University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2021-12

Enhancing the Reproducibility of Health Technology Assessments

Wagner, Daniel

Wagner, D. (2022). Enhancing the reproducibility of health technology assessments (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. http://hdl.handle.net/1880/114250

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Enhancing the Reproducibility of Health Technology Assessments

by

Daniel Jonathan Wagner

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

JANUARY, 2022

© Daniel Jonathan Wagner 2022

Abstract

Health systems around the world depend on Health Technology Assessment (HTA) programs to provide policy guidance on many factors, including value-for-money. To ensure decisions are made with current information, methods of evidence synthesis and economic evaluation are used to inform a continuous process of evidence gathering and decision making. While computers are used in almost every part of this process, the act of updating an existing HTA often involves a duplication of the original effort. The experience of other scientific fields suggests this is attributable to a lack of reproducibility. This refers to the ability to obtain consistent computational results, using the same set of files and processes. The objective of this thesis was to explore how an emphasis on reproducibility can support the effective development and maintenance of HTAs.

Satisfaction of this objective required the identification and implementation of computing strategies to enhance the reproducibility of HTAs. A literature review was used to identify techniques for reproducibility which had proven successful in other fields. The identified strategies encouraged the creation of an accurate and complete record of the research process in human and machine-readable formats. These findings were subsequently applied to a case study which redeveloped an existing appraisal of biologic treatment for psoriatic arthritis. The first part of the case study summarized the development and execution of an automated workflow. The second part explored how the computing strategies affected the programming of the economic model. Outcomes from the case study included improved quality control, more efficient updating, and the elimination of barriers to the characterization of uncertainty. With enough investment, enhancing the reproducibility of HTAs will enable improved transparency, better decision making, and ultimately population health gains.

Preface

This thesis is original, unpublished, work by the author, Daniel Jonathan Wagner.

The body of research included in this thesis was developed under the guidance and support of the supervisory committee. Each member contributed intellectually to the content and provided critical review of each manuscript. At the time of this writing, these works have not been submitted for peer-review.

Acknowledgements

The research presented in this thesis was supported by a Graduate Studentship in Health Innovation from Alberta Innovates. In addition to their generous financial support, I would like to extend my appreciation to Alberta Innovates for reaching out to ensure I was well supported at the beginning of the pandemic. Prior to the initiation of this award, I was supported by a scholarship jointly funded by the Department of Community Health Sciences and the Svare Chair in Health Economics at the University of Calgary. This unique arrangement was the result of many months of advocacy on my behalf from Dr. Braden Manns and Dr. Elizabeth Oddone Paolucci. Their counsel and advice transformed my doctoral training experience, and I am grateful for their efforts.

I have been fortunate to have received support of many additional faculty members over the course of this degree. I'd like to express my appreciation to Drs. Paul Ronksley and Jesse Hendrikse, both of whom helped to guide my thinking on the initial ideas which inspired this thesis. In addition, Dr. Noah Ivers continues to be an invaluable source of feedback and mentorship to me. To this day, his care for my success has been that of a true friend.

I must extend my sincere gratitude to the members of my supervisory committee: Dr. Reed Beall and Dr. Laura Bojke. Their thoughtful feedback helped me clarify many of the complex ideas captured in the thesis and strengthened this body of work. I must also take a moment to acknowledge Dr. Bojke for her trust and reassurance as I proceeded with the redevelopment of the case study.

Developing this body of work would not have been possible without the advice and tutelage of Dr. Eldon Spackman and Dr. Glen Hazlewood. From the start of our working relationship, I have benefitted from their willingness to act as a sounding board for my *many*

different project ideas or questions. I am proud of the fact that this dynamic resulted in a collaborative effort which allowed all three of us to learn new things together. Furthermore, the ambitious nature of this thesis is a testament to both supervisors' insight, intuition, expertise, and encouragement.

Through their limitless encouragement and dedication, my family has played a critical role in my success. In addition to being wonderful role models, my parents' guidance and advice help me navigate my way through this degree. I'd also like to thank Steven, Kira, and Ari for their unconditional support throughout this process. Our daily, hours-long, video calls in the early parts of the pandemic were the perfect antidote to the isolating nature of writing a thesis.

Finally, I can think of no larger contributor to my current and future accomplishments than the time I spent at Calgary Academy. In one short year, this special school taught me how to leverage my strengths to overcome my learning challenges and changed my life in the process. I take great pride in being able to serve as a role model for those who face similar challenges as I did.

