. Considering the unprecedented rollout of clinical trials investigating potential treatments for COVID-19, objective research prioritization seems paramount.[3,5] Our model can be updated with further evidence from trials and (cumulative) meta-analyses as they become available to continuously evaluate the optimal overall strategy as the pandemic evolves.

Input parameters were based on best-available evidence. Some parameters, such as the quality of life of COVID-19 patients in 5 years and the costs of research, had to be estimated based on previous studies considering other diseases. Where necessary, we chose conservative approaches to our model parameters and settings. For example, treatment effects were only applied for reported trial duration and not extrapolated beyond and wide distributions were chosen to represent large uncertainty. Additionally, we calculated the net value for the US population alone, whereas globally more patients would benefit from determining the optimal overall strategy.

Unidentified bias in studies may, however, have affected our results. Previous papers have discussed the potential disagreements between meta-analyses and large trials.[79,80] Single trials may not consider heterogeneity that is likely to exist across trials and centers. An advantage of single trials, however, is the more detailed group-specific information. In the case of Tocilizumab, we explicitly decided not to use the results of a meta-analysis due to the clinically relevant differences between hospital ward and ICU patients. Large network meta-analyses, such as the living WHO guideline on COVID-19 treatments,[69] do not currently provide sufficient details to enable distinctions in treatment effect and this needs to be considered in future analyses.

With respect to the methods, our paper builds on the work by McKenna and Claxton by calculating net value and considers not only the decision for conducting further research versus implementing treatment options[7,81–83], but also accounts for the effect of immediate implementation on patients that would have "missed out" on treatment whilst awaiting trial results.[83] The effect of the strategy decision on current patients is especially important given the pace of hospitalizations during the pandemic. The foregone benefit of delaying approval whilst awaiting further trial results could be exceptionally high. Whilst our paper focusses on COVID-19, creating an infrastructure to investigate rapidly emerging existing trials and value of additional trials in real-time using evidence synthesis and VOI models in potential future pandemics could form the basis for informing clinical practice, research and policy decisions going forward.

The results of any VOI depend on the underlying choices and assumptions.[84] Therefore, the limitations of our analysis need to be considered. We estimated the EVPPI with a linear-regression meta-model and EVSI with a Gaussian approximation approach as proposed by Jalal and Alarid-Escudero[44,45]. This is, however, one out of several existing estimation methods[44,45], which differ in approaches but none of which has shown computational or statistical

superiority[85]. The advantage of the chosen approximation method is the computational efficiency without introducing substantial bias46. A potential limitation of the linear-regression meta-model is that this normal approximation of the prior and pre-posterior distributions for parameters of interest in EVPPI and EVSI computations may introduce bias if severely non-normal.[46] As our analysis contains parameters with sufficiently large sample sizes to approach normal distributions as per the Central Limit Theorem, we did not consider this a cause for an introduction of bias in our analysis. Finally, the current EVSI estimation methods do not consider structural uncertainty[45], meaning that even if the true values for all input parameters are known, we are still not certain that the model reflects reality.

The key uncertainty investigated in this analysis is the estimated precision of the treatment effectiveness, as this uncertainty would be reduced by acquiring additional RCT data. Other uncertainties that could be addressed with further research include long-term morbidity and mortality, heterogeneity, adverse events of treatment, recovery time, quality of life after recovery from COVID-19, the number of hospitalizations, and costs. Furthermore, our models' findings are not appropriate for comparing investigated treatments to each other. Our analysis does not aim to prioritize drug treatment by comparing active treatments to each other but rather focuses on the research and approval health policy questions of each studied drug regimen. Through different mechanisms and when applied in different contexts, these treatments may be useful in sequence or in combination. A head-to-head comparison of treatments based on currently available evidence would require the assumption of independent effects and comparability of study populations, which would likely strongly bias the results. Key differences in considered trials include patient populations, e.g., only ICU patients for Dexamethasone and Tocilizumab, the duration for which the treatment effect is applied, and evolving usual care as the pandemic unfolded. For example, when Remdesivir and Dexamethasone became incorporated in usual care, patients in both treatment and control arms in subsequent trials received these drugs, and accordingly our model investigated the incremental effect of the newly introduced treatment. Future network meta-analyses that identify head-to-head treatment-specific effects could provide new input for the model to help distill comparative cost-effectiveness.

Our model is based on several assumptions. The unavailability of appropriate U.S. cohort data necessitated the assumption that the large cohort of UK patients was sufficiently representative. It is likely that care for COVID-19 patients has improved since May 2020. We additionally made assumptions on the utilities of patients recovered from the ICU and Hospital ward over their lifetime. Projections of hospitalized patients[47] may also be altered due to the rollout of vaccinations, new virus variants, (reversal of) lockdown measures, or the effect of treatment on COVID-19 transmission.

Our model did not account for age-, gender-, serostatus-, or comorbidity-specific treatment effects, and investigated reductions in mortality, but not severity of disease. Our analysis also assumed treatments were available to all patients in our simulated cohort. This assumption ignores a potential shortage of treatments in certain areas or to specific patient groups when there is a need for prioritization of resources. The analysis could, however, be repeated to investigate subgroup specific costs, effects and strategy recommendations for specific subgroups of interest.

Conclusion

Our study demonstrates that using Hydroxychloroquine only in research; approval and implementation of Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Baricitinib-Remdesivir and Tocilizumab; and the rejection of Lopinavir-Ritonavir and Interferon beta-1a provide the highest net value per November 2021. In the case of Hydroxychloroquine, this highest net value arises from decremental-cost-effectiveness, and for ethical reasons, we would not recommend hydroxychloroquine to be investigated further. Performing ongoing VOI analyses using updated research results during the pandemic can help define the optimal moment to implement emerging therapies and whether further clinical trials are justified.

Supplemental material

Key input parameters are provided in the manuscript and Appendix. Model code, datafiles containing all input data and distributions including supplementary files will be made publicly available via https://github.com/krijkamp/Emerging-Therapies-for-COVID-19.

Acknowledgement

The authors would like to thank the Society for Medical Decision Making COVID-19 Decision and the Gordon and Betty Moore Foundation for their support through the COVID-19 Decision Modeling Initiative. We additionally thank the Decision Analysis in R for Technologies in Health (DARTH) workgroup for providing the coding framework in R.

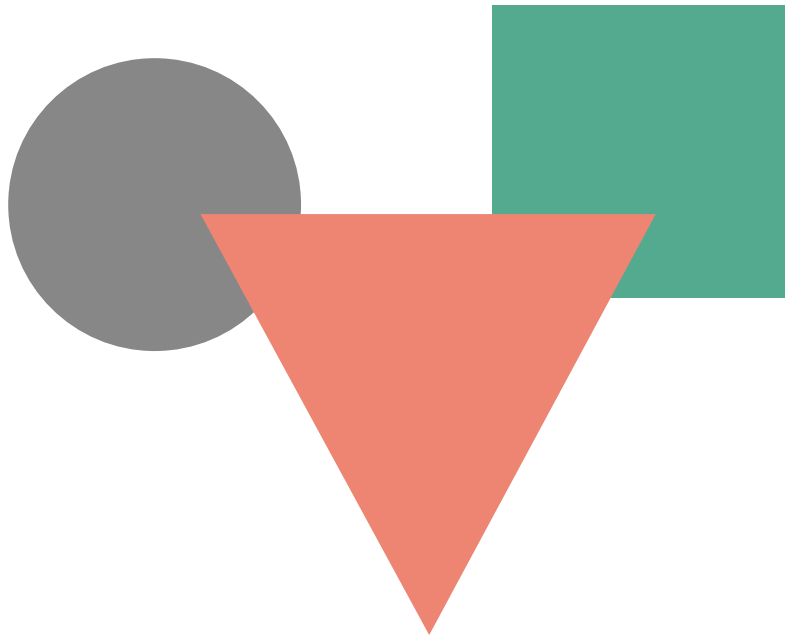
References

1. Ritchie H, Mathieu E, Rodés-Guirao L, Appel C, Giattino C, Ortiz-Ospina E, Hasell J, Macdonald B, Beltekian D, Roser M. Coronavirus Pandemic (COVID-19) Hospitalizations. *Our World Data.*, https://ourworldindata.org/covid-hospitalizations (2020, accessed January 15, 2021).
2. Johns Hopkins University. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. *2020*, https://github.com/CSSEGISandData/COVID-19. (accessed November 11, 2020).
3. Van Nguyen T Van, Ferrand G, Cohen-Boulakia S, Martinez R, Kapp P, Coquery E. The COVID-NMA initiative: A living mapping and living systematic review of Covid-19 trials. RCT studies on preventive measures and treatments for COVID-19. *COVID-NMA consortium*, https://covid-nma.com/dataviz/ (2021, accessed November 1, 2021).
4. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, Spackman E, Rothery C. A Comprehensive Algorithm for Approval of Health Technologies With, Without, or Only in Research: The Key Principles for Informing Coverage Decisions. *Value Health*. 2016; 19(6):885–891.
5. Mullard A. Flooded by the torrent: the COVID-19 drug pipeline. *Lancet*. 2020; 395(10232):1245–1246.
6. U.S. Food and Drug Administration (FDA). Emergency Use Authorization, https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization (accessed January 16, 2021).
7. Claxton K, Griffin S, Koffijberg H, McKenna C. How to estimate the health benefits of additional research and changing clinical practice. *Bmj*. 2015; 351:h5987.
8. Heath A, Myriam Hunink MG, Krijkamp E, Pechlivanoglou P. Prioritisation and design of clinical trials. *Eur J Epidemiol*. 2021; 36(11):1111–1121.
9. Keisler JM, Collier ZA, Chu E, Sinatra N, Linkov I. Value of information analysis: The state of application. *Environ Syst Decis*. 2014; 34(1):3–23.
10. Wilson ECF. A practical guide to value of information analysis. *Pharmacoeconomics*. 2015; 33(2):105–121.
11. Robinson M, Palmer S, Sculpher MJ, Philips Z, Ginnelly L, Bowens A, Golder S, Alfakih K, Bakhai A, Packham C, Cooper N, Abrams K, Eastwood A, Pearman A, Flather M, Gray D, Hall A. Cost-effectiveness of alternativestrategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling. *Heal Technol Assess*. 2005; 9(27):iii–iv, ix–xi, 1–158.
12. Iglesias CP, Claxton K. Comprehensive decision-analytic model and Bayesian value-of-information analysis. *Pharmacoeconomics*. 2006; 24(5):465–478.
13. Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess*. 2004; 8(27):iii–1.

14. Gurusamy K, Wilson E, Burroughs AK, Davidson BR. Intra-operative vs pre-operative endoscopic sphincterotomy in patients with gallbladder and common bile duct stones: cost-utility and value-of-information analysis. *Appl Health Econ Health Policy*. 2012; 10(1):15–29.
15. Wade R, Sideris E, Paton F, Rice S, Palmer S, Fox D, Woolacott N, Spackman E. Graduated compression stockings for the prevention of deep-vein thrombosis in postoperative surgical patients: a systematic review and economic model with a value of information analysis. *Health Technol Assess*. 2015; 19(98):1–220.
16. Heath A, Manolopoulou I, Baio G. A Review of Methods for Analysis of the Expected Value of Information. *Med Decis Making*. 2017; 37(7):747–758.
17. Tuffaha HW, Gordon LG, Scuffham PA. Value of information analysis in oncology: the value of evidence and evidence of value. *J Oncol Pract*. 2014; 10(2):e55–e62.
18. RECOVERY Collaborative Group. Effect of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med*. 2020; 383(21):2030–2040.
19. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S. Remdesivir for the Treatment of Covid-19—Preliminary Report. *N Engl J Med*. 2020; (383):992–994.
20. Consortium WHOST. Repurposed antiviral drugs for COVID-19—interim WHO SOLIDARITY trial results. *N Engl J Med*.
21. Horby PW, Mafham M, Peto L, Campbell M, Pessoa-Amorim G, Spata E, Staplin N, Emberson JR, Prudon B, Hine P. Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *medRxiv*. Epub ahead of print 2021. DOI: https://doi.org/10.1101/2021.06.15.21258542.
22. RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. *N Engl J Med*. 2021; 384(8):693–704.
23. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, Marconi VC, Ruiz-Palacios GM, Hsieh L, Kline S. Baricitinib plus remdesivir for hospitalized adults with COVID-19. *N Engl J Med*. 2020; 384(9):795–807.
24. Investigators R-C. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N Engl J Med*. 2021; 384(16):1491–1502.
25. Horby PW, Mafham M, Bell JL, Linsell L, Staplin N, Emberson J, Palfreeman A, Raw J, Elmahi E, Prudon B. Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet*. 2020; 396(10259):1345–1352.
26. Hunink MGM, Weinstein MC, Wittenberg E, Drummond MF, Pliskin JS, Wong JB, Glasziou PP. *Decision making in health and medicine: integrating evidence and values. 2nd Edition*. Chapter 9 and 12. Cambridge University Press, 2014.
27. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, Holden KA, Read JM, Dondelinger F, Carson G, Merson L, Lee J, Plotkin D, Sigfrid L, Halpin S, Jackson C, Gamble C, Horby PW, Nguyen-Van-Tam JS, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ*. 2020; 369:m1985.
28. Wilmoth JR S V. The Human Mortality Database. *University of California, Berkeley (US), and Max Planck Institute for Demographic Research (Germany)*, http://www.mortality.org/ (accessed May 29, 2020).
29. R Core Team. R: A language and Environment for Statistical Computing, https://www.r-project.org (2013).
30. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics*. 2019; 37(11):1329–1339.
31. Alarid-Escudero F, Krijkamp EM, Enns EA, Yang A, Hunink MGM, Pechlivanoglou P, Jalal H. A Tutorial on Time-Dependent Cohort State-Transition Models in R using a Cost-Effectiveness Analysis Example, https://arxiv.org/abs/2108.13552 (2021, accessed October 28, 2021).
32. 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. *Eur J Heal Econ*. 2013; 14(3):367–372.
33. Briggs AH, Weinstein MC, Fenwick EAL, 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.
34. Ward ZJ. Amua: An open source modeling framework. *https://github.com/zward/Amua*.
35. US Department of Labor Statistics. Bureau of labor statistics. *Consumer Price Index*, https://www.bls.gov/cpi/ (2020, accessed November 1, 2020).
36. Zorginstituut Nederland. Zorginstituut adviseert tijdelijk ruimere vergoeding paramedische herstelzorg voor patiënten met ernstige COVID-19. *zorginstituutnederland.nl*, https://www.zorginstituutnederland.nl/actueel/nieuws/2020/07/13/zorginstituut-adviseert-tijdelijk-ruimere-vergoeding-paramedische-herstelzorg-voor-patienten-met-ernstige-covid-19 (2020, accessed July 13, 2020).
37. Agency for Healthcare Research and Quality. Use, expenditures and population. US Medical Expenditure Panel Survey 1996-2017, meps.ahrq.gov (accessed June 14, 2021).
38. Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated costs of pivotal trials for novel therapeutic agents Approved by the US food and drug administration, 2015-2016. *JAMA Intern Med*. 2018; 178(11):1451–1457.
39. Emanuel EJ, Schnipper LE, Kamin DY, Levinson J, Lichter AS. The costs of conducting clinical research. *J Clin Oncol*. 2003; 21(22):4145–4150.
40. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, Kuntz KM, Meltzer DO, Owens DK, Prosser LA, Salomon JA, Sculpher MJ, Trikalinos TA, Russell LB, Siegel JE, Ganiats TG. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. JAMA - *Journal of the American Medical Association* 2016; 316(10):1093–1103.
41. Weinstein MC. How much are Americans willing to pay for a quality-adjusted life year? *Med Care*. 2008; 46(4):343–345.