Table of Contents

List of Tables

List of Illustrations, Figures, & Graphics

List of Symbols, Abbreviations, & Nomenclature

Abbreviation	Definition
ADA	Adalimumab
APR	Apremilast
BSC	Best Supportive Care
CADTH	Canadian Agency for Drugs and Technology in Health
CBA	Cost-Benefit Analysis
cDMARD	Conventional Disease Modifying Anti-Rheumatic Drug
CEA	Cost-Effectiveness Analysis
CPU	Central Processing Unit
CSV	Comma Separated Values
CUA	Cost-Utility Analysis
CZP	Certolizumab Pegol
ENBS	Expected Net Benefit of Sampling
ETN	Etanercept
EVPI	Expected Value of Perfect Information
EVPPI	Expected Value of Partial Parameter Information
EVSI	Expected Value of Sample Information
GOL	Golimumab
HAQ-DI	Health Assessment Questionnaire – Disability Index
HTA	Health Technology Assessment
ICER	Incremental Cost-Effectiveness Ratio
INF	Infliximab
MSAC	Australian Medical Services Advisory Committee
MTA	Multiple Technology Appraisal
NHB	Net Health Benefit
NICE	National Institute for Health and Care Excellence
NMA	Network Meta-Analysis
NMB	Net Monetary Benefit
PASI	Psoriasis Area Severity Index
PsA	Psoriatic Arthritis
PsARC	Psoriatic Arthritis Response Criteria
QALY	Quality Adjusted Life Year
SEC	Secukinumab
UST	Ustekinumab
VoI	Value of Information

Chapter 1 Introduction

Evidence-based decision making has been embraced by jurisdictions around the world as a mechanism to improve health system quality and efficiency. This is reflected in the design of Health Technology Assessment (HTA) programs which support the managed entry of new technologies (1–3). Prominent examples of institutions which administer these activities include Australia's Medical Services Advisory Committee (MSAC), the United Kingdom's National Institute for Health and Care Excellence (NICE), and the Canadian Agency for Drugs and Technology in Health (CADTH) (4,5). A distinguishing feature of the HTA process is the commissioning of technology appraisals which leverage methods of evidence synthesis to compile the information required by the decision maker. These reports and the guidelines used to prepare them tend to be publicly available to increase accountability and transparency in decision making (6–8).

A technology appraisal may be commissioned for any policy, product, or intervention which is meant to improve health. The purpose of this effort is to systematically evaluate the relevant evidence relating to the use of the technology. As a result, most works will include a comparison of the technology's safety and clinical effectiveness against relevant alternatives. Pressures on health budgets have also forced policymakers to consider the effects of a health technology in the context of its costs. These trade-offs are estimated through methods of economic evaluation (1–8).

1.1 Economic Evaluation

Economic evaluation provides a mechanism to compare mutually exclusive alternatives with respect to their resource use (costs) and expected outcomes (benefits) (4,6,7). Making an

informed decision on these grounds will require the identification, measurement, and valuation of both the benefits and costs associated with each alternative included in an appraisal.

The approach used to measure the benefits of a health technology will inform the classification of an economic evaluation (4). For example, Cost-Benefit Analysis (CBA) is a type of economic evaluation where the benefits are measured in monetary terms. This approach is common in other sectors of the economy such as transportation or finance (4). While CBA is rarely applied in the health context, a measure of benefit can be estimated using stated preference methods like Discrete Choice Experiments (9,10). Another type of economic evaluation is a Cost-Effectiveness Analysis (CEA). This designation applies to situations where benefits are measured in terms of natural units like gains in life-expectancy and number of cases prevented, among others. The choice of which measure to use will depend on the objective of the analysis. A major limitation with this approach in decision-making is that it does not facilitate the comparison of benefits across programs. Such an analysis necessitates a generic measure of benefit which is relevant to all the interventions for which a decision maker is responsible (4).

Variants of CEA which use a generic measure of benefit are the most common form of economic evaluation in the HTA process. Often defined as a Cost-Utility Analysis (CUA), HTA-commissioned studies tend to use the Quality Adjusted Life Year (QALY) as a standardized approach to promote comparability across interventions. (4,6,11,12). The QALY is a generic measure of disease burden in which the length of life for a specific health state is adjusted to reflect the quality of life for that state (11). As a measure of benefit, it is dependent on two underlying assumptions. First, that a major objective of decision makers is to maximize health across the population. Second, that health can be measured or valued based on the amount of time spent in various health states (11).

1.1.1 Decision Modelling as a Vehicle for Economic Evaluation

The recommended approach for assembling the costs and effects for an economic evaluation is known as decision modelling. The model itself will take the form of an explicit set of mathematical relationships between different clinical events and health states. Costs and effects are estimated from a series of pathways specific to each mutually exclusive alternative considered in the economic evaluation (4,13–15). While many model structures are available, the most popular tend to be decision trees and Markov models (4,13–15).

Development of a decision model will require a comprehensive understanding of the decision problem and its context. One must determine the alternatives to be compared, the perspective of the analysis, which costs and outcomes to consider, the time horizon, and the target population (4,13–15). This information will aid in the identification of the full range of evidence using well-established methods of evidence synthesis. Drawing on current evidence to estimate such parameters allows the decision model to fulfill several requirements which cannot be accommodated in a single empirical study (4,14,16). In addition to evaluating all relevant alternatives, this includes the need to consider a standardized follow-up period and the ability to assess patient heterogeneity (4,14,16). Input parameters requiring compilation and synthesis will often include the relative effectiveness of alternative treatments, baseline risks for specific clinical events, resource use, unit costs, and estimates of health-related quality of life (4,13,14).

1.1.1.1 *Decision Rules*

The expected costs and effects estimated by a decision model can be used to inform a decision regarding a technology's widespread adoption. Pairwise comparisons are needed to determine the cost to gain an extra QALY by choosing a more effective alternative (4,13). This statistic is known as the incremental cost-effectiveness ratio (ICER; Equation 1) (4,13).

Equation 1. Incremental Cost-Effectiveness Ratio

$$\text{ICER} = \frac{\text{Cost}_2 - \text{Cost}_1}{\text{Effect}_2 - \text{Effect}_1} = \frac{\Delta\text{Cost}}{\Delta\text{Effect}}$$

Any assessment of cost-effectiveness will involve a determination of whether the incremental effects justify the incremental costs. This choice is clear when a technology is more effective and less costly than the next best option. Likewise, one may choose to reject a technology when it is less effective and more expensive (4). However, the decision becomes much harder to make when an alternative offers incremental benefits at some additional cost. In a resource constrained health system, this cost represents health loss borne by patients with competing claims on the same resources. Therefore, the maximum amount the decision maker is willing to pay for an additional unit of effect (QALY) represents the cost-effectiveness threshold (λ) (4,17). A decision to adopt an alternative will then be based on whether the ICER is less than the value of this threshold, expressed in the inequality in Equation 2.

Equation 2. ICER Decision Rule

$$\frac{\Delta\text{Cost}}{\Delta\text{Effect}} < \lambda$$

However, the ICER is not well suited for decision problems which include more than two mutually exclusive alternatives. While the decision rule can still be used in such circumstances, the calculation of each ICER will need to be made using appropriate pairwise comparisons. This will require the identification and exclusion of strongly and extendedly dominated alternatives. The complex procedures of conducting an incremental analysis can be avoided by placing the costs, effects, and cost-effectiveness threshold on a common scale (4,13,18,19). Direct comparisons of each alternative can be expressed in terms of Net Health Benefit (NHB; Equation 3) or Net Monetary Benefit (NMB; Equation 4). Under a net-benefit framework, the adoption

decision will involve selecting the alternative which offers the maximum expected net-benefit (Equation 5). This will return the same conclusions about cost-effectiveness as the ICER decision rule following the exclusion of dominated alternatives.

Equation 3. Net Health Benefit

$$\mathrm{NHB}_j = \mathrm{Effect}_j - \frac{\mathrm{Cost}_j}{\lambda}$$

Equation 4. Net-Monetary Benefit

$$\mathrm{NMB}_j = \lambda \times \mathrm{Effect}_j - \mathrm{Cost}_j$$

Equation 5. Net-Benefit Decision Rule

$$\max_j \{E[NB(j)]\}$$

where: $j =$ mutually exclusive alternative considered in decision model.