42. Braithwaite RS, Meltzer DO, King Jr JT, Leslie D, Roberts MS. What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? *Med Care*. 2008; 46(4):349–356.
43. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness—the curious resilience of the $50,000-per-QALY threshold. *N Engl J Med*. 2014; 371(9):796–797.
44. Jalal H, Alarid-Escudero F. A Gaussian Approximation Approach for Value of Information Analysis. *Med Decis Making*. 2018; 38(2):174–188.
45. Kunst N, Wilson ECF, Glynn D, Alarid-Escudero F, Baio G, Brennan A, Fairley M, Goldhaber-Fiebert JD, Jackson C, Jalal H. Computing the expected value of sample information efficiently: practical guidance and recommendations for four model-based methods. *Value Health*. 2020; 23(6):734–742.
46. Jalal H, Goldhaber-Fiebert JD, Kuntz KM. Computing expected value of partial sample information from probabilistic sensitivity analysis using linear regression metamodeling. *Med Decis Making*. 2015; 35(5):584–595.
47. Institute of Health Metrics and Evaluation. COVID-19 Projections - USA. *https://covid19.healthdata.org/*.
48. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: Combined and uneven development. *Med Decis Making*. 2008; 28(1):21–32.
49. Griffin SC, Claxton KP, Palmer SJ, Schulpher MJ. Dangerous Omissions: The consequences of ignoring decision uncertainty. *Health Econ*. 2011; 20(2):212–224.
50. U.S. Food and Drug Administration. Coronavirus (COVID-19) Update: FDA Authorizes Monoclonal Antibodies for Treatment of COVID-19. *www.fda.gov*.
51. U.S. Food and Drug Administration. Emergency Use Authorization, https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization (2020, accessed January 16, 2021).
52. U.S. Food and Drug Administration (FDA). Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine. *FDA NEWS RELEASE*, https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and (2020, accessed January 16, 2021).
53. Turone F. Ruling gives green light for controversial COVID-19 therapy. *Nature News*, December 2020.
54. Brito J, Darlington S. How Brazil gambled on unproven drugs to fight Covid-19. *CNN*, 2021.
55. Manivannan E, Karthikeyan C, Moorthy NS, Chaturvedi SC. The Rise and Fall of Chloroquine/Hydroxychloroquine as Compassionate Therapy of COVID-19. *Front Pharmacol*. 2021; 12:1057.
56. National Institutes of Health. NIH halts clinical trial of hydroxychloroquine, https://www.nih.gov/news-events/newsreleases/ nih-halts-clinical-trial-hydroxychloroquine (2020, accessed November 26, 2021).
57. Nelson AL, Cohen JT, Greenberg D, Kent DM. Much cheaper, almost as good: decrementally cost-effective medical innovation. *Ann Intern Med*. 2009; 151(9):662–667.
58. Sodeifian F, Nikfarjam M, Kian N, Mohamed K, Rezaei N. The role of type I interferon in the treatment of COVID-19. *J Med Virol*. 2021; 94(1):63–81.
59. Sa Ribero M, Jouvenet N, Dreux M, Nisole S. Interplay between SARS-CoV-2 and the type I interferon response. *PLoS Pathog*. 2020; 16(7):e1008737.
60. Kmietowicz Z. Covid-19: Selected NHS patients will be treated with remdesivir. *BMJ*. 2020; 369:m2097.
61. Ader F, Bouscambert-Duchamp M, Hites M, Peiffer-Smadja N, Poissy J, Belhadi D, Diallo A, Lê M-P, Peytavin G, Staub T. Remdesivir plus standard of care versus standard of care alone for the treatment of patients admitted to hospital with COVID-19 (DisCoVeRy): a phase 3, randomised, controlled, open-label trial. *Lancet Infect Dis*. 2021; 22(2):209–221.
62. Gyselinck I, Janssens W. Remdesivir, on the road to DisCoVeRy. *Lancet Infect Dis*. 2021; 22(1):153–155.
63. Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Undertaking network meta-analyses. *Cochrane Handb Syst Rev Interv*. 2019; :285–320.
64. National Institutes of Health. The COVID-19 Treatment Guidelines Panel's Statement on Baricitinib for the Treatment of Adults with COVID-19.
65. Abani O, Abbas A, Abbas F, Abbas M, Abbasi S, Abbass H, Abbott A, Abdallah N, Abdelaziz A, Abdelfattah M. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet*. 2021; 397(10285):1637–1645.
66. Lee CK, Linder JA, Gates KL. Management of severe covid-19: progress and promise. *Br Med J*.; 373(n1147).
67. Salama C, Han J, Yau L, Reiss WG, Kramer B, Neidhart JD, Criner GJ, Kaplan-Lewis E, Baden R, Pandit L. Tocilizumab in patients hospitalized with Covid-19 pneumonia. *N Engl J Med*.
68. Tanzi MG. FDA has authorized these therapies to manage patients with COVID-19. *Pharm Today*. 2021; 27(2):18–20.
69. Zhou W, Liu Y, Tian D, Wang C, Wang S, Cheng J, Hu M, Fang M, Gao Y. Potential benefits of precise corticosteroids therapy for severe 2019-nCoV pneumonia. *Signal Transduct Target Ther*. 2020; 5(1):1–3.
70. Raju R, Prajith V, Biatris PS. Therapeutic role of corticosteroids in COVID-19: a systematic review of registered clinical trials. *Futur J Pharm Sci*. 2021; 7(1):1–18.
71. Lamontagne F, Agoritsas T, Macdonald H, Leo Y-S, Diaz J, Agarwal A, Appiah JA, Arabi Y, Blumberg L, Calfee CS. A living WHO guideline on drugs for covid-19. *bmj*. 2020; 370:m3379.
72. World Health Organization. On new recommendation for treatment of COVID-19 patients: WHO calls for equitable access to casirivimab and imdevimab for COVID-19. *www.who.int/news*.
73. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Med Decis Making*. 2002; 22(3):218–227.

74. Vanness DJ, Lomas J, Ahn H. A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Ann Intern Med*. 2021; 174(1):25–32.
75. Institute for Clinical and Economic Review. ICER Provides First Update to Pricing Models for Remdesivir as a Treatment for COVID-19, https://icer.org/news-insights/press-releases/updated_icer-covid_models_june_24/.
76. Lomas J, Claxton K, Martin S, Soares M. Resolving the "cost-effective but unaffordable" paradox: estimating the health opportunity costs of nonmarginal budget impacts. *Value Health*. 2018; 21(3):266–275.
77. Grosse SD. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. *Expert Rev Pharmacoecon Outcomes Res*. 2008; 8(2):165–178.
78. Ebm C, Carfagna F, Edwards S, Mantovani A, Cecconi M. Potential harm caused by physicians'a-priori beliefs in the clinical effectiveness of hydroxychloroquine and its impact on clinical and economic outcome–A simulation approach. *J Crit Care*. 2021; 62:138–144.
79. Cappelleri JC, John PA, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J. Large trials vs meta-analysis of smaller trials: how do their results compare? *Jama*. 1996; 276(16):1332–1338.
80. Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials*. 1997; 18(6):568–579.
81. Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med Decis Making*. 2008; 28(3):300–305.
82. Conti S, Claxton K. Dimensions of design space: a decision-theoretic approach to optimal research design. *Med Decis Making*. 2009; 29(6):643–660.
83. McKenna C, Claxton K. Addressing adoption and research design decisions simultaneously: the role of value of sample information analysis. *Med Decis Making*. 2011; 31(6):853–865.
84. Koffijberg H, Rothery C, Chalkidou K, Grutters J. Value of information choices that influence estimates: a systematic review of prevailing considerations. *Med Decis Making*. 2018; 38(7):888–900.
85. Heath A, Kunst N, Jackson C, Strong M, Alarid-Escudero F, Goldhaber-Fiebert JD, Baio G, Menzies NA, Jalal H. Calculating the Expected Value of Sample Information in Practice: Considerations from 3 Case Studies. *Med Decis Making*. 2020; 40(3):314–326.



11

GENERAL DISCUSSION

“The more we discover, the more we know there is to discover.”

Nederburg Wines, South Africa

GENERAL DISCUSSION

The overall aim of this thesis was to expand knowledge on the use of open-source modeling in decision analysis in healthcare and to demonstrate how open-source decision modeling can contribute to resource prioritization in healthcare. This final chapter describes and discusses the main findings of each part of this thesis and provides recommendations to further optimize the use of open-source modeling and research optimization methods in healthcare.

Part I
Open-source decision modeling in healthcare

In this part we established a practical approach to develop (advanced) decision models in the open-source statistical software R. First, we introduced a high-level framework that can be used to construct R-based decision models (**Chapter 2**). This was followed by an evaluation of the use of R for decision analysis in healthcare and included a comprehensive description of R functionalities useful for decision analysis (**Chapter 3**). Part I concluded with a set of methodological papers that outline how decision-analytic models can be constructed in R (**Chapters 4-7**).

Main findings and interpretation

Open-source software in decision analysis in healthcare

In the first two chapters, we found that the popularity of R is increasing in the field of decision analysis in healthcare (**Chapter 3**) and explained why R has the potential to facilitate model transparency, reproducibility and shareability. We also outlined that

the fragmentation of R material (**Chapter 3**) and lack of guidance about best coding practices (**Chapter 2**) make it hard to reach the full potential of R for healthcare decision analysis. In **Chapter 2** we described that decision-analytic models can be conceptualized as structures with the same high-level model components, regardless of the specific structure of the model. These common model components were further discussed within the DARTH framework; a generic methodological approach to construct open-source decision-analytic models. Adoption of this framework will facilitate the sharing, review and readability of decision-analytic models implemented in R as model users will become familiar with and follow the component structure. In **Chapter 3**, we provided an overview of R packages that can be useful in various stages of model design. An R package is a collection of functions and/or datasets developed by the community. This overview can be used as a baseline reference point for modelers interested in navigating through the large number of R packages. Our DARTH framework together with the comprehensive overview of useful R packages contributed to fill the gap in guidance of the use of R in decision analysis in healthcare. However, this is just a start. Modeling-specific guidance is needed as well. Therefore, the aim of **Chapters 4-7** was to develop a generic programming structure and code that can be applied to a wide range of decision problems. This has resulted in a set of papers that provide guidance to decision modelers in the medical field to build both cohort and individual based state-transition models in R.

Cohort state-transition models in R

In **Chapter 4** we described the development of time-*independent* cohort state-transition models in R using three core components: (1) the state vector, (2) the cohort trace matrix and (3) the transition probability matrix. In **Chapter 5** we proposed that to incorporate simulation-time dependency (e.g., age dependent mortality rates), the transition probability matrix should become a three-dimensional structure with the third dimension being time. This third dimension represents the transition probability matrix at different timepoints during the simulation. To account for time dependency on state residence (e.g., the risk of side effects from a drug given to patients diminishes over time), the model structure should be expanded with extra health states, so-called tunnel states. In both chapters, we first use mathematical notation to explain the model components needed for the methods we developed. This is followed by the R code applied to our Sick-Sicker case example. This case example is a comprehensive cost-effectiveness analysis with four strategies along with a probabilistic sensitivity analysis. For the model structure, we use base R packages because this makes it easier for decision modelers to understand the concept and structure of a cohort state-transition model built in R. In addition, this structure is flexible and avoids any package-specific structural requirements. We demonstrated visualization of the cost-effectiveness analysis results via the R package `dampack`.

In **Chapter 6** we introduced the dynamics-array approach, which we developed to structure the output of cohort state-transition models. The dynamics-array approach can compute and store both the state occupancy and transition dynamics of a cohort state-transition model. This is something the traditional cohort trace is not capable of. While not possible with the cohort trace, the dynamics-array approach facilitates applying state and transition rewards and makes it possible to obtain several epidemiological measures. From this structure, it is still possible to obtain the conventional cohort trace. Therefore, the dynamics-array approach is an efficient, simple, and convenient method of summarizing the model dynamics of cohort state-transition models.

Individual-based state-transition models in R

In **Chapter 7** we described a general coding structure that can be used to build an individual-based state-transition model, also called a microsimulation model, in R. We first provided a generic algorithm to explain the required steps to construct a microsimulation model in a programming language. Next, we demonstrated how to construct a model in R with code applied to our Sick-Sicker case example. The general code structure makes use of four key functions: (1) a

function to estimate probabilities, (2) a function to estimate costs, (3) a function to estimate the health effects and (4) the overall microsimulation function that executes the first three functions throughout the simulation time. This general code structure can be used to build flexible microsimulation models in R.

Current innovation practices and implications

We developed an extensive open-source modeling framework and coding structure that can be applied to a wide variety of decision-analytic modeling problems. These tailored tools for decision-analysis in healthcare is inspired by reproducible research guidelines from other fields[1,2] and general coding recommendations[3–5]. Our tailored work narrows the gap of the lack of guidance of the use of R in decision analysis in healthcare.

This thesis work contributes to the foundations to develop open, transparent and reproducible decision-analysis in healthcare via R-based modeling. In addition, it contributes to the education of the next generation decision-analytic modelers. Familiarity with R-based decision-analytic models among various stakeholders in medical decision making has the potential to improve decision making in healthcare along several stages, including, but not limited to, (collaborative) model development, validation, reporting, reviewing, testing and adapting.

Recent publications indicate the usefulness of our methods, as other researchers cite our work as the basis for their modeling methods. More specifically, our cohort-state transition approach is used for a Norwegian meningocele vaccination evaluation.[6] The microsimulation work was used to project the current HPV vaccination uptake on the incidence of oropharyngeal cancer in the USA[7] and to assess the health effect of a menthol cigarette ban in New York city.[8] The microsimulation method also made it possible to efficiently simulate multiple treatment sequences for multiple sclerosis.[9] The visualization methods in our `dampack` packages was of benefit to a cost-effectiveness analysis of cancer treatment in Malawi.[10] This wider implementation of our work and reflection on the usefulness of our methods in different settings will help to further develop modeling methods for R-based decision models that can facilitate efficient and streamlined (international) collaboration between researchers, manufacturers in healthcare and HTA-agencies.

This thesis work can also be used as the reference point for standardization of R-based modeling guidelines. These guidelines would be relevant to develop R-based decision-analytic models for submission to HTA-agencies and academic journals. Key in the development of these guidelines should be to balance standardization and flexibility. Homogeneity would be preferred for the back-bone of the models (e.g., suggestion for code organizing, naming conventions, reporting standards, approved packages, unit testing and checklist to identify and avoid errors), while the flexibility of R should facilitate the implementation of novel methodologies to best inform the decision-making process.

Alongside our work, there have been other initiatives to explore which role open-source software can play to improve transparency in decision-analytic models for healthcare. Some of these initiatives will be discussed below.

In 2019, the journal PharmacoEconomics had a special themed issue titled “Improving Transparency in Decision models”.[11] This issue stressed that model transparency is a highly desirable goal for the field. Nevertheless, this desired transparency comes with many-sided challenges, concerns, costs and requires efforts for all stakeholders involved in the decision-making process. They concluded that increasing transparency will remain a relevant topic for debate in the field.

The journal Medical Decision Making contributes to the required efforts that are needed in the field through a call for tutorials (educational material about the use of cutting edge methodological

topics) and explainer articles (best-practice research examples).[12] Such materials, when focused on R-based decision-analytic modes like ours (**Chapters 2-7**), could bridge the gap between novel decision-analytic methods and the lack of implementation guidance. Academic journals can also contribute to improved model transparency by for example requesting model source code to be shared in the supplementary material of papers, or through external sources like GitHub. This way the materials will be freely available to others. Other journals make all their publications completely open-access.

The work presented in the first part of this thesis originates from the collaborative effort with the DARTH (Decision Analysis in R for Technologies in Health) workgroup.[13] The DARTH workgroup consists of an international group of researchers with a passion for open-source decision-analytic models. DARTH contributes to the field via educational material in the form of papers (**Chapters 2-7**) as well as via courses. Similar initiatives that contribute to the transition in the field are the R for Health Technology Assessment group (R-HTA)[14] and Open-Source Value Project[15,16] from the Innovation and Value Initiative. Applied research examples come from groups at BresMed[17,18] and The University of Sheffield[19–21]. Furthermore the Dark Peak Analytics group provides consultancy for health economic modeling and data-analyses using R.[22] In addition, many valuable discussions are held in interest groups and seminars at the ISPOR[23] and SMDM[24] conferences.

HTA-agencies are also exploring the use of R for model-based economic evaluations. With DARTH, I contributed to the R skills of HTA-agencies worldwide by providing workshops to some of their employees. Familiarity with R for decision modeling within their teams is required to review R-based model submission. Some HTA-agencies are now exploring when and how model-based economic evaluations in R will be accepted and how their submission guidelines can facilitate an efficient submission and review process.

Together, all of these examples emphasize the relevance of the debate of increased transparency and the use of open-source modeling in decision-analysis in healthcare.

Drawbacks of R in healthcare decision analysis

Although this thesis research has been able to demonstrate that R can be used to construct decision-analytic models for a wide range of decision problems in an efficient and transparent manner, we also identified some drawbacks of R for decision modeling. Some challenges are discussed in the section below.

Visualization

The first major drawback is that R lacks built-in functionalities to visualize the model structure. A visual representation of the model is desirable because it helps the decision analyst in the model development process as well as in the process to communicate the model to stakeholders. Visualization is also very valuable when teaching how to use R for decision modeling in healthcare. Coding in R can seem quite abstract. Modelers new to R for decision modeling can benefit a lot from a visualization tool because coding a model is much easier after you first envisioned what the model should look like and how it should perform. A clear graphical representation also makes it easier to ask others for help and guidance.

A schematic representation of the sequence of important events and combined data sources that form conditional probabilities can be especially helpful. Events in a decision tree and the cycle tree of a state-transition model can be intuitively shown using a tree diagram structure. These tree structures in turn add face validity and can prevent calculation errors when estimating the likelihood of a branch outcome of a decision tree or the transition probabilities in a state-transition model.

Luckily, there are several initiatives to facilitate the visualization of the model in R. One such initiative is the R package `Arvore` (**Chapter 3**), and our methods provide code to create state-transition diagrams

in R (**Chapters 4 & 5**). More recently, OpenTree[25] and Amua[26] became available. These software packages fill the visualization gap of R. OpenTree can be used to construct decision trees and cohort state-transition models. These files can be saved locally and after finalizing the model OpenTree can convert the model structure into R code. For the model analysis, R is required to run the exported code from OpenTree. Like OpenTree, Amua can export the model to R. In addition, Amua can also export the model to C++, Java and Python code. Although exporting the model to programming languages is possible, none of these languages is required per se. Amua works as an installed program on your computer and can do the analysis as well as the production of graphs within the software environment itself. The trade-off is that the exported code of Amua is harder to read compared to the OpenTree code. Regardless, both of these programs facilitate the need for visualization of decision-analytic models while still being able to benefit from the strengths of R and facilitate the use of state-of-the-art decision-analytic methods.

Steep learning curve

A second limitation of R is its steep learning curve. There are several factors that contribute to this feature. One of them is that R is open-source but this is also one of the favorable characteristics of R. However, open-source means that there is no owner of the software that is responsible for all documentation or for providing a help desk you can call in case you need help. As a consequence, and therefore a second reason for the steep learning curve, is that the available documentation is scattered. There is plenty of information available via books[27], platforms (e.g., R-blog[28], Reddit[29] and Stack Overflow[30]), educational companies (e.g., DataCamp[31]), or in packages in R's main repository (CRAN)[32] or GitHub[33]. However, decision modelers new to R need to become familiar with these sources and need to be able to search for and select what they need. For this reason, academic consortia provide papers, conference workshops[34,35] and code examples[13,14,36]. Examples of these consortia are R-HTA[14], The Collaborative Network for Value of Information (ConVOI)[36] and the DARTH workgroup[13].

Another reason is that R extensions for model communication have an even steeper learning curve compared to base R. Examples of very useful extensions are R-Shiny[37], an R package to build interactive web apps, R-Markdown[38], a notebook interface to integrate code and text to make fully reproducible reports, or R-Bookdown[39], which facilitates producing reproducible books or long articles. These extensions require additional knowledge from the users.

Finally, the generally experienced steep learning curve of R may cause additional difficulties for decision modelers who have not had previous exposure to programming languages. Healthcare decision modelers are likely to have a background in health sciences, epidemiology, medicine or (health) economics. These degree programs often provide limited tools and experience with programming languages compared to for example statistics, engineering or computer sciences.

Package and version dependency

Our developed code guidance (**Chapters 4-7**) uses base R only. For the visualization of results, we make use of the package `dampack`. R-packages are very useful in the development of decision-analytic models, for example for survival analysis or value of information methods. Packages reduce the need to duplicate functionalities and can standardize code. They do, however, also have limitations.

A third limitation of the use of R for decision-analytic models could be dependency on (specific versions of) packages. Package dependency means that in order to run the code users need to have these packages activated in their R environment. Version dependency means that a specific version of that package is required. Different versions of packages exist, because functionalities are constantly improved. For CRAN packages, R packages hosted on the formal platform from R, extensive testing to check for potential code incompatibility is done before a new version is released. However, this is never a guarantee for compatibility. In order to guarantee reproducibility of the decision-analytic model results, modelers should always report the package version they used.

The development of a CRAN package requires extensive testing and documentation. Therefore, the state-of-the-art decision-analytic method packages are more likely to be first available through GitHub. GitHub is a website where modelers can store code and helps to track and control changes to this code.[33] GitHub does not require developers to perform compatibility tests. Therefore, there is a risk of broken code after an update. Since GitHub is a version control system it is always possible to go back to the version as was used in the model. However, restoring that specific package version requires extra work, and likely frustration, for code users.

Furthermore, some functionalities of packages could have programming errors ('bugs'). The use of packages in research projects increases the risk that in case of a bug in a package, this bug influences the results of multiple modes. Identifying bugs in R code is likely to be easier given the open-source character of the code. Nevertheless, the risk of bugs stresses the need for extensive testing of package code before widespread use. In the long run, the use of properly tested packages reduces the risk of mistakes. Continuously writing new code is more prone to mistakes than the formally tested functionalities of a package.

Lack of guidelines of submissions of R-based decision-analytic models

A forth limitation is the lack of guidelines for the submission of R-based decision-analytic models to HTA-agencies. Although some HTA-agencies accept R-based models nowadays, they provide no or limited guidance on how to structure those models. Therefore, it is not clear for decision-modelers that aim to inform HTA-agencies how to use the currently available guidance on R-based decision-analytic models for their analyses.

Benefits of R in healthcare decision analysis

Throughout the development of this thesis material, we have found several benefits of R relevant for healthcare decision analysis. These benefits are highlighted in this section.

- R can be used to make decision-analytic methods more computationally efficient. One example of this is the vectorized sampling functionality we developed for microsimulation models (**Chapter 7**). Another example is the use of parallel processing for uncertainty analysis (applied in **Chapter 5 & 10**). During parallel processing, multiple processors of a computer are used to separately perform simulation tasks, which can be applied in an uncertainty analysis since the simulations do not depend on each other.
- R allows for integration of statistical analysis with the development of decision-analytic models. Statistical survival analyses are frequently required to populate decision-analytic models with accurate estimates about the survival benefits of an intervention.[40,41] These survival analyses require specific methods.[42] R has many functionalities that support using statistical methods[43,44] frequently applied in decision analysis (e.g., meta-analysis, calibration, regression analysis or direct analysis from public datasets)[44]. Since R can do both, statistical analysis and decision-analysis, the model process can be less fragmented between different software programs.
- R removes methodological constraints that specialized software or spreadsheets have. The flexibility of R makes it possible to incorporate all methods necessary to reflect the decision problem accurately. For example, with R modelers can select the most appropriate calibration method[45] for their analysis. This in contrast to TreeAge which supports calibration via three calibration algorithms (e.g., Bound Optimization by Quadratic Approximation, Nelder-Mead, and By Deterministic Parameter Ranges).[46] Although these algorithms are likely to be useful for a wide variety of decision problems, the software dictates the available methods and the way they are used.
- R can efficiently perform the recently developed approximation methods for value of information (VOI) analysis, as demonstrated by The Collaborative Network for Value of Information (ConVOI).[47,48] Since 2019, ConVOI aims to remove the barriers to use VOI in

practice by providing training and guidance to use VOI. Based on their developed methods, they created several packages to reduce the coding burden for other users (e.g., EVSI[49], BCEA[50], VOI[51]). They also developed several web-tools (e.g., The Sheffield Accelerated Value of Information (SAVI) app[20,52], Visualization for EVSI[53] and BCEAweb[54]) and their GitHub repository provides the source code of the methods they explain in their papers.[55]

- R can integrate all steps required to build a comprehensive decision-analytic model including statistical analysis, model development, calibration, sensitivity- and value of information analysis. R can even be used for reporting via the packages R-Shiny[56], R-Markdown[38] and R-Bookdown[39]. These packages can best used within R Studio, a free integrated development environment for R.
    - R-based decision-analytic models can be used to build interactive web-applications via R-Shiny. The original data and source code from the decision-analytic model can form the basis of the Shiny web-application. Therefore, there is no need to recode the functionalities of the model. They just need to be integrated into the Shiny structure. These web-applications are useful to communicate results with decision makers. In addition, these web-applications can be used to allow users to change model input parameters and see the effect of these changes to the results, without the need to modify the source code in the R environment.
- The source code of R-based decision-analytic models can be shared as reproducible scripts. This facilitates extensive model review. These source files can also be structured as R-packages, which makes distributing the model even easier. Making model source files available can avoid expensive duplication efforts.
- In R script for unit testing can be written. Unit testing is a way to create functions that test if a specific unit of the code behaves as expected. Unit testing increases confidence in the model code and prevents unintended incompatibilities between functions. Tests of the code are useful during the development phase to check whether changes to existing functions have potential downstream issues. In addition, it facilitates model sharing because these tests will identify whether code modifications to one part of the code requires adjustments in other parts as well.
- The open-source nature of R enables collaboration. This is strongly affected by the positive attitude towards model transparency and code sharing by decision-modelers working in R. This also contributes to model expansion or further methods development. For example, Smith and Schneider used the same DARTH Sick-Sicker case example to write a tutorial about using R-Shiny for decision modeling.[19] This way new components are added to existing code which makes the field move forward rapidly.

Limitations of methodological R papers

One of the limitations of the methodological coding papers (**Chapters 4-7**) could be that they were written in collaboration with the DARTH workgroup of which the members are mainly active in academia. Therefore, the methods are based on code mainly applied and tested in the academic setting via courses and (own) research projects. Wider adoption of our methods and reflection on the usefulness of our methods in different settings will contribute to identify points of improvement.

Second, the overview of useful packages (**Chapter 3**) is not comprehensive anymore. More packages should be in this list these days, of which some examples are: `hesim`[57], `EVSI`[49], `voi`[51], `HEdtree`[58] and packages developed by the DARTH group: `OpenTree`[25], `DARTHpack`[59], `darthtools`[60], and `dampack`[61]. The list should also be updated with an overview of useful web-applications (e.g., `BCEAweb`[54] and `SAVI`[20]) and maybe extended with GitHub repositories from researchers that contribute a lot to the field. The fact that our work is rapidly outdated is an indication of the fast transition and developments in the field. We anticipated this at the time of publication and created a Wikiversity page[62] that ideally would be updated on an ongoing and regular basis. The fact that it has not been updated over the previous years, stresses the point

that the incentive to continuously update published work is lacking and more value is considered in the development of new material. Similarly, the code from our individual-based state-transition work (**Chapter 7**) provided on GitHub has not been updated according to our own naming convention (**Chapter 2**). As a consequence, the newest material is still fragmented and requires additional efforts from decision-modelers to collect the newest available material.

Third, the provided materials require time investment of the users. The materials are well structured from an educational point of view to guide readers step-by-step through the methods as well as the code, but reading the work once is not enough. It requires practice to become familiar with the material. The best way to gain experience with the code is by applying the general coding concepts to a new case example. However, it is likely that questions come up or modelers might get stuck in the application of some model-specific code. There is currently no formal platform to get help with R-based decision-analytic models. To support modelers, the DARTH workgroup continues to be approachable and we try to help where and when possible. In addition, we host courses where participants can get help with their code. Ideally, there will be a new subcommunity within the R community where healthcare modelers find guidance with R related questions for their R-based decision-analytic models.

Challenges of open-source modeling for healthcare decision-analysis

Although open-source modeling has several benefits for healthcare decision-analysis there are also several challenges. In this section some of these challenges are highlighted.

- There are currently limited incentives for decision modelers to create transparent open-source models. It requires more time and effort to build self-explanatory open-source decision-analytic models compared to making a model intended to only inform decision making or publication of the results. This extra effort is currently not requested nor rewarded and should therefore come from the intrinsic motivation of the modelers and stakeholders working on the project.
- Reviewers lack the time or knowledge to fully review open-source decision models. As a result, access to the model source code could unintentionally imply high model credibility. Some form of a quality mark or test status that indicates if and how the work is reviewed might be appropriate to add.[63]
- There are concerns about intellectual property protection related to open-source model sharing.[24,64,65] It requires years to develop good decision-analytic models. Some organizations want to benefit from their efforts and use the models for other indications, which makes them hesitant to share the full models. Sampson and Wrightson describe model registration as a solution for this issue.[66] A model registration platform could be a place to provide detailed information about the model by the developer and could facilitate a formal citation system that can be used to give credit to the developer. Model owners can also provide licensing details on this platform. This may help to protect intellectual property because it can be used to detect plagiarism or use without appropriate citation of the original model.
- There are also concerns about model misuse.[24] This can be misuse in the context of new research but also public misuse via misinterpretation or blaming researchers or their institutions (e.g., during heated times like the COVID-19 pandemic).
- Another concern relates to data sharing issues. Sensitive data (e.g., patient-level data or proprietary pricing information) used to inform the decision-analytic models cannot be publicly shared, although they are key to get to generate model results. Aggregate data (a summary of the original data) or a pseudo dataset (artificially generated dataset with comparable information) can be used to demonstrate the functionalities of the model to increase model credibility without the need to share the original dataset.
- It is not clear what the future role of other open-source software could be for healthcare decision analysis. R is currently discussed and used a lot as open-source modeling platform but some healthcare decision modelers work with Julia[67], Python[68] or C++[69].

- Transparency is desirable, but it is debatable what the level of transparency should be. Should everyone have access to the source code? Should it be available upon request or do only reviewers have access? In 2012 the ISPOR-SMDM Task Force recommended that every model should have "technical documentation" that should be made available openly or under agreements that protect intellectual property, at the discretion of the modelers.[70] Regarding open-source modeling this recommendation should be revisited to add the role of model source code.
- Most HTA-agencies currently do not allow R-based model submission. This delays the uptake of R by healthcare manufacturers because they often develop and evaluate new interventions for multiple countries at the same time. For these evaluations they prefer to use the same model structure. Acceptance of R-based models by HTA-agencies worldwide is essential in order to achieve a global uptake of R for decision-analysis in healthcare.

Despite these challenges, many would agree that increased model transparency is desired and that the discussion about the role of open-source modeling for decision analysis in healthcare should continue. Every discussion brings us a little closer to transparent decision-analytic models and together these challenges can be turned into solutions.

Part II
Resource prioritization in healthcare

Main findings and interpretation

In the second part of this thesis, we demonstrated how open-source decision modeling methods can contribute to resource prioritization in healthcare.

In **Chapter 8**, we showed how a decision-analytic model can be used to prioritize scarce surgical capacity. We used a cohort state-transition model to quantify expected health loss due to surgical delay. This outcome measure can be used as an urgency measure to prioritize surgeries in case of scarce surgical capacity. The use of an open-source decision analytic model for priority setting is more evidence-based and more transparent than the alternative strategy of triaging based on expert opinion.