$\mathrm{NB} =$ Net Benefit Statistic, either NMB or NHB

1.1.2 Decision Making Under Conditions of Uncertainty

Any decision regarding the use of a technology will need to be made with expected costs and effects. Estimation of these values will be subject to two distinct sources of uncertainty. Parameter uncertainty refers to the fact that the model inputs represent imprecise estimates of the true parameter values. Meanwhile, structural uncertainty refers to the impact that different assumptions or scientific judgements can have on the model results (20–22). An adoption decision will be based on incomplete information if a decision model is evaluated using the expected (mean) parameter values and default structural assumptions. This will be especially true for model structures with non-linear relationships between inputs and outputs.

Expected costs and effects can be estimated using data generated from the probabilistic evaluation of a decision model (13,21–23). Monte Carlo simulation is the most common approach to characterize uncertainty in this manner. It involves the repeated evaluation of the

decision model using randomly sampled parameter values (13,24). Assumptions regarding the type of distribution to assign to each parameter should be determined by its respective type (probabilities, relative risk, costs, and utilities) (13). In most situations, the impact of structural uncertainty will be explored through separate scenario analyses (20–22). By generating distributions of cost and effect for each specified alternative, the model results will reflect the full range of values each parameter will be likely to take (21,22,24). From this the ICER can be calculated from the expected (mean) costs and benefits for each alternative. Meanwhile, expected net-benefits must be calculated from distributions of net-benefit for each value of the cost-effectiveness threshold.

While the uncertainty will affect how the decision model is evaluated, it has no impact on how the adoption decision should be made. Under Bayesian Decision Theory, the alternative which offers the maximum expected net benefit should be selected (18). However, the nature of the uncertainty implies there is a risk the adoption decision will be incorrect. If proven to be true, the opportunity cost from this decision uncertainty will be expressed as a health loss to current and future patients (18). Value of Information (VoI) methods offer a mechanism to determine if reducing future decision uncertainty through the acquisition of additional information could be a worthwhile activity (18,25–27). This suggests that there are two decisions to be made: i) whether current evidence supports adopting the technology; and ii) whether further evidence is required to support this decision in the future (13,18).

1.1.2.1 *Value of Information Methods*

Value of Information (VoI) analysis is an approach to justify commissioning additional research for a specific decision problem (6,7). This methodology values the return on investment in future research to reduce decision uncertainty (19,25,26,28). The statistics used to inform a

research decision will be the expected value of perfect information (EVPI), the expected value of partial perfect information (EVPPI), the expected value of sample information (EVSI) and the expected net benefit of sampling (ENBS) (13,19,25,26,28).

Due to the different sources of uncertainty in the decision model, there will be some probability that the adoption decision is wrong. However, with perfect information there would be no decision uncertainty. As a result, there would also be no opportunity cost (in terms of health to current and future patients) from making the wrong decision. This expected loss from uncertainty represents the EVPI. It is equivalent to an expected gain from reducing the uncertainty. Detailed in Equation 6, the EVPI statistic can be calculated from the Monte Carlo simulation output which informed the adoption decision (19,25). To calculate the upper boundary for all research into a decision problem, the population-level statistic will be required (Equation 7) (19,25). Additional research will not be worthwhile if the population EVPI is greater than the cost of that research. In the event this inequality is reversed, it will be necessary to determine the parameter(s) for which more research would be worthwhile.

Equation 6. Expected Value of Perfect Information

$$\text{EVPI} = \text{E}_{\theta}\left\{\max_j [NB(j,\theta)]\right\} - \max_j \{\text{E}_{\theta}[NB(j,\theta)]\}$$

where: $j =$ alternative

$\theta =$ uncertain parameters

Equation 7. Population EVPI

$$\text{Pop. EVPI} = \text{EVPI} \times \sum_{t=0}^{T} \frac{I_t}{(1+r)^t}$$

where: $T =$ effective lifetime of a technology

$I_t =$ disease incidence at time t

$r = \text{discount rate}$

The EVPPI, Equation 8, provides an upper boundary for research to reduce uncertainty for a specific (group) of parameter(s) (19,25,26). As with the EVPI, it is the difference between the expected value of a decision made with perfect and current information. Therefore, the EVPI can be viewed as a special case of the EVPPI where the parameters of interest are all the parameters in the model (19,25,26). However, the output of the original Monte Carlo simulation cannot be used to calculate EVPPI according to Equation 8. Capturing the effect of the parameter(s) of interest will typically involve a nested Monte Carlo simulation. For each random draw of the parameter(s) of interest (φ), the decision model will be repeatedly evaluated over many random draws for the remaining parameters (ψ) (19,25,26). Following calculation of EVPPI, the population EVPPI (Equation 9) can be used to determine if research for the parameter(s) of interest will be worthwhile (19,25,26). A major limitation of this approach to calculating EVPPI is the time-consuming nature of the nested Monte Carlo simulations. Recent developments in VoI methodology have introduced non-parametric regression-based approaches to evaluate the EVPPI without re-running the model (29,30).