Next, we demonstrated how state-of-the-art decision modeling methods can inform clinicians and policy makers about the optimal moment for approval of an intervention or funding of future research. In **Chapter 9**, we outlined an iterative research cycle, which we call the *value-driven* approach. In this chapter we described how decision-analytic models can be used to optimize healthcare and research resources. Although this *value-driven* approach requires more data and time to design a study, we concluded that this approach, if correctly applied, can reduce research waste by guiding efforts towards the performance of only research that is valuable to society.

In **Chapter 10**, we used the *value-driven* approach to evaluate the optimal overall implementation strategy (e.g., treatment implementation or further research) for eight interventions (Baricitinib-Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Hydroxychloroquine, Interferon beta-1a, Lopinavir-Ritonavir, Remdesivir and Tocilizumab) for hospitalized COVID-19 patients. Based on the results as of November 2021, we concluded that the optimal strategy for Dexamethasone, Remdesivir, Casirivimab-Imdevimab, Baricitinib-Remdesivir and Tocilizumab was to implement the intervention in clinical practice, while Interferon beta-1a and Lopinavir-Ritonavir should be rejected without further research. Although our analysis showed that the optimal implementation strategy for Hydroxychloroquine is use of the drug in a research setting only, we do not recommend this implementation strategy. Hydroxychloroquine is likely to be decrementally cost-effective due to cost savings, which are mainly due to reduced long-term healthcare expenditures because less patients survive. We do not consider it an ethical approach to perform additional research to explore the likely decremental cost-effectiveness of Hydroxychloroquine. Our study

showed how the *value-driven* approach combined with s*tate-of-the-art* decision analytic modeling can inform policy makers and clinicians about drug or research approval during a pandemic.

Resource prioritization challenges

Although decision-analytic models have many benefits for the prioritization of resources, their methods also have limitations. Some of them are discussed in the below.

Decision-analytic models require evidence. This required evidence can be limited, biased or lacking. For example, the estimates about the survival probabilities could not be derived from high-quality evidence (e.g., biased observational data or unethical to estimate relative treatment effect of surgery versus no surgery, **Chapter 8**) and limited evidence about the health utility value of hospitalized and recovered COVID-19 patients was available (**Chapter 10**). In situations of limited evidence, especially during a pandemic, modelers are recommended to use the best alternative available evidence.[71] In our analyses, relative treatment effect of surgical procedure from the literature were used to reconstruct survival without surgery (**Chapter 8**), and proxy utility values based on utility estimates from comparable health conditions were used (**Chapter 10**). The potential bias in the model results due to the use of less robust evidence should be reflected in the decision uncertainty as much as possible. In addition, modelers should clearly explain the implications and generalizability of the results with respect to the data limitations.

Evidence can also become fast outdated. Healthcare system conditions or our understanding of the clinical disease pathways might change (e.g., fluctuation of incidence, supply shortages, new intervention alternatives to consider). This is especially the case in a novel pandemic situation. In the light of changing situations and new evidence, the models should be updated and decision makers should consider re-evaluating the decision. This requires organizing decision-analysis in a "living" approach where modelers continuously keep track of the available resources and evidence and update the model results accordingly. Especially, transparent open-source models are suitable for this "living" approach as they are available for modifications or they can quickly obtain new results if new evidence becomes available. However, this "living" approach is time-consuming and sometimes the incentive or funding to do this work is missing. International collaborations between groups working on the same topics, like the COVID-19 best practice guidance effort[71], can help to efficiently inform each other about the best available evidence.

This "living" evaluation of decision models and consequently transparent dynamic decision making about resource allocation requires a strong integration of the models with daily practice. This requires changes to the current infrastructure of the way decisions are made. For example, implementation of value-based surgery planning (**Chapter 8**) or bringing together organizations responsible for decisions about intervention implementation with those responsible for research funding (**Chapter 10**). An ongoing close collaboration between modelers and decision makers is required to guarantee correct interpretation of the model results given the chosen methods and their limitations.

It is time-consuming to develop decision-analytic models for resource prioritization. Under rapidly changing conditions, like a pandemic, we are often too early until we are too late. In the first phase it is too early to have a reliable model that reflects the decision problem mainly because of lack of data. When a reliable model has been developed, the information comes too late as (irreversible) resource allocation decisions (e.g., funding of large clinical trials or implementation of an intervention) have either already been made based on less explicit weighting of factors that contribute to the decision or the results have become irrelevant due to changed conditions (e.g., lower incidence of disease, new treatment alternatives).

Finally, personalized decision-analytic models for resource prioritization are methodologically more challenging to develop. Subgroup analyses (e.g., incorporating co-morbidities of patient's

(**Chapter 8**) or separately analyzing antibody sero-positive and sero-negative COVID-19 patients (**Chapter 10**)) can help identify for which subgroups the intervention or additional research is worthwhile. However, these subgroups analyses are only valuable when the model itself and its comparators are correctly selected and informed by sufficiently detailed data.

Recommendations and future research direction

Based on this thesis and the research experience gained during my doctoral work, I would make the following recommendations. These recommendations are organized in three overall categories: (1) education and communication, (2) collaboration and (3) standardization.

Education and communication

- The current and next generation of healthcare decision modelers should be educated in state-of-the-art open-source modeling methods as well as data science skills.
  - Health science programs should consider a data science and open-source programming course in their curriculum.
  - More tutorials in scientific journals and books about open-source decision modeling in healthcare should be written.
- Clinicians should be educated about decision-analytic methods and how they can be used for recourse prioritization.
- The field should organize international reproducible research initiatives (i.e., events where open-source models are critically reviewed with the aim to provide constructive feedback and learn from each other).
- Decision scientist should put more effort in the communication and interpretation of model-based scientific results, especially focused on decision uncertainty.
- Clear communication about the implications of decision uncertainty and why intervention implementation and research funding decisions can change over time to the public and journalists could avoid harm and mistrust in science.

Collaboration

- A concerted international effort between modelers, manufactures, decision makers, scientific journals and HTA-agencies is needed to achieve a wider implementation of R-based decision-analytic models.
  - Round tables with all stakeholders should be organized in which everyone can communicate their needs and concerns.
  - It is needed to explore incentives for modelers to put effort in model sharing.
- Decision sciences is often already an interdisciplinary collaboration of statisticians, clinicians, health economists and decision scientists, but should be expanded with experts about data- and computer sciences to reach the full potential of open-source modeling for the field.
- All stakeholders in healthcare should work towards a more resilient and circular healthcare system where healthcare and research resource allocation methods are integrated to guide dynamic decision making under fast fluctuating scenarios.

Standardization

- HTA-agencies worldwide should develop guidelines for R-based model submission. Ideally, this development will be a collaborative effort. Their recommendations regarding good coding practices and reporting should be streamlined as much as possible. Consistency in the recommendations will facilitate increased transparency, efficiency and international collaboration of decision modelers.
- HTA-agencies globally together with the ISPOR-SMDM Task Force should develop R packages targeted to HTA-agencies which includes tested and proven R functions.
  - In these packages, unit testing (i.e., check if code works as expected) should have an important role. Unit testing facilitates model review, further model development and re-usability because it helps to identify potential code issues or downstream consequences because of added code functionalities.

- It should be explored if mandatory model registration can avoid duplication efforts, support modeling sharing and encourage further development of existing models.
  - Such registry might facilitate international research collaborations, model comparisons and model re-reuse, while protecting intellectual property.
- The review process of model-based submission to academic journals should be restructured.
  - Academic journals should explore the possibility of a special peer-review process with dedicated code-review for model-based publications.
  - Academic journals should make it easier to update publications after new evidence became available.
- The value-driven approach for trial design should be promoted in research proposals to guide the choice of study type, inclusion/exclusion criteria, sample size, allocation ratio, and criteria for adaptive designs.

Overall conclusion

This thesis aimed to expand knowledge on the use of open-source modeling for decision analysis in healthcare and to demonstrate how open-source decision models can contribute to resource prioritization in healthcare.

The work presented in this thesis, together with initiatives presented in the section about "Current innovation practices and implications", flatten the steep learning curve of R for decision models and empowers them to construct decision models in open-source software. In addition, the work contributes to the discussion on how open-source models can increase transparency and improve resource prioritization in the field.

R-based decision-analytic models have the potential to facilitate model transparency, reproducibility, and shareability in healthcare. Nevertheless, R is too flexible to implement for large scale cost-effectiveness and health technology assessment submissions without clear modeling guidelines and proper coding education to decision modelers in healthcare. Without clear guidance and training, the field risks getting lost in code that is neither well documented nor systematically organized in a comprehensible and shareable manner.

The methodological coding framework presented in this thesis contributes to the foundation for open, transparent, and reproducible decision modeling in healthcare. Specific modification to the described framework presented are likely to be required in order to facilitate the construction and submission of open-source model-based evaluation to either scientific journals and/or HTA agencies. The required modifications depend on the aim of the models (methodological vs. application) and why these models are evaluated (scientific contribution, collection of new evidence or health (policy) decisions). Nevertheless, these modifications should be based on good coding practice guidelines, fit the national modeling guidelines and are, ideally, as much as possible globally streamlined.

Finally, we demonstrated that R-based decision-analytic models can be used for a dynamic value-based allocation of scarce resources, both healthcare resources and research resources, which is especially relevant in times of a pandemic.

References

1. Cooper N, Hsing P-Y, editors. *A Guide to Reproducible Code in Ecology and Evolution*. London, UK: British Ecology Society, 2017.
2. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. *PLoS Computational* Biology; 9(10). Epub ahead of print October 2013. DOI: 10.1371/journal.pcbi.1003285.
3. Wickham H. The tidyverse style guide, https://style.tidyverse.org (2019, accessed July 19, 2019).
4. Google. Google's R Style Guide, https://google.github.io/styleguide/Rguide.html (2019, accessed July 24, 2021).
5. Martin RC. *Clean code: a handbook of agile software craftsmanship*. Boston, MA: Pearson Education, 2009.
6. Watle SV, Næss LM, Tunheim G, Caugant DA, Wisløff T. Cost-effectiveness of meningococcal vaccination of Norwegian teenagers with a quadrivalent ACWY conjugate vaccine. *Hum Vaccines Immunother*. 2021; 17(8):2777–2787.
7. Damgacioglu H, Sonawane K, Chhatwal J, Lairson DR, Clifford GM, Giuliano AR. Long-term impact of HPV vaccination and COVID-19 pandemic on oropharyngeal cancer incidence and burden among men in the USA: A modeling study. *Lancet Reg Heal*. 2021; Dec 15:100143.
8. Li Y, Sisti J, Flórez KR, Albrecht SS, Viswanath A, Davila M, Cantrell J, Brahmbhatt D, Thompson AB, Jasek J, Chambers EC. Assessing the Health and Economic Impact of a Potential Menthol Cigarette Ban in New York City: a Modeling Study. *J Urban Heal*. 2021; 98(6):742–751.
9. Huygens S, Versteegh M. Modeling the Cost-Utility of Treatment Sequences for Multiple Sclerosis. *Value Health*. 2021; 24(11):1612–1619.
10. Painschab MS, Kohler R, Kimani S, Mhango W, Kaimila B, Zuze T, Mithi V, Kasonkanji E, Mumba N, Nyasosela R, Wheeler S, Gopal S. Comparison of best supportive care, CHOP, or R-CHOP for treatment of diffuse large B-cell lymphoma in Malawi: a cost-effectiveness analysis. *Lancet Glob Heal*. 2021; 9(9):e1305–e1313.
11. Tappenden P, Caro JJ. Improving Transparency in Decision Models: Current Issues and Potential Solutions. *Pharmacoeconomics*. 2019; 37(11):1303–1304.
12. Zikmund-Fisher BJ, Editors on behalf of the. A Call for Explainer/Tutorial Articles and Changes to Manuscript Submission and Review at MDM and MDM P&P. *MDM Policy Pract*. 2020; 5(2):1–3.
13. Decision Analysis in R for Technologies in Health (DARTH) workgroup. Decision Analysis in R for Technologies in Health, http://darthworkgroup.com (2019).
14. R for Health Technology Assessment, https://r-hta.org (accessed December 16, 2021).
15. Innovation and Value initiative. Open-source Value Project: Transforming the Process of Value Assessment. 2019, http://www.thevalueinitiative.org/open-source-value-project/ (accessed December 12, 2021).
16. Jansen JP, Incerti D, Linthicum MT. Developing Open-Source Models for the US Health System: Practical Experiences and Challenges to Date with the Open-Source Value Project. *Pharmacoeconomics*.; 2019;37(11):1313-1320.
17. Hart R, Burns D, Ramaekers B, Ren S, Gladwell D, Sullivan W, Davison N, Saunders O, Sly I, Cain T, Lee D. R and Shiny for Cost-Effectiveness Analyses: Why and When? A Hypothetical Case Study. *Pharmacoeconomics*. 2020; 38(7):765–776.
18. BresMed, https://www.bresmed.com (accessed December 12, 2021).
19. Smith R, Schneider P. Making health economic models Shiny: A tutorial. *Wellcome Open Res*. 2020; 5:1–26.
20. Strong M, Oakley J, Brennan A. Sheffield Accelerated Value of Information, https://savi.shef.ac.uk/SAVI/ (accessed December 21, 2021).
21. The University of Sheffield, https://www.sheffield.ac.uk (accessed December 12, 2021).
22. Sheffield E. Dark Peak Analytics Limited, https://darkpeakanalytics.com (2020, accessed December 19, 2021).
23. Pouwels X, Hart R, Thom H. Open Source Health Economics and Outcomes Research: Why and How to Do It? How to Connect with Users? In: 2021-11, ISPOR Europe 2021, *Copenhagen, Denmark,* https://www.ispor.org/heor-resources/presentations-database/presentation/euro2021-3380/13343 (2021, accessed December 10, 2021).
24. Alarid-Escudero F, Garbayo L. Open Science and Model Transparency in the context of COVID-19 Pandemics. In: *43rd Annual North American Meeting Symposia*, *Society of Medical Decision Making*, https://smdm.org/meeting/page/43rd-annual-north-american-meeting-symposia/43rd-annual-north-american-meeting (2021).
25. Decision Analysis in R for Technologies in Health (DARTH) workgroup. OpenTree: An Open-Source Visual Tool for Decision Modeling and Cost-Effectiveness Analysis in R, https://github.com/DARTH-git/OpenTree (2021, accessed November 8, 2021).
26. Ward ZJ. Amua: An open source modeling framework. *2019*, https://github.com/zward/Amua. https://github.com/zward/Amua (accessed November 8, 2021).
27. R project. Books related to R, https://www.r-project.org/doc/bib/R-books.html (accessed November 16, 2021).
28. R project. R blog, https://developer.r-project.org/Blog/public/ (accessed December 5, 2021).
29. Reddit, https://www.reddit.com (accessed December 5, 2021).
30. Stack Overflow, https://stackoverflow.com (2008, accessed December 5, 2021).
31. DataCamp, https://www.datacamp.com (accessed November 12, 2021).
32. The Comprehensive R Archive Network, https://cran.r-project.org.

33. Github. GitHub, https://github.com/ (2020).
34. Societey of Medical Decision Making, https://smdm.org (accessed December 28, 2021).
35. ISPOR—The Professional Society for Health Economics and Outcomes Research, https://www.ispor.org (accessed December 28, 2021).
36. Collaborative Network for Value of Information, https://www.convoi-group.org (2019, accessed December 21, 2021).
37. Chang W, Cheng J, JJ Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B. shiny: Web Application Framework for R, https://cran.r-project.org/package=shiny (2021).
38. Allaire J, Horner J, Xie Y, Marti V, Porte N. markdown: Render Markdown with the C Library "Sundown," https://cran.r-project.org/package=markdown (2019).
39. Xie Y. *Bookdown: Authoring Books with R Markdown*. Boca Raton, FL: CRC Press, 2016.
40. Latimer NR. Survival analysis for economic evaluations alongside clinical trials - Extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013; 33(6):743–754.
41. Williams C, Lewsey JD, Mackay DF, Briggs AH. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. *Med Decis Making*. 2017; 37(4):427–439.
42. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, Sharples L. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. *Med Decis Making*. 2017; 37(4):377–390.
43. Jackson CH. flexsurv: A Platform for Parametric Survival Modeling.*J Stat Softw*. 2016; 70(May):i08.
44. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making*. 2017; 37(7):735–746.
45. Vanni T, Karnon J, Madan J. Calibrating models in economic evaluation. *Pharmacoeconomics*. 2011; 29(1):35–49.
46. Software T. *TreeAge Pro Healthcare* 2020 User ' s Manual. 2020.
47. Heath A, Kunst N, Jackson C, Strong M, Alarid-Escudero F, Goldhaber-Fiebert JD, Baio G, Menzies NA, Jalal H. Calculating the Expected Value of Sample Information in Practice: Considerations from 3 Case Studies. *Med Decis Making*. 2020; 40(3):314–326.
48. Kunst NR, Wilson E, Alarid-Escudero F, Baio G, Brennan A, Fairley M, Glynn D, Goldhaber-Fiebert JD, Jackson C, Jalal H, Menzies NA, Strong M, Thom H, Heath A. Computing the Expected Value of Sample Information Efficiently: Expertise and Skills Required for Four Model-Based Methods. *Value Health*. 2020; 20(6):18–26.
49. Heath A, Baio G. EVSI, https://github.com/annaheath/EVSI (accessed December 23, 2021).
50. Baio G. BCEA: A R Package to Perform Bayesian Cost-Effectiveness Analysis. In: *Value in Health*, p. A550.
51. Jackson C. voi: a generic package to calculate the expected value of information, https://github.com/chjackson/voi (accessed December 23, 2021).
52. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast, Nonparametric Regression-Based Method. *Med Decis Making*. 2015; 35(5):570–583.
53. Heath A, Baio G. Visualisations for the EVSI, https://egon.stats.ucl.ac.uk/projects/EVSI/Example1/ (accessed August 2, 2019).
54. Baio G, Hadjipanayiotou P, Berardi A, Heath A. BCEA web, https://egon.stats.ucl.ac.uk/projects/BCEAweb/ (accessed December 23, 2021).
55. ConVol Group GitHub repository, https://github.com/convoigroup (accessed December 23, 2021).
56. Beeley C. *Web Application Development with R using Shiny*. Birmingham, UK: Packt Publishing Ltd, 2013. Epub ahead of print 2013. DOI: 10.1017/CBO9781107415324.004.
57. Incerti D, Jansen JP. hesim: Health Economic Simulation Modeling and Decision Analysis. *arXiv*; 2102.09437, https://cran.r-project.org/web/packages/hesim/index.html (2021).
58. Dodd P. HEdtree: Utilities for decision tree like models in health economics, https://github.com/petedodd/HEdtree (accessed December 23, 2022).
59. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics*. 2019; 37(11):1329–1339.
60. Decision Analysis in R for Technologies in Health (DARTH) workgroup. darthtools: an R package that contains tools developed by the Decision Analysis in R for Technologies in Health (DARTH) workgroup to construct model-based cost-effectiveness analysis in R. *2020*, https://github.com/DARTH-git/darthtools.
61. Enns EA, Knowlton G, Easterly C, Kao S-YZ, Krijkamp E, Pechlivanoglou P, Jalal H, Alarid-Escudero F. 'Dampack': A Flexible R Package for Analyzing and Visualizing Cost-Effectiveness Analysis Results, https://github.com/DARTH-git/dampack (2019).
62. Decision Analysis in R for Technologies in Health (DARTH) workgroup. Wikiversity: R in Health Decision Sciences. *2018*, https://en.wikiversity.org/wiki/R_in_Health_Decision_Sciences (accessed December 19, 2021).

63. Daniel R, Zhang J, Farewell D. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical J*. 2021; 63(3):528–557.
64. Rutter CMC, Zaslavsky AAAM, Feuer EEJ. Dynamic microsimulation models for health outcomes: a review. *Med Decis Making*. 2011; 31(1):10–18.
65. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, Hawkins N, Langham S, Marshall D, Sadatsafavi M, Sullivan W, Wilson ECF, Wrightson T. Transparency in Decision Modelling: What, Why, Who and How? *Pharmacoeconomics*. 2019; 37(11):1355–1369.
66. Sampson CJ, Wrightson T. Model Registration : A Call to Action. *PharmacoEconomics - Open*. 2017; 1(2):73–77.
67. Lee TY. Integrating Julia and R (Studio) for microsimulation modeling: challenges and rewards, https://r-hta.org/events/workshop/2021/ (2021).
68. Chartash D, Bennett W. An introduction to (PYTHON) programming for the medical decision scientist. In: *37th Annual Meeting of the Society for Medical Decision Making*, https://smdm.confex.com/smdm/2015mo/webprogram/Session2046.html (2015).
69. University of Minnesota and Massachusetts General Hospital. Simulation Model of Colorectal Cancer (SimCRC), https://resources.cisnet.cancer.gov/registry/packages/simcrc-minnesota/#details.
70. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, on Behalf of the ISPOR-SMDM Modeling Good Research Practices Task Force. 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.
71. Elvidge J, Summerfield A, Knies S, Németh B, Kaló Z, Goettsch W, Dawoud D, On behalf of the COVID-19 HTA best-practice guidance development group. *Best-practice guidance for the health technology assessment of diagnostics and treatments for COVID-19*. 2021. Epub ahead of print 2021. DOI: 10.5281/zenodo.5530468.

“Not that the story need be long, but it will take a long while to make it short.”

Henry David Thoreau it in a letter of 1857

SUMMARY

The overall aim of this thesis is to expand knowledge on the use of open-source modeling in decision analysis and to demonstrate how open-source decision modeling can contribute to resource prioritization in healthcare. In this thesis, the open-source programming language R is used to demonstrate our conceptual framework and methods. Nevertheless, our methodological modeling approaches can be generalized to other open-source programming languages. To demonstrate how open-source decision analysis models can be used for resource prioritization we use two COVID-19 related decision problems, of which one includes a value of information analysis.

General introduction

The first chapter gives an overview of the background and describes the aim of the thesis.

Decision makers in healthcare (e.g., healthcare professionals, policy makers, patients and their family, healthcare organizations, health technology assessment (HTA) agencies) frequently face decisions on the allocation of scarce resources under conditions of uncertainty. Healthcare resources include all materials, personnel, facilities, funds, and anything else that is needed to provide healthcare services. Regardless of decision uncertainty, a decision needs to be made about the use of these resources for the implementation of an intervention or the collection of further evidence. This requires decision makers in healthcare to evaluate the consequences of their decisions regarding both health outcomes and resource utilization. To maximize the value

that can be achieved from the available resources, healthcare decisions can be supported by decision-analytic models.

Decision-analytic models can evaluate both clinical and economic aspects of a decision problem. They are considered the most comprehensive and suitable approach to inform decision-making in healthcare. However, how to build a reliable decision-analytic model is not always straightforward. Throughout the development process, modelers and stakeholders have to make many choices and assumptions regarding both model structure and model inputs. This requires good collaboration between decision makers and modelers to develop a decision-analytic model that appropriately informs the decision.

To make a well-informed decision, decision makers need to gain trust in the accuracy and absence of bias in the model results. Decision makers can gain this trust via review of the model structure, inputs, and assumptions. This requires that the decision-analytic models are available and transparent enough for adequate review. Unfortunately, the model source files from specialized commercial software or spreadsheet calculators are not self-explanatory and can be cumbersome to review. These models are therefore sometimes considered as a "black-box". Other limitations of these software are the need of a paid license and insufficient facilitation of all state-of-the art decision analysis methods, like calibration and value of information analysis.

Scripts from open-source programming languages, such as R or Python, have the potential to overcome many of the limitations that specialized software programs and spreadsheets have. However, no clear guidance for the implementation of decision-analytic models in these open-source languages exists. To realize the full potential of open-source programming languages for healthcare decision making, this thesis explores how to implement decision-analytic models in R.

In the second part of this thesis, we demonstrate how open-source decision-analytic models can be used to optimize the allocation of scarce resources in healthcare. Decision-analytic models evaluate the expected outcomes of different strategies. These outcomes can be health effects, costs or both. Either by themselves, or as an aggregated measure in which costs and effects are combined, these model outcomes can be used to identify strategies that justify the consumed resources (i.e., how we can get the "best value for money").

Similar to healthcare resources, resources to fund research are limited. Resource funds can be defined as all facilities, materials, human resources, study participants, funding and all other sources that are required to perform a (clinical) study. Cutting-edge methods in decision analysis can be used to prioritize both healthcare and research resources. These methods are called value of information (VOI) analyses. VOI analyses are used to quantify the monetary value of reducing decision uncertainty via the collection of additional evidence. We say that the additional evidence has value when it has the potential to reduce the costs associated with making the wrong decision.

Historically, these calculations were very computationally demanding or not possible to perform in commonly used software. However, there have been several efforts to make these calculations more efficient and to develop approximation methods. These new methods in combination with the use of open-source decision-analytic models facilitate an efficient implementation of VOI analyses. Despite the improvements in methods, decision makers are still not familiar with these methods leading to the limited application of VOI in the design of clinical studies. Therefore, we aim to communicate the potential value of VOI methods to a broad audience and present it within a COVID-19 related decision problem.

In summary, this thesis aims to expand knowledge on the use of open-source modeling in decision analysis and to demonstrate how open-source decision models can contribute to resource prioritization in healthcare.

Part I: Open-source decision modeling in healthcare

In **Chapter 2** we present a generic methodological framework to construct open-source decision-analytic models. This methodological framework is a collaboration of the "Decision Analysis in R for Technologies in Health" (DARTH) workgroup. The DARTH framework provides guidance for the implementation of decision-analytic models in R and separates decision-analytic models into five components that are common across traditional model-based decision analyses. The first four components, (1) model inputs, (2) decision model implementation, (3) model calibration and (4) model validation, form the model development phase. The fifth component, (5) analysis, involves the analysis of the final model (developed in components 1-4). The analysis is used to answer the policy or research question(s) of interest given current information, to assess decision uncertainty, and/or to determine the value of future research through VOI analysis. The flexibility of the model framework was confirmed via our case study example, called the Sick-Sicker model. All material is provided on GitHub for easy adaptation to other applications.

In **Chapter 3** we conducted a literature review to investigate how frequently R is used in studies related to decision making in healthcare compared to other commonly used software. Based on our results, R appears to be increasing in popularity. In the scientific journals we reviewed, the proportion of studies that use R increased by nearly 50% between 2010 and 2015. In this chapter, we also discuss R packages and user-defined functions that are relevant to decision analysis in healthcare. These packages and functions are grouped by stages of model development and analysis for easy navigation.

Building on the findings from Chapter 2 and 3, the aim of the **Chapters 4–7** was to establish a practical approach to develop and implement R-based decision-analytic models in healthcare. The flexibility of R implies that decision-analytic models can be implemented in multiple ways depending on the experience and programming style of the modeler and the research question that needs to be addressed. Therefore, we aimed to provide a generic approach with a reasonable balance between clarity and efficiency.

In **Chapter 4** we show how to build time-independent cohort state-transition models in R. Time-independent means that the transition probability among health states or rewards associated with being in a health state (e.g., cost or health outcomes) remains constant over time. First, we conceptualize how a time-independent cohort state-transition model can be implemented in a programming language. We use math equations to describe the three core components of a cohort state-transition model: (1) the state vector, which describes the distribution of the cohort among the different health states at one point in time, (2) the cohort trace matrix, which describes the distribution of the cohort across health states over the model time horizon, and (3) the transition probability matrix, which specifies the transition probabilities of moving from one of the model health states to any other health state in any given point in time. In addition, we describe state rewards, discounting and within-cycle correction. Second, we provide a template to implement a time-independent cohort-state transition model in R. The case example in the template, the Sick-Sicker model, is a comprehensive cost-effectiveness analysis with four strategies along with a probabilistic sensitivity analysis and visualization of the model results.

Time-independent models are more straightforward to implement than time-dependent ones. However, most problems in healthcare are best modeled with some form of time-dependency. In **Chapter 5**, we demonstrate how time-dependency can be incorporated in R-based cohort state-transition models.

We distinguish between two types of time-dependency that require different approaches: time since the start of the simulation (simulation-time dependency) and time spent in a health state (state-residence dependency). Simulation-time dependency affects parameters that vary with time for the entire cohort in the same way during the simulation (e.g., age-specific background mortality over time as the cohort ages). We show that simulation-time dependency can be incorporated into a model by defining a transition probability matrix as a function of time. State-residence dependency captures the time dependence of events that members of the cohort could

experience at different times (e.g., the risk of side effects from a drug given to patients diminishes over time). However, throughout the simulation, healthy individuals may experience disease onset at different times. This requires keeping track of the disease onset time of the cohort members. We implement state-residence dependency by expanding the model state space to include disease states that encode the time since an event has occurred. We used the same Sick-Sicker model as in Chapter 4 to showcase our developed methods. In addition, we demonstrate how to generate epidemiological and economic measures, account for transition rewards, and conduct a cost-effectiveness analysis and a corresponding probabilistic sensitivity analysis. Together, Chapters 4 and 5, provide a novel methodological approach that guide step-by-step implementation of cohort state-transition models in R.

In **Chapter 6**, we introduce the dynamics-array approach. The dynamics-array approach can compute and store cohort state-transition model outcomes that capture both state occupancy and transition dynamics. Transition dynamics are not documented in the traditionally generated cohort trace. Because all model dynamics are documented with the dynamics-array approach, this method allows us to apply both state and transition rewards. State rewards are rewards associated with spending time in a particular health state, while transition rewards are rewards that are only associated when a specific transition from one health state to another occurs. In addition, the dynamics-array approach facilitates computation of epidemiological measures (e.g., survival, prevalence and lifetime risk of events). These epidemiological outcomes can be helpful for calibration and validation. In a simulation study, we demonstrate that the dynamics-array approach is faster than the traditional cohort-trace approach, but requires slightly more memory (1 - 3.5 times) for models with an average number of health states (<50 health states). This relative memory difference is unlikely to become a modeling limitation. Therefore, structuring the output of cohort state-transition models using the dynamics-array approach is an advantageous method of summarizing the model dynamics.

In **Chapter 7** we provide a novel methodological structure to construct an individual-based state-transition model, often referred to as a microsimulation model, in R. We first provide a conceptual algorithm for the model implementation. Subsequently, we demonstrate the implementation with R code applied to our Sick-Sicker case example. We demonstrate that a microsimulation model in R can be built using a general framework that relies on of four key functions: (1) a function to define and update probabilities, (2) a function to define costs, (3) a function to define the health effects and (4) the overall microsimulation function. This microsimulation function executes the first three functions throughout the simulation time and samples the next health state of each individual. For educational purposes, we first present the microsimulation code as an iterative process at an individual level over the model's time horizon. In addition, we show a novel vectorized implementation of a microsimulation, where at each time iteration, all simulated individuals are transitioned simultaneously through the model. This vectorized approach achieves a 97% reduction in computational time.

Part II: Resource prioritization in healthcare

Prioritization of healthcare resources

In **Chapter 8** we focus on prioritization of scarce surgical resources in times of a pandemic. We developed a decision-analytic model in R to quantify expected health loss due to surgical delay. From a utilitarian perspective, which aims to maximize health for the population, this outcome measure can be used as an urgency measure to prioritize surgeries in case of scarce surgical capacity. This open-source decision-analytic model is more evidence-based and more transparent than the alternative strategy of triaging based on expert opinion. In addition, this approach prioritizes patients from all specialties, instead of the current approach which prioritizes patients within the silos of each specialty first. Although the available evidence about relative treatment effects is limited and one might disagree with the utilitarian perspective, this modeling approach can help to minimize health losses when trying to overcome delay in surgical procedures across disciplines in times of scarce surgical capacity.