Equation 8. Expected Value of Partial Perfect Information

$$\text{EVPPI}_{\varphi} = \text{E}_{\varphi}\left\{\max_j\left[\text{E}_{\psi|\varphi} NB(j,\varphi,\psi)\right]\right\} - \max_j\{\text{E}_{\theta}[NB(j,\theta)]\}$$

where: $\varphi = \text{parameter(s) of interest;}$

$\psi = \text{remaining uncertain parameters;}$

$\theta = \text{all uncertain parameters: } \psi \text{ and } \varphi$

Equation 9. Population EVPPI

$$\text{Pop.EVPPI}_{\varphi} = \text{EVPI}_{\varphi} \times \sum_{t=0}^{T} \frac{I_t}{(1+r)^t}$$

The EVSI represents the sufficient condition regarding the commissioning of a particular data collection exercise (19,26–28,31). As illustrated in Equation 10, the statistic measures the expected reduction in expected loss from information acquired in a study of sample size n. Generating distributions of cost and effect will require a different implementation of the nested Monte Carlo method described for EVPPI (27,28,32). For each random draw of the parameter(s) of interest (φ), a pre-posterior distribution will need to be defined which combines the predicted results of a trial with sample size n (D) with the prior value of the target parameter(s) (φ). Randomly sampled values from the pre-posterior distribution (D|φ) and the remaining uncertain parameters (ψ) will then be used in the repeated evaluation of the decision model. As with the EVPPI process, the simulation-based method for EVSI will be vulnerable to long execution times (30,33). In response, alternate approaches to the estimation of EVSI have been developed and published. Prominent examples include regression-based methods, importance sampling, Gaussian approximation, and moment matching (30,33). Web-based applications like the Sheffield Accelerated Value of Information tools have been developed to promote the implementation of these methods (29,30,34).

Equation 10. Expected Value of Sample Information

$$\text{EVSI}_{\varphi,n} = E_D\left\{\max_j\left[E_{\varphi|D}NB(j,\varphi)\right]\right\} - \max_j\{E_\theta[NB(j,\theta)]\}$$

where: $\varphi = \text{single uncertain parameter}$

$\theta = \text{all uncertain parameters}$

$D = s\text{ampled value of } \theta \text{ from trial with sample size } n$

Specific research will represent an efficient use of resources if the EVSI for a population of current and future patients is less than the expected costs of the research. In other words, a positive ENBS (Equation 11) indicates that the benefits of a proposed study outweigh its costs

(19,26–28,31). On the other hand, a negative ENBS will indicate that current evidence is insufficient for decision making. A thorough assessment of this criteria will therefore involve calculating the EVSI and ENBS for a range of study designs. In the event a positive ENBS is realized for multiple designs, the study which offers the maximum ENBS should be commissioned. The results from this research will then be used to update the decision model to consider new adoption and research decisions (22,27,28).

Equation 11. Expected Net Benefit of Sampling

$$\mathrm{ENBS}_{\varphi,n} = \mathrm{EVSI}_{\varphi,n} - \mathrm{TC}_{\varphi,n}$$

where: $\mathrm{TC}_{\varphi,n} = \text{total cost for research of parameter } \varphi \text{ with sample size } n.$

1.2 The HTA Cycle

Altogether, the use of economic evaluation within the process of HTA should reflect a continuous cycle of evidence gathering and decision making (19,31–33). As illustrated in Figure 1-1, the specific decision problem will direct the assembly of current evidence as well as a decision model. Afterwards, Monte Carlo simulation will be used to determine if current evidence supports the adoption of the health technology. Subsequently, Value of Information analysis will inform whether the decision uncertainty warrants the commissioning of additional research. The process will end if new information is unlikely to change the adoption decision. If the opposite proves true, new research will be commissioned for the parameters which had the greatest impact on decision uncertainty. Once the commissioned research is complete, the findings will be incorporated into an updated systematic review and decision model and the process will be repeated (19,31–33).