Prioritization of research resources and design of clinical trials

In **Chapter 9** we outline an iterative research cycle, which we call the *value-driven* approach. The *value-driven* approach is a state-of-the-art approach for research prioritization and trial design because it evaluates the value that research can provide to clinical and policy decision makers. The *value-driven* approach is an alternative to the more traditional but frequently critiqued trial design approach, that is focused on controlling statistical errors and therefore called the error-driven approach. Via the *value-driven* approach we outline how decision-analytic models can be used to optimize research resources. The first step in the iterative research cycle of the *value-driven* approach is to determine the health decision making problem that is relevant to the research question. Next, a decision-analytic model is used to integrate available evidence on benefits and costs and to calculate the expected value of each intervention. This is followed by a probabilistic sensitivity analysis and VOI analysis. These analyses are done to determine the decision uncertainty in both the probability and the consequences of making the wrong decision about the optimal intervention. The results from these analyses inform whether there is value in collecting further evidence to reduce decision uncertainty or whether it is better to make the decision now given the current available evidence. Although the *value-driven* approach requires more data and time to design a study, we conclude that this approach, if correctly applied, can reduce research waste by helping us prioritize research that is most valuable to society. The upfront extra time investment of this approach also pays off after data collection. The *value-driven* approach reduces the time between data collection and results publication, because the model to perform the analysis has already been developed.

In **Chapter 10** we demonstrate how the *value-driven* approach can be used to inform policymakers and clinicians about the optimal strategy of a new drug with respect to both treatment implementation and research in times of a pandemic. During the COVID-19 pandemic, decision makers are facing time-sensitive decisions regarding new interventions because they have to deal with limited and uncertain clinical evidence. In addition, this evidence is likely to be updated rapidly because of the evolving scientific understanding of the disease pathway and its clinical management. We evaluated eight therapies for hospitalized COVID-19 patients: Baricitinib-Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Hydroxychloroquine, Interferon beta-1a, Lopinavir-Ritonavir, Remdesivir and Tocilizumab. For this evaluation, we developed a cohort state-transition model in R combined with a value of information analysis to quantify the consequences of approving these therapies or pursuing further research; either immediate approval, use only in further research, approval with research (e.g., Emergency Use Authorization), or reject. At the time of publishing, we concluded that the optimal strategy for Dexamethasone, Remdesivir, Casirivimab-Imdevimab, Baricitinib-Remdesivir and Tocilizumab was to approve the drugs and implement them in daily practice, while Interferon-B1a and Lopinavir-Ritonavir should be rejected without further research. Hydroxychloroquine is a controversial case. The optimal strategy based on our analysis would be to use Hydroxychloroquine only in a research setting. However, this comes with the warning that this study would be performed to further explore the decremental cost-effectiveness of the drug. This drug saves costs through the reduction of long-term healthcare expenditure due to the decreased survival of hospitalized COVID-19 patients. We do not consider this an ethical path to follow. In conclusion, we show how the *value-driven* approach combined with a state-of-the-art open-source decision-analytic model can be used for resource prioritization in healthcare in times of a pandemic. Especially, the use of open-source software can facilitate a "living" character of the evaluation to identify the optimal moment for drug or research approval.

General discussion

Finally, in **Chapter 11** I discuss the main results of each part of this thesis and provide recommendations for future research directions.

This thesis, together with recent efforts from others, has proven that R can be used to develop transparent decision-analytic models for healthcare decision making. We have also demonstrated that R-based decision-analytic models can be used for a dynamic value-based allocation of scarce resources, which is especially relevant in times of rapidly changing conditions, such as a pandemic.

Although decision-analytic models are often used to help decision makers with efficient prioritization of resources, they are also associated with a number of limitations, which I discuss in this chapter. More specifically, our established decision-analytic modeling approach in R comes with a steep learning curve and requires from users a sufficient time investment to become familiar with the code. Furthermore, development of this type of models used for resource prioritization is time consuming and the data required to inform them may be limited, biased or lacking. These potential limitations can be an issue in situations such as the first phase of a pandemic, when the available data are limited but time-sensitive decisions have to be made.

We further conclude that R is too flexible to implement for large scale cost-effectiveness evaluations and health technology assessment submissions without clear modeling guidelines and proper coding education to decision modelers in healthcare. Without clear guidance and training, the field risks getting lost in code that is neither well documented nor systematically organized in a comprehensible and shareable manner.

Based on these findings, I recommend to explore how R-based decision-analytic models can be wider implemented for resource prioritization. To achieve this, I suggest to focus on the following three points: (1) the education of the current and next generation stakeholders of decision analysis in healthcare in cutting-edge open-source modeling methods; (2) international collaboration between various stakeholders participating in medical decision making to develop guidelines for open-source decision-analytic modeling and to integrate dynamic value-based allocation of scarce resources in the healthcare system; (3) standardization of international good practice guidelines for open-source based model submissions to HTA-agencies, academic journals and research funding agencies.

“Verveling kun je met nieuwsgierigheid genezen. Nieuwsgierigheid is niet te genezen.”

Dorothy Parker

SAMENVATTING

Het doel van dit proefschrift is om de kennis over het gebruik van open-source software voor het maken van beslismodellen uit te breiden en om aan te tonen hoe deze modellen kunnen bijdragen aan het prioriteren van beschikbare middelen in de gezondheidszorg. In dit proefschrift wordt de open-source programmeertaal R gebruikt om ons conceptuele raamwerk en de methoden te demonstreren. Onze methoden zouden evengoed gegeneraliseerd kunnen worden naar andere open-source programmeertalen. Om te laten zien hoe open-source beslismodellen kunnen worden gebruikt voor het prioriteren van beschikbare middelen gebruiken we twee COVID-19 gerelateerde problemen in de zorg.

Algemene inleiding

Dit eerste hoofdstuk introduceert het onderwerp en legt het doel van dit proefschrift uit.

Besluitvormers in de gezondheidszorg (zoals zorgprofessionals, patiënten en familie, beleidsmakers, zorgorganisaties en health technologie assessment (HTA) organisaties) worden vaak geconfronteerd met besluitvorming omtrent het toewijzen van schaarse middelen onder onzekere omstandigheden. Ongeacht de onzekerheid over de uitkomst van beslissingen moeten er keuzes gemaakt worden over het bestedingsdoel van de beschikbare middelen. Middelen in de gezondheidszorg omvatten alles dat nodig is om gezondheidsdiensten te kunnen verlenen, zoals materialen, faciliteiten en personeel. Deze middelen kunnen ingezet worden

voor het implementeren van een interventie. Ook kunnen deze middelen gebruikt worden voor onderzoek. Het extra bewijs dat via onderzoek verzameld wordt kan bijdragen aan het verkleinen van de onzekerheid rondom de keuzes. Om een goede keuze te kunnen maken over de inzet van beschikbare middelen, is het voor besluitvormers in de gezondheidszorg noodzakelijk voorafgaand aan hun beslissing de gevolgen (m.a.w. verwachte (gezondheids) uitkomsten) zo goed mogelijk te evalueren. Beslismodellen kunnen helpen om deze evaluatie uit te voeren.

Beslismodellen worden beschouwd als de meest uitgebreide en geschikte benadering om de besluitvorming in de gezondheidszorg te ondersteunen. Deze modellen kunnen zowel klinische als economische aspecten van een beslissing evalueren. Het maken van een betrouwbaar beslismodel is echter niet altijd eenvoudig. Gedurende het ontwikkelingsproces moeten veel keuzes en aannames gemaakt worden met betrekking tot zowel de structuur van het model als de gebruikte data. Een goede samenwerking tussen de modelleurs en besluitvormers is nodig om tot een beslismodel te komen dat de besluitvorming op de juiste manier kan ondersteunen.

Voor besluitvormers is het belangrijk om vertrouwen te krijgen in de nauwkeurigheid van de modelresultaten. Hiervoor dienen zij inzicht te vergaren in de modelstructuur, de model input waardes en de gemaakte aannames. Zo kunnen besluitvormers de modellen controleren op systematische fouten (bias) of andere gebreken. Deze beoordeling vereist dat de beslismodellen beschikbaar en transparant genoeg zijn voor adequate toetsing. Helaas zijn de uitwerkingen van de modellen (gemaakt in gespecialiseerde commerciële software of spreadsheetprogramma's) niet vanzelfsprekend en vaak lastig te beoordelen. Deze modellen worden daarom vaak als "black box" beschouwd. Daarnaast zijn de vereiste licenties bij deze software ook een beperkende factor. Bovendien faciliteren ze onvoldoende de nieuwste besliskundige methodes, zoals kalibratie en 'value of information' analyse.

Scripts van open-source programmeertalen, zoals R of Python, hebben het potentieel om veel van de uitdagingen die gespecialiseerde softwareprogramma's en spreadsheets hebben te overwinnen. Beslismodellen gemaakt in deze open-source programmeertalen kunnen worden gebruikt om het besluitvormingsproces op een transparante manier te informeren. Er bestaan echter geen duidelijke handleidingen voor de implementatie van beslismodellen in deze open-source talen. Om het volledige potentieel van deze programmeertalen voor besluitvorming in de gezondheidszorg te realiseren, onderzoekt dit proefschrift hoe beslismodellen in R kunnen worden geïmplementeerd.

In het tweede deel van dit proefschrift wordt aangetoond hoe open-source beslismodellen kunnen worden gebruikt voor het optimaal inzetten van schaarse middelen in de gezondheidszorg. Middels beslismodellen evalueren we de verwachte resultaten van verschillende strategieën. Deze verwachte resultaten worden vaak uitgedrukt als gezondheidseffecten, al dan niet gecombineerd met kosten. Op basis van de verwachte gezondheidseffecten, kosten of een uitkomstmaat die effecten en kosten combineert, kan worden bepaald welke strategieën het gebruik van de beschikbare middelen rechtvaardigen (d.w.z. hoe de "beste prijs-kwaliteitverhouding" kan worden verkregen).

Net als de middelen voor gezondheidszorg, zijn ook de middelen om onderzoek te financieren beperkt. Onderzoeksmiddelen kunnen worden gedefinieerd als alle faciliteiten, materialen, arbeid, studiedeelnemers en alle andere middelen die nodig zijn om een (klinisch) onderzoek uit te voeren. Geavanceerde besliskundige methoden kunnen ook worden gebruikt om te berekenen hoe onderzoeksmiddelen het beste ingezet kunnen worden. Deze methoden worden "value of informatie" (VOI) analyses genoemd. Door middel van VOI-analyses kunnen we bepalen hoeveel het ons waard is om aanvullend bewijs te verzamelen dat helpt om de onzekerheid rondom een beslissing die gemaakt moet worden te verminderen. We zeggen dat het aanvullende bewijs waarde heeft

wanneer het de kosten die gepaard gaan met het nemen van een potentieel verkeerde beslissing kan verlagen.

Historisch gezien waren deze VOI-analyses rekenkundig veeleisend of was het niet mogelijk om ze uit te voeren in de gebruikelijke software. In de afgelopen tien jaar zijn verschillende pogingen ondernomen om deze berekeningen efficiënter te maken en hebben onderzoekers verschillende benaderingsmethoden ontwikkeld. Ondanks deze verbeteringen zijn lang niet alle beleidsmakers of ontwerpers van onderzoeksplannen bekend met deze methoden. De toepassing van VOI-analyses voor het opzetten van klinische studies is hierdoor beperkt. Daarom communiceren we in het tweede deel van dit proefschrift de potentiële waarde van VOI-analyses aan een breed publiek. Ook laten we met behulp van een COVID-19 voorbeeld zien dat deze nieuwe methoden, in combinatie met het gebruik van open-source beslismodellen, een efficiënte implementatie van VOI-analyse mogelijk maken.

Samenvattend, dit proefschrift draagt bij aan het uitbreiden van kennis over het gebruik van open-source software voor het maken van besliskundige modellen en demonstreert hoe open-source beslismodellen kunnen bijdragen aan het prioriteren van beschikbare middelen in de gezondheidszorg.

Deel I: Open-source beslismodellen in de gezondheidszorg

In **Hoofdstuk 2** presenteren we een generiek methodologisch raamwerk om open-source beslismodellen op te stellen. Dit methodologische raamwerk is door de "Decision Analysis in R for Technologies in Health" (DARTH) werkgroep gemaakt. Het DARTH-raamwerk biedt richtlijnen voor de implementatie van beslismodellen in R en verdeelt veelvoorkomende beslismodellen in vijf gemeenschappelijke componenten. De eerste vier componenten, (1) model input, (2) implementatie van het beslismodel, (3) modelkalibratie en (4) modelvalidatie, vormen de modelontwikkelingsfase. In het vijfde component, (5) de analyse, wordt het model (ontwikkeld in componenten 1-4) daadwerkelijk toegepast om de onderzoeksvragen te beantwoorden. De analyse wordt ook gebruikt om de onzekerheid rondom de verschillende keuzeopties in kaart te brengen. Daarnaast kan de analyse gebruikt worden om te berekenen of verder onderzoek zinvol is door middel van VOI-analyse. De flexibiliteit van het raamwerk konden we bevestigen door het toe te passen op een voorbeeld, genaamd het "Sick-Sicker" model. Alle data en codes die gebruikt zijn voor dit voorbeeld zijn beschikbaar via GitHub. Hierdoor is het voor andere mogelijk het raamwerk te gebruiken voor andere toepassingen.

In **Hoofdstuk 3** hebben we literatuuronderzoek verricht om inzicht te krijgen in de vraag hoe vaak R, in vergelijking met soortgelijke software, wordt gebruikt in studies met betrekking tot besluitvorming in de gezondheidszorg. Op basis van onze resultaten lijkt R in populariteit toe te nemen. In de wetenschappelijke tijdschriften die zijn beoordeeld, is het aandeel van studies dat R gebruikt tussen 2010 en 2015 met bijna 50% toegenomen. Ook bespreken we in dit hoofdstuk welke R-packages (een bundel gedocumenteerde code en functies dat gemakkelijk te delen is voor gebruik door andere) en welke functies relevant zijn voor besliskundige analyses in de gezondheidszorg. Deze R-packages zijn gegroepeerd per stadium van modelontwikkeling en analyse zodat ze gemakkelijk te vinden zijn.

Hoofdstuk 4-7 bouwen verder op de bevindingen van Hoofdstuk 2 en 3, door het ontwikkelen van een praktische benadering voor het implementeren van (geavanceerde) beslismodellen voor de gezondheidszorg in R. De flexibiliteit van R maakt dat beslismodellen op meerdere manieren kunnen worden geïmplementeerd. De manier van implementatie is sterk afhankelijk van de onderzoeksvraag die moet worden beantwoord en de programmeerervaring en -stijl van de modelleur. Daarom streefden we bij de ontwikkeling van een generieke aanpak waarbinnen modelleurs kunnen werken naar een balans tussen duidelijkheid en efficiëntie.

In **Hoofdstuk 4** laten we zien hoe we tijdonafhankelijke cohort-transitie modellen in R kunnen bouwen. Tijdonafhankelijk betekent dat de overgangskans tussen gezondheidstoestanden of waarderingen die geassocieerd zijn met die gezondheidstoestand (zoals kosten of gezondheidsverbeteringen) niet veranderen over de tijd. Eerst leggen we conceptueel uit hoe een tijdonafhankelijk cohort-transitie model kan worden geïmplementeerd in een programmeertaal. We gebruiken wiskundige formules om de drie kerncomponenten van een cohort-transitie model te beschrijven: (1) de gezondheidstoestandsvector, deze beschrijft de verdeling van het cohort over de verschillende gezondheidstoestanden op een bepaald moment in de tijd, (2) de "cohort trace matrix", deze beschrijft de verdeling van het cohort over de gezondheidstoestanden over de gehele tijdshorizon van de analyse, en (3) de transitiekans matrix, die de overgangskansen specificeert van het verplaatsen van één van de gezondheidstoestanden in het model naar een andere gezondheidstoestand op een bepaald punt in de tijd. Daarnaast beschrijven we de concepten verdiscontering, de correctie voor het gebruik van een tijdsinterval tijdens de simulatie (m.a.w. stapgrootte i.p.v. een continu proces) en het toepassen van waarderingen die horen bij de gezondheidstoestanden waarin individuen zich bevinden. Vervolgens bieden we een template aan, om een tijdonafhankelijk cohort-transitie model in R te implementeren. In het template gebruiken we ons "Sick-Sicker" casusvoorbeeld voor een uitgebreide kosteneffectiviteitsanalyse met vier strategieën, waarbij we ook een probabilistische sensitiviteitsanalyse demonstreren en de resultaten visualiseren.

Tijdonafhankelijke modellen zijn eenvoudiger te gebruiken dan *tijdafhankelijke* modellen. Echter, de meeste problemen in de gezondheidszorg kunnen het beste worden gemodelleerd op basis van een vorm van tijdafhankelijkheid. **Hoofdstuk 5** bouwt dan ook voort op de kennis van Hoofdstuk 4 door te laten zien hoe tijdafhankelijkheid kan worden opgenomen in cohort-transitie modellen gemaakt in R.

We onderscheiden twee soorten tijdafhankelijkheid: tijd sinds het begin van de simulatie (simulatietijd-afhankelijkheid) en de tijd doorgebracht in een gezondheidstoestand (toestand-duur afhankelijkheid). Bij simulatietijd-afhankelijkheid beïnvloedt de tijd van de simulatie de tijdafhankelijke modelparameters op dezelfde manier voor het gehele cohort (bijvoorbeeld leeftijd-specifieke sterfte van het cohort). Simulatietijd-afhankelijkheid kan in het model worden opgenomen door de transitiematrix (de wiskundige structuur die gebruikt wordt om de verplaatsingen van het cohort over de gezondheidstoestanden te bepalen) tijdafhankelijk te maken.

Bij afhankelijkheid van de duur van de gezondheidstoestanden (toestand-duur afhankelijkheid) krijgen de leden van het cohort op verschillend*e* momenten gedurende de simulatie te maken met veranderingen van de tijdafhankelijke modelparameters. Deze veranderingen in de parameters (bijvoorbeeld het risico op bijwerkingen van een medicijn dat afneemt bij langere gebruiksduur) zijn namelijk afhankelijk van hoe lang iemand zich in een bepaalde gezondheidstoestand bevindt. Het moment waarop individuen van het cohort zich naar een andere gezondheidstoestand verplaatsen (bijvoorbeeld ziek worden) is echter niet voor iedereen hetzelfde. Om de modelparameters op de juiste wijze aan te passen, is het nodig om bij te houden op welk moment elk lid van het cohort ziek werd. Dit kan gedaan worden door het aantal gezondheidstoestanden in het model uit te bereiden. Net als in Hoofdstuk 4, gebruiken we in dit hoofdstuk ons "Sick-Sicker" voorbeeld om onze ontwikkelde methoden te demonstreren. Op basis van dit voorbeeld laten we zien hoe verschillende epidemiologische en economische uitkomsten berekend kunnen worden en hoe een kosteneffectiviteits- en probabilistische sensitiviteitsanalyse uitgevoerd kunnen worden.
Samen geven de Hoofdstukken 4 en 5 een nieuwe methodologische benadering voor het stap voor stap implementeren van cohort-transitie modellen R.

In **Hoofdstuk 6** introduceren we een nieuwe methode genaamd de dynamische-array-benadering. Door middel van de dynamische-array-benadering kunnen we de uitkomsten van cohort-transitie modellen berekenen. Daarnaast kan deze benadering informatie opslaan over zowel de verdeling van het cohort over de gezondheidstoestanden in het model, als hoe ze zich gedurende de tijd tussen deze gezondheidstoestanden verplaatst hebben. Het bijhouden van de overgangsdynamiek tussen de gezondheidstoestanden is nieuw ten opzichte van de traditionele manier voor het opslaan van uitkomsten van een cohort-transitie model, via de zogenaamde “cohort trace”. Met de dynamische-array-benadering blijft alle dynamiek in het model bewaard. Deze methode maakt het mogelijk om zowel gezondheidstoestandwaarderingen als veranderingswaarderingen toe te passen. Gezondheidstoestandwaarderingen zijn waarderingen die toegewezen worden voor het doorbrengen van tijd in een bepaalde gezondheidstoestand, bijvoorbeeld ziek zijn. Veranderingswaarderingen zijn waarderingen die alleen worden toegepast wanneer een specifieke overgang van de ene gezondheidstoestand naar de andere plaatsvindt, bijvoorbeeld ziek worden.

Bovendien vergemakkelijkt de dynamische-array-benadering de berekening van epidemiologische ziektefrequenties, zoals overleving, prevalentie en ziektekansen gedurende het leven. Deze epidemiologische uitkomsten kunnen van belang zijn voor kalibratie en validatie. In een simulatiestudie laten we zien dat de dynamische-array-benadering sneller is dan de traditionele cohort-trace-benadering. Wel vergt deze nieuwe benadering meer geheugen (1-3,5 keer) voor modellen met een gemiddeld aantal gezondheidstoestanden (<50 gezondheidstoestanden). Het is echter onwaarschijnlijk dat dit relatieve geheugenverschil een beperking zal vormen voor het uitvoeren van de analyse. Daarom is het structureren van de output van cohort-transitie modellen aan de hand van de dynamische-array-benadering een efficiënte en compacte methode om alle modeldynamiek samen te vatten.

In **Hoofdstuk 7** beschrijven we een nieuwe methode om transitie modellen te construeren in R waarbij individuen worden gesimuleerd. Deze modellen noemen we ook wel microsimulatie modellen. Eerst beschrijven we conceptueel het benodigde algoritme voor de implementatie van een microsimulatie. Vervolgens demonstreren we aan de hand van ons “Sick-Sicker” voorbeeld hoe deze methode in R-code kan worden toegepast. We laten zien dat een microsimulatiemodel in R gebouwd kan worden met behulp van een algemeen raamwerk dat gebruik maakt van vier sleutelfuncties: (1) een functie om kansen te definiëren en bij te werken, (2) een functie om kosten te definiëren, (3) een functie om de gezondheidseffecten te definiëren en (4) de algemene microsimulatiefunctie. Deze microsimulatiefunctie voert de eerste drie functies uit gedurende de simulatie en bepaalt vervolgens op basis van de berekende kansen de volgende gezondheidstoestand van elk individu. Om de lezer mee te nemen, presenteren we eerst de microsimulatiecode als een iteratief proces waarbij we één voor één elk individu volgen over de gehele tijdshorizon van het model. Vervolgens demonstreren we een nieuwe – gevectoriseerde - implementatie van een microsimulatie, waarbij bij elke iteratie voor alle gesimuleerde individuen gelijktijdig bepaald wordt wat hun volgende gezondheidstoestand is. Deze gevectoriseerde aanpak reduceert de tijd die nodig is om de analyse in R uit te voeren met 97%.

Deel II: Prioritering van beschikbare middelen in de gezondheidszorg

In het tweede deel van dit proefschrift demonstreren we hoe open-source beslismodellen kunnen worden gebruikt voor het optimaal inzetten van schaarse middelen in de gezondheidszorg.

Prioritering van middelen in de gezondheidszorg

In **Hoofdstuk 8** richten we ons op het verdelen van schaarse operatiekamercapaciteit ten tijde van een pandemie. We ontwikkelden een besliskundig model in R dat berekent hoeveel gezondheid de patiënt verliest door uitstel van een operatie. De

mate van verwacht gezondheidsverlies gebruiken we als maat voor urgentie. Dit wil zeggen, patiënten die snel veel gezondheid verliezen door uitstel van een operatie hebben een hogere prioriteit dan patiënten die minder gezondheid verliezen door uitstel. Deze urgentiemaat kan gebruikt worden vanuit een utilitair perspectief. Binnen dit gestelde ethische perspectief is het van belang gezondheidsverlies voor de bevolking als geheel door uitgestelde operaties te minimaliseren. Operatieplanning op basis van dit model is een objectievere en transparantere werkwijze dan een alternatieve strategie waarin triage op basis van inschattingen van klinische experts plaatsvindt. Bovendien kunnen we door deze aanpak keuzes maken die specialismen overstijgen, doordat de aanpak de urgentie van patiënten uit alle specialismen vergelijkt. Dit is anders dan een veelvoorkomende aanpak waarbij geprioriteerd wordt binnen de muren van elk specialisme. Hoewel het beslismodel beperkingen heeft kan deze specialisme-overstijgende waardegedreven aanpak helpen om gezondheidsverlies door operatie uitstel te minimaliseren in tijden van schaarse operatiekamercapaciteit.

Prioritering van onderzoeksmiddelen en het ontwerp van klinische studies

In **Hoofdstuk 9** schetsen we een iteratieve onderzoekscyclus, die we de *waardegedreven* benadering noemen. Deze benadering is een geavanceerde benadering voor het prioriteren en opzetten van onderzoeken waarbij gekeken wordt naar welke waarde het onderzoek heeft in de besluitvorming. Deze *waardegedreven* benadering is een alternatief voor de meer traditionele - en vaak bekritiseerde - manier om onderzoek op te zetten waarbij men gericht is op het beperken van statistische fouten, namelijk de *foutgestuurde* benadering. We beschrijven aan de hand van de *waardegedreven* benadering hoe beslismodellen kunnen bijdragen aan het optimaal gebruik van onderzoeksbudgetten. De eerste stap in de iteratieve onderzoekscyclus van de *waardegedreven* benadering is het vaststellen van een relevant onderzoeksvraag voor het keuzeprobleem binnen de gezondheidszorg. Vervolgens wordt een besliskundig model gebruikt om het beschikbare bewijs over kosten en baten te integreren en om de verwachte uitkomsten van elke interventie te berekenen. Deze berekeningen worden gevolgd door een probabilistische sensitiviteitsanalyse en VOI-analyse. Deze analyses worden gedaan om de beslissingsonzekerheid te bepalen. De resultaten van deze analyses geven aan of het waardevol is om verder bewijs te verzamelen om de onzekerheid over de beslissing te verminderen of dat het beter is om de beslissing nu te nemen, gezien het huidige beschikbare bewijs.

Hoewel deze *waardegedreven* benadering meer gegevens en tijd vereist om een onderzoek te ontwerpen, concluderen we dat deze benadering, indien correct toegepast, verspilling van beschikbare onderzoeksmiddelen kan verminderen door te helpen prioriteit te geven aan onderzoek dat het meest waardevol is voor de samenleving. De extra tijdsinvestering voorafgaand aan deze aanpak loont na het verzamelen van gegevens. Omdat het model dat nodig is om de analyses uit te voeren al ontwikkeld is, is de tijd tussen het verzamelen van de onderzoeksgegevens en de publicaties van onderzoeksresultaten bij de *waardegedreven* benadering korter.

In **Hoofdstuk 10** laten we zien hoe een *waardegedreven* benadering kan worden gebruikt om beleidsmakers en clinici te informeren over de optimale strategie van een nieuw medicijn ten tijde van een pandemie. Tijdens een pandemie worden besluitvormers geconfronteerd met tijdgevoelige beslissingen over nieuwe interventies waarover nog veel onzekerheid is. Dit komt vaak doordat er slechts beperkt en onzeker klinisch bewijs beschikbaar is. Bovendien wordt dit bewijsmateriaal waarschijnlijk snel bijgesteld door het vergaren van nieuwe wetenschappelijke inzichten over de ziekte. In ons onderzoek analyseerden we de volgende acht medicijnen voor gehospitaliseerde COVID-19-patiënten: Baricitinib-Remdesivir, Casirivimab-Imdevimab, Dexamethasone, Hydroxychloroquine, Interferon bèta-1a, Lopinavir-Ritonavir, Remdesivir en Tocilizumab. Voor deze evaluatie hebben we een cohort-transitie model in R ontwikkeld en gecombineerd met een VOI-analyse. De VOI-analyse werd gebruikt om de optimale strategie te bepalen. VOI-analyse

kan dat door te kwantificeren wat de verwachte consequenties zijn van zowel het implementeren van het medicijn (d.w.z. het wel of niet gebruiken) als het doen van verder onderzoek. Dit geeft ons vier keuzes (1) onmiddellijke goedkeuring van het medicijn voor gebruik, (2) het medicijn alleen gebruiken in verder onderzoek, (3) goedkeuring voor gebruik met daarnaast ook onderzoek (bijvoorbeeld autorisatie voor noodgebruik), of (4) geheel afwijzen. Op het moment van publicatie concludeerden we dat voor Dexamethasone, Remdesivir, Casirivimab-Imdevimab, Baricitinib-Remdesivir en Tocilizumab de optimale strategie was om het medicijn goed te keuren en in de dagelijkse praktijk toe te passen, terwijl Interferon bèta-1a en Lopinavir-Ritonavir geheel afgewezen moesten worden. Hydroxychloroquine was controversieel. De optimale strategie op basis van onze analyse zou zijn om Hydroxychloroquine alleen te gebruiken in een onderzoek setting. Het is echter van belang om te vermelden dat deze studie gedaan zou moeten worden om verder te onderzoeken in hoeverre Hydroxychloroquine resulteert in kostenbesparingen en gezondheidsverlies. Het is waarschijnlijk dat Hydroxychloroquine kosten bespaart doordat het zorgt voor een verminderde overleving van gehospitaliseerde COVID-19-patiënten die daardoor minder langdurig zorgkosten maken. We beschouwen het uitvoeren van een studie om aan te tonen dat we kosten kunnen besparen doordat mensen sneller overlijden niet als een ethisch acceptabele aanpak. Ons onderzoek laat zien hoe een *waardegedreven* benadering in combinatie met *state-of-the-ar*t open-source beslismodel gebruikt kan worden voor het prioriteren van middelen in de gezondheidszorg en onderzoek in tijden van een pandemie. Vooral het gebruik van open-source software kan een 'levend' karakter van de evaluatie vergemakkelijken om zo het optimale moment te identificeren voor goedkeuring van geneesmiddelen of onderzoek.

Algemene discussie

Tot slot worden in **Hoofdstuk 11** de belangrijkste resultaten van de thesis belicht en worden aanbevelingen en toekomstige onderzoeksrichtingen besproken.

Deze thesis, samen met de inspanningen van anderen, heeft bewezen dat R kan worden gebruikt om transparante beslismodellen voor besluitvorming in de gezondheidszorg te ontwikkelen. We hebben bovendien aangetoond dat beslismodellen gemaakt in R gebruikt kunnen worden voor een dynamische, op *waardegedreven* verdeling van schaarse middelen. Deze aanpak is vooral relevant ten tijde van een pandemie.

Hoewel beslismodellen veel voordelen hebben voor het prioriteren van schaarse middelen, bespreken we in dit hoofdstuk ook hun beperkingen. Zo gaat onze methodologische modelleringsaanpak gepaard met een steile leercurve voor de modelleurs en vereist deze een aanzienlijke tijdsinvestering van gebruikers om vertrouwd te raken met de code. Bovendien is het ontwikkelen van beslismodellen voor het prioriteren van schaarse middelen tijdrovend en kunnen de gegevens die nodig zijn om deze modellen te informeren beperkt, vertekend of niet beschikbaar zijn. Dit is met name een beperking in de eerste fase van een pandemie, waarbij tijdgevoelige beslissingen genomen dienen te worden.