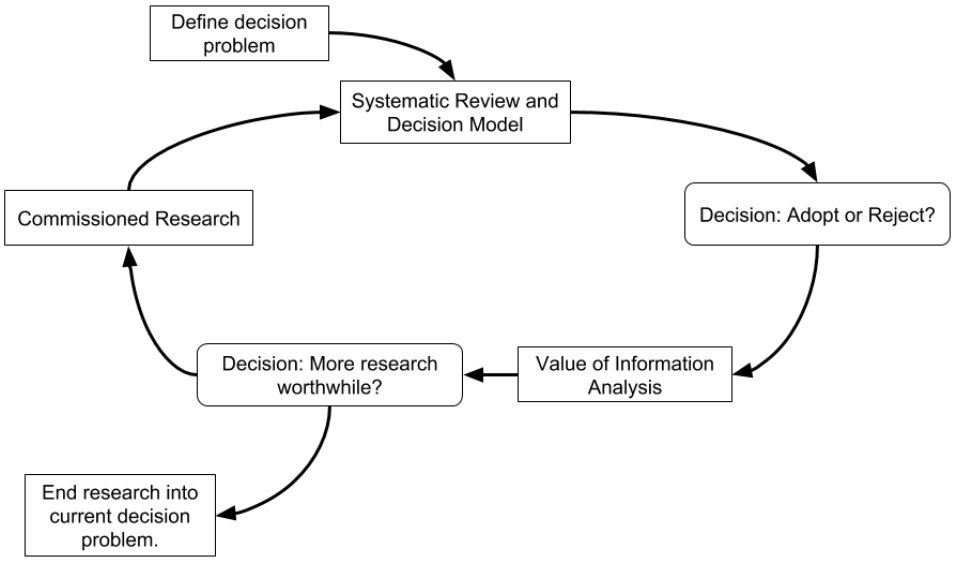


Figure 1-1. Health Technology Assessment Cycle. Adapted from Wilson and Abrams (19)

New information identified from the VoI analysis will not hold its value forever. This can occur because of three different factors. First, there may be a change in the effective price of a technology. Second, a new technology may be introduced requiring an expansion of the mutually exclusive alternatives (and the supporting evidence) included in the decision model. Third, new information related to the model inputs or structure may be published in advance of, or shortly after, an adoption decision. These issues have raised concerns about assumptions relating to the effective lifetime of a technology in VoI analysis (34). In addition, policy frameworks have been proposed which use the decision uncertainty to consider coverage with evidence development schemes (21,35,36).

One methodological issue which has not received much attention is the challenge of keeping a systematic review and economic evaluation up to date. This is particularly relevant given that much work in HTA happens when the evidence base for a technology is least mature (25). While the literature on this topic is scarce, there appears to be a consensus that most models are developed for a single analysis and are rarely updated. When updates do happen, they

typically involve a duplication of the original effort (37–44). This challenge has been acknowledged in the evidence synthesis literature as well. In response, organizations like the Cochrane Collaboration have been promoting living systematic reviews which attempt to incorporate new evidence as soon as it becomes available (45,46). While efforts have been made to automate some review tasks, the resulting efficiency gains have been poorly described (45,46). The persistence of this challenge suggests that the costs incurred to update an existing systematic review and decision model could be allocated to other activities. Moreover, the delayed decisions which result from inefficient iteration risk health losses to patients and increased system costs.

1.3 Reproducibility

The duplicative and inefficient nature of HTA iteration can be attributed to a lack of emphasis on reproducibility in its initial development. When defined narrowly, reproducibility refers to an ability to obtain consistent computational results given the same set of files and processes (47,48). More broadly, it reflects a strategic approach to directly validate, repeat, improve, and potentially re-purpose part (or all) of a project (47,48). This conceptualization of reproducibiltiy is distinguished from replication in terms of the availability of the original computer files to generate the results. Furthermore, it treats reproducibility as a matter of degree – with a project's level of reproducibility ranging from irreproducible to exactly reproducible (47,48). This suggests an HTA's level of reproducibiltiy will be determined by the ability to re-use a specific version of a decision model and the efficiency with which new information can be incorporated within the model's structure or input parameters. Therefore, the consensus surrounding the duplicative and inefficient nature of HTA iteration may indicate that HTAs are consistently developed to sub-optimal levels of reproducibiltiy.

The principle of reproducibility has been a long-standing component of science. However, interest in the practice of reproducibility only began to gain traction with the widespread use of computing in scientific research. In fact, the first documented use of the phrase “reproducible research” was in relation to the way a seismology lab used computer files to efficiently re-create scientific artifacts (tables and figures) for inclusion in thesis or journal manuscripts (47,48). Today, computers are used in almost every part of scientific research – data collection, data analyses of varying degrees of sophistication, and manuscript development. Despite this ubiquity, there has been a lack of training in the effective use of computing in the research process across many scientific fields (47,48). As a result, the production of research at sub-optimal levels of reproducibility is not exclusive to HTA.

Growing interest in reproducibility has led to the emergence of literature relevant to academics, policy makers, and trainees. Two of the most prominent examples of such efforts include a book of reproducible case studies (47) as well as a report commissioned by the United States’ National Academies of Science, Engineering, and Medicine (48). Together these works summarize the features, challenges, and opportunities associated with reproducibility (47,48). The experiences documented in both works confirm that enhancing the reproducibility of HTAs can reduce the marginal cost of iteration. An emphasis on reproducibility will also complement ongoing efforts to improve transparency, such as open-source modelling (38–43,49–52). Both outcomes will depend on the implementation of computing strategies which capture the provenance of the results in machine- and human-readable formats (47,48). Thus, efforts to improve transparency through reproducibility and open science can be distinguished in terms of the benefits offered to the original author.

As of this writing, there are two primary barriers to the production of reproducible HTAs. First, the specific computing strategies which require implementation have not been well described for the HTA context. For example, the recommended strategies from the book of case studies were targeted to an audience with some expertise in programming and data management. Fortunately, descriptions of factors which can affect reproducibility can guide the classification of strategies applied in the reproducible research literature (47,48). Second, efforts to enhance the reproducibility of HTAs will represent a significant behavioural change in research practice. As a result, the impact of the identified strategies on the methods and process of HTA, including the characterization of uncertainty, will need to be explored.

1.4 Objectives

The objective of this thesis project was to explore how an emphasis on reproducibility can support the effective development and maintenance of Health Technology Assessments. Specific research objectives included:

1. To identify relevant computing strategies from the reproducible research literature which can be adopted to enhance the reproducibility of Health Technology Assessments. (Chapter 2)
2. To explore the procedural and methodological impacts from the implementation of the reproducible research strategies using a case study of Biologic Treatment for Psoriatic Arthritis. (Chapter 3)
3. To describe how an emphasis on reproducibility can enable the consistent and thorough characterizations of uncertainty in economic evaluations of Health Technologies. (Chapter 4)

1.5 References

1. Luce BR, Drummond M, Jönsson B, Neumann PJ, Schwartz JS, Siebert U, et al. EBM, HTA, and CER: Clearing the Confusion. Milbank Q. 2010;88(2):256–76.

2. Banta D. The development of health technology assessment. Health Policy. 2003 Feb 1;63(2):121–32.

3. O'Donnell JC, Pham SV, Pashos CL, Miller DW, Smith MD. Health Technology Assessment: Lessons Learned from Around the World—An Overview. Value Health. 2009 Jun;12(s2):S1–5.

4. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. 461 p.

5. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Serv Res. 2013 Sep 30;13(1):371.

6. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada [Internet]. Ottawa, ON; 2017 Mar. Report No.: 4th ed. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition

7. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013. 94 p. (NICE Process and Methods Guides). Available from: http://www.ncbi.nlm.nih.gov/books/NBK395867/

8. Medical Services Advisory Committee. Technical Guidelines for Preparing Assessment Reports for the Medical Services Advisory Committee - Service Type: Investigative [Internet]. Canberra, Australia: Department of Health, Government of Australia; 2017 Jul. Available from: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/MSAC-Guidelines

9. McIntosh E. Using Discrete Choice Experiments within a Cost-Benefit Analysis Framework. PharmacoEconomics. 2006 Sep 1;24(9):855–68.

10. McIntosh E, Clarke P, Frew EJ, Louviere J, editors. Applied Methods of Cost-Benefit Analysis in Health Care [Internet]. Oxford, UNITED KINGDOM: Oxford University Press USA - OSO; 2010. Available from: http://ebookcentral.proquest.com/lib/ucalgary-ebooks/detail.action?docID=975570

11. Weinstein MC, Torrance G, McGuire A. QALYs: The Basics. Value Health. 2009 Mar;12:S5–9.

12. Bleichrodt H. QALYs and HYEs: Under what conditions are they equivalent? J Health Econ. 1995 May 1;14(1):17–37.

13. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006. 237 p. (Briggs A, Gray A, editors. Oxford handbooks in health economic evaluation).

14. Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, Drummond M. Evidence synthesis, economics and public policy. Res Synth Methods. 2010 Apr 1;1(2):126–35.

15. Stahl JE. Modelling Methods for Pharmacoeconomics and Health Technology Assessment. PharmacoEconomics. 2008 Feb 1;26(2):131–48.

16. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15(7):677–87.

17. Sculpher M, Claxton K. Establishing the Cost-Effectiveness of New Pharmaceuticals under Conditions of Uncertainty—When Is There Sufficient Evidence? Value Health. 2005 Jul;8(4):433–46.

18. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999 Jun 1;18(3):341–64.

19. Wilson E, Abrams K. From evidence-based economics to economics-based evidence: using systematic review to inform the design of future research. In: Shemilt I, Donaldson C, Marsh K, Mugford M, Vale L, editors. Event-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated; 2010.

20. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. Value Health. 2009 Jul 1;12(5):739–49.

21. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. PharmacoEconomics. 2008 Sep 1;26(9):781–98.

22. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 11: Characterizing, Reporting, and Interpreting Uncertainty. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. p. 353–88.

23. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? Value Health. 2006;9(4):244–52.

24. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation: A Practical Approach. Med Decis Making. 1985 Jun;5(2):157–77.

25. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research. PharmacoEconomics. 2006 Nov 1;24(11):1055–68.

26. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, et al. 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 Feb 1;23(2):139–50.

27. Wilson ECF. A Practical Guide to Value of Information Analysis. PharmacoEconomics. 2015 Feb;33(2):105–21.

28. Ades AE, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. Med Decis Making. 2004 Mar 1;24(2):207–27.

29. Strong M, Oakley JE, Brennan A. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach. Med Decis Making. 2014 Apr 1;34(3):311–26.

30. Tuffaha H. Value of Information Analysis: Are We There Yet? PharmacoEconomics Open. 2020 Aug 11;5(2):139–41.

31. Welton NJ, Thom HHZ. Value of Information: We've Got Speed, What More Do We Need? Med Decis Making. 2015 Jul;35(5):564–6.

32. Heath A, Baio G. Calculating the Expected Value of Sample Information Using Efficient Nested Monte Carlo: A Tutorial. Value Health. 2018 Nov 1;21(11):1299–304.

33. Kunst N, Wilson ECF, Glynn D, Alarid-Escudero F, Baio G, Brennan A, et al. Computing the Expected Value of Sample Information Efficiently: Practical Guidance and Recommendations for Four Model-Based Methods. Value Health. 2020 Jun 1;23(6):734–42.

34. Strong M, Breeze P, Thomas C, Brennan A. SAVI - Sheffield Accelerated Value of Information [Internet]. 2021 [cited 2021 Dec 6]. Available from: https://github.com/Sheffield-Accelerated-VoI/SAVI

35. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

36. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessent: the role of iterative decision analytic modelling. 2000 May; Available from: https://core.ac.uk/reader/6406440

37. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

38. Philips Z, Claxton K, Palmer S. The Half-Life of Truth: What Are Appropriate Time Horizons for Research Decisions? Med Decis Making. 2008 May 1;28(3):287–99.

39. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, et al. A Comprehensive Algorithm for Approval of Health Technologies With, Without, or Only in Research: The Key Principles for Informing Coverage Decisions. Value Health. 2016 Sep 1;19(6):885–91.

40. McKenna C, Soares M, Claxton K, Bojke L, Griffin S, Palmer S, et al. Unifying Research and Reimbursement Decisions: Case Studies Demonstrating the Sequence of Assessment and Judgments Required. Value Health. 2015 Sep 1;18(6):865–75.

41. Arnold RJG, Ekins S. Time for Cooperation