We concluderen ook dat R te flexibel is om over te gaan op grootschalige implementatie voor kosteneffectiviteitsanalyses wanneer duidelijke richtlijnen en cursussen voor besliskundige modelleurs ontbreken. Zonder duidelijke instructies over hoe modellen aangeleverd moeten worden en training voor de modelleurs loopt het vakgebied het risico te verdwalen in niet goed gedocumenteerde codes met alle gevolgen van dien.

Op basis van de bevindingen in deze thesis raad ik aan om te onderzoeken hoe beslismodellen gemaakt in R breder kunnen worden geïmplementeerd voor het prioriteren van middelen en onderzoek.

Om dit te bereiken, stel ik voor om te focussen op drie punten. Laten we ons ten eerste richten op het opleiden van de huidige en de komende generatie die werkzaam is in de medische besliskunde in *state-of-the-art* open-source modelleringsmethoden. Ten tweede is internationale samenwerking tussen alle belanghebbenden op het gebied van medische besluitvorming noodzakelijk (bijvoorbeeld modelbouwers, fabrikanten, besluitvormers, datawetenschappers, clinici, financieringsinstanties, wetenschappelijke tijdschriften en HTA-organisaties) voor het ontwikkelen van richtlijnen voor open-source besluitvormingsanalyse en voor het implementeren van dynamische *waardegedreven* verdeling van schaarse middelen in de gezondheidszorg en onderzoek. Ten derde raad ik aan om te onderzoeken hoe internationale richtlijnen over het aanleveren van open-source beslismodellen bij HTA-organisaties, academische tijdschriften en financieringsorganitaties gestandaardiseerd kunnen worden.

“If you want to go fast, go alone, if you want to go far, go together.”

African Proverb

LIST OF CO-AUTHORS

List of authors that contributed to the manuscripts in this thesis, including their affiliations during the conduct of research.

▼ Fernando Alarid-Escudero

Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN, USA

Division of Public Administration, Center for Research and Teaching in Economics (CIDE)-CONACyT, Aguascalientes, AGS, Mexico

Rob Baatenburg de Jong ●

Department of Otorhinolaryngology (ENT), Erasmus University Medical Center, Rotterdam, the Netherlands

▶ Jan van Busschbach

Department of Medical Psychology, Erasmus University Medical Center, Rotterdam, the Netherlands

Netherlands Institute for Health Sciences, Erasmus University Medical Center, Rotterdam, the Netherlands

Sophie Bruinsma ■

Department of Quality and Patient Care, Erasmus University Medical Center, Rotterdam, the Netherlands

▼ Stijntje Dijk

Department of Epidemiology and Department of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands

Eva Enns ●

Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN, USA

▶ Geert Geleijnse

Department of Otorhinolaryngology (ENT), Erasmus University Medical Center, Rotterdam, the Netherlands

Benjamin Gravesteijn ■

Department of Otorhinolaryngology (ENT) and Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

▼ Carry Gross

Section of General Internal Medicine, Yale University School of Medicine, New Haven, CT, USA

Anna Heath ●

Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada
Division of Biostatistics, University of Toronto, Toronto, ON, Canada
Department of Statistical Science, University College London, London, UK

▶ Myriam Hunink

Department of Epidemiology and Department of Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands
Center for Health Decision Sciences, Harvard T.H. Chan School of Public Health, Boston, USA
Netherlands Institute for Health Sciences, Erasmus University Medical Center, Rotterdam, the Netherlands

Natalia Kunst ■

Department of Population Medicine, Harvard Medical School & Harvard Pilgrim Health Care Institute, Boston, USA
Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale University School of Medicine, New Haven, CT, USA

▼ Hester Lingsma

Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

Céline van Lint ●

Department of Quality and Patient Care, Erasmus University Medical Center, Rotterdam, the Netherlands

▶ Hawre Jalal

Department of Health Policy and Management, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

Isabel Retel Helmrich ■

Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

▼ Jan van Saase

Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

Ewout Steyerberg ●

Department of Biostatistics, Leiden University Medical Center, Leiden, the Netherlands
Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands

▶ Ernest van Veen
Department of Public Health and Department of Intensive Care, Erasmus University Medical Center, Rotterdam, the Netherlands

Kees Verhoef ■
Department of Surgical Oncology and Gastrointestinal Surgery, Erasmus University Medical Center, Rotterdam, the Netherlands

▼ Petros Pechlivanouglou
Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada
Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada

John Wong ●
Division of Clinical Decision Making, Tufts Medical Center, Boston, USA

▶ Alan Yang
Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada

Szu-Yu Zoe Kao ■
Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN, USA

“The way a team plays as a whole determines its success.”

Babe Ruth

LIST OF PUBLICATIONS

This thesis

Chapter 2: Alarid-Escudero F, **Krijkamp E**, Pechlivanoglou P, Jalal H, Kao S-YYZ, Yang A, Enns EA. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. *Pharmacoeconomics* 2019; 37: 1329–1339.

Chapter 3: Jalal H, Pechlivanoglou P, **Krijkamp E**, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making* 2017; 37: 735–746.

Chapter 4: Alarid-Escudero F, **Krijkamp E**, Enns EA, Yang A, Hunink MGM, Pechlivanoglou P, Jalal H. An Introductory Tutorial on Cohort State-Transition Models in R Using a Cost-Effectiveness Analysis Example. - available on arXiv & Under review at *Med Decis Making*

Chapter 5: Alarid-Escudero F, **Krijkamp E**, Enns EA, Yang A, Hunink MGM, Pechlivanoglou P, Jalal H. A Tutorial on Time-Dependent Cohort State-Transition Models in R using a Cost-Effectiveness Analysis Example. – *available on arXiv & Under review at Med Decis Making*

Chapter 6: **Krijkamp E***, Alarid-Escudero F*, Enns EA, Pechlivanoglou P, Hunink MGM, Yang A, Jalal HJ. A Multidimensional Array Representation of State-Transition Model Dynamics. *Med Decis Making* 2020; 40: 242–248.

Chapter 7: **Krijkamp E**, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MGM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making* 2018; 38: 400–422.

Chapter 8: Gravesteijn B*, **Krijkamp E***, Busschbach J, Geleijnse G, Retel Helmrich I, Bruinsma S, van Lint C, van Veen E, Steyerberg E, Verhoef C, van Saase J, Lingsma H, Baatenburg de Jong R. Minimizing Population Health Loss in Times of Scarce Surgical Capacity. *Value Health* 2020; 24: 648–657.

Chapter 9: Heath A*, Hunink MGM*, **Krijkamp E***, Pechlivanoglou P*. Prioritisation and design of clinical trials. *Eur J Epidemiol* 2021; 36(11):1111–1121.

Chapter 10: Dijk SW, **Krijkamp E**, Kunst N, Gross CP, Wong J, Hunink MGM. Emerging Therapies for COVID-19: the value of information from more clinical trials. – *Accepted Value Health*

* = these authors contributed equally to the manuscript

Other publications

- Broekhoff TF, Sweegers CCG, **Krijkamp E**, Mantel-Teeuwisse AK, Leufkens HGM, Goettsch WG, Vreman RA. Early Cost-Effectiveness of Onasemnogene Abeparvovec-xioi (Zolgensma) and Nusinersen (Spinraza) Treatment for Spinal Muscular Atrophy I in The Netherlands With Relapse Scenarios. *Value Health* 2021; 24: 759–769.
- Gravesteijn B, van Hof K, **Krijkamp E**, Asselman F, Leemans R, van der Horst H, Widdershoven G, Baatenburg de Jong L, Lingsma H, Busschbach J, Baatenburg de Jong R. Minimizing population health loss due to scarcity in OR capacity: a validation study – *Under review Plos One*
- Knowlton G, Alarid-Escudero F, Easterly CW, Kao SYZ, **Krijkamp E**, Yang A, Pechlivanoglou P, Jalal H, Jenness S, Enns EA. `dampack`: Visualizing and analyzing decision analytic modeling results in R. – *Ready to submit*

“Rivers know this: there is no hurry. We shall get there some day.”

A. A. Milne (Winnie-the-Pooh)

ABOUT THE AUTHOR

Eline Krijkamp was born on August 9th, 1991 in Haarlem, the Netherlands. After she completed her secondary school at the Haarlemmermeer Lyceum Hoofddorp in 2010, she studied Health and Life Sciences at the Vrije Universiteit (VU) in Amsterdam. She obtained her bachelor’s degree in 2013.

During her study, Eline was heavily involved at the European Board of an international student-run organization (STeLA) that aimed to create an international network of the next generation of leaders in science and technology. Part of her work consisted of the organization of the Forum themed “Managing 10 Billion people” in Delft and the Health and Bioethics Forum at Stanford University.

In 2013 she started the Master Health Economics, Policy and Law at the Erasmus University Rotterdam. Her master thesis focused on the internal inconsistency of the standard gamble under supervision of Arthur Attema and Prof. Han Bleichrodt. Eline obtained her Health Economics master's degree in June 2015.

After this graduation, Eline continued working on her second master thesis for the Master of Epidemiology at the Netherlands Institute for Health Sciences (NIHES). Under supervision of Prof. Myriam Hunink she did research in the field of medical decision making. A field that combines her interests in both health economics and epidemiology. During this time Eline became part of the Decision Analysis in R for Technologies in Health (DARTH) workgroup. With DARTH she explored how R-based decision-analytic models can be constructed. Eline obtained her Master in Epidemiology in July 2016.

Eline found a temporary job as a junior teacher for the Health and Life sciences program at the VU. After she traveled through South-America, she joined the NIHES educational innovation team in 2017. Eline became actively involved in education at NIHES, as well as in courses from the Center for Health Decision Science from Harvard. With DARTH she also had the opportunity to teach several courses at conferences, universities and public and private organizations.

The research endeavors with DARTH and Prof. Hunink ultimately led her to further pursue a scientific career with a focus on open-source decision analysis in healthcare. In 2018 Eline started as a PhD student at the Department of Epidemiology at the Erasmus MC. In 2019 Eline was awarded the SMDM Fellowship in Medical Decision Making funded by the Gordon and Betty Moore Foundation. The work she did in these positions is included in this thesis.

Although Eline loves her professional career, she likes to prioritize sports, good food and enough sleep to stay energized. In the summer of 2021 Eline moved from Amsterdam to Woerden. After graduation, Eline will start a new job at the Erasmus School of Health Policy & Management.

“Van tegenwind op de fiets ga je harder trappen, van meewind ga je denken dat je harder trapt.”

Michiel de Hoog

PHD PORTFOLIO

PhD training	Year	Organizer	Workload (EC)
Short course - Hands-on Model Calibration in R	2018	SMDM	0.20
Scientific Integrity	2020	Erasmus MC	0.30
Photoshop and Illustrator CC Workshop for PhD-students and other researchers	2019	MolMed	0.30
Basisdidactiek voor docenten (TtTI)	2019	Erasmus MC	0.40
Individuele begeleiding training	2019	Erasmus MC	0.10
Optimal Research Design Using Value of Information	2019	SMDM	0.20
Causal Diagrams: Draw Your Assumptions Before Your Conclusions	2020	edX	1.00
Causal Inference	2020	NIHES	1.40
Short course - Introduction to the Psychology of Medical Decision Making	2020	SMDM	0.15
Short course - Causal Inference and Causal Diagrams in Medical Decision Making	2020	SMDM	0.20
Short course - Quantifying and Valuing Health Inequality Impacts in Economic Evaluation	2020	SMDM	0.20
Core course - Introduction to Share Decision Making and Patient Decisions	2021	SMDM	0.20
Core course - Introduction to Medical Decision Analysis	2021	SMDM	0.20

	Year	Organizer	Workload (EC)
Core course - Introduction to the Psychology of Medical Decision Making	2021	SMDM	0.20
Short course - Metamodeling for simulation-based optimization of strategies in healthcare	2021	SMDM	0.20
Short course - Discrete Event Simulation in R to Support Healthcare Decision Making	2021	SMDM	0.20
Visual Communication workshop	2021	Promeras	0.10
Writing efficient R code	2021	Datacamp	0.15
Operations Management	2021	NIHES	1.00
Joint Models for Longitudinal and Survival Data	2021	NIHES	0.70
Data Science in Epidemiology	2021	NIHES	0.70

Conferences

	Year	Organizer	Workload (EC)
17th Biennial European Conference • *Host session, The DARTH Initiative: Promoting the Use of Open-Source Software in Medical Decision Making*	2018	SMDM	1.00
40th Annual North American Meeting • *Poster: Retrieving Intermediate Outcomes from a State Transition Model Using Multidimensional Matrices*	2018	SMDM	1.00
41th Annual North American Meeting	2019	SMDM	1.00
42nd Annual North American Meeting • *Poster - Utilitarian Distribution of Scarce Surgical Capacity during the COVID-19 Crisis and Beyond*	2020	SMDM	1.00
43rd Annual North American Meeting	2021	SMDM	1.00

Symposia and workshops

	Year	Organizer	Workload (EC)
RESET - Projectmanagement in de (semi)-publieke sector	2019	Over/Nieuw	0.10
Health(y) science - First Health Sciences Research Day	2019	Erasmus MC	0.30
European Cooperation on Healthcare	2019	Erasmus School of Health Policy & Management	0.25
PhD career day 2019	2019	Erasmus MC	0.30
Epidemiology of the future!	2020	VVE	0.10
PhD day 2021	2021	Erasmus MC	0.20
European Spring Virtual Event	2021	SMDM	0.20
R for HTA workshop	2021	University College London	0.30

Seminars

	Year	Organizer	Workload (EC)
CQM-seminars	2018-2021	Department of Biostatistics	1.40
Research seminars	2018-2021	Department of Epidemiology	3.00
Journal Club	2018-2021	Department of Epidemiology	1.00
Organize Decision Science Methods Club	2019-2021	Department of Epidemiology	3.00
Harvard Decision Science Journal Club and Research-in-progress meetings	2021	Harvard T.H. Chan School of Public Health	0.50

Reviewing papers

	Year	Organizer	Workload (EC)
Reviewing activities for European Journal of Epidemiology	2020		1.00
Reviewing activities for Value in Health	2021		2.00

Teaching activities

	Year	Organizer	Workload (EC)
Clinical Epidemiology	2018	NIHES	3.00
Topics in Medical Decision Making	2018-2021	NIHES	5.60
Advanced Decision Modeling	2018-2021	NIHES	4.20
Using R for Decision Modeling in Health Technology Assessment	2019-2021	NIHES	5.60
Methods for Decision Making in Medicine	2018-2021	Harvard T.H. Chan School of Public Health	4.80
Decision Analysis in R for Technologies in Health at La Roche Basel	2018	DARTH	0.50
Decision Modelling Using R at University of Oslo	2018	DARTH	0.50
Decision Modeling Using R	2018	DARTH	0.60
Beginner's Guide to Decision Modelling in R	2018	DARTH – SMDM	0.30
Microsimulation Modeling in R	2019	DARTH – SMDM	0.30
Advanced Decision Modeling in R	2020	DARTH	0.60
Microsimulation in R	2020	DARTH – SMDM	0.30
Decision Modeling for Public Health at CDC	2020	DARTH	0.60
Health Economic Modelling in R for Zorginstituut experts: a hands-on online training	2020	DARTH	0.60
Cohort Modeling in R	2021	DARTH – SMDM	0.30
Microsimulation Modeling in R	2021	DARTH – SMDM	0.30
Decision Modeling for Public Health at CDC	2021	DARTH	0.60

Other activities

	Year	Organizer	Workload (EC)
Fellowship	2019-2022	SMDM	
		Total EC	55.75

CDC, Centers for Disease Control and Prevention, United States;
DARTH, Decision analysis in R for Technologies in Health Workgroup;
MolMed, Erasmus Postgraduate school Molecular Medicine;
NIHES, Netherlands Institute for Health Sciences;
SMDM, Society for Medical Decision Making;
VVE, Vereniging voor Epidemiologie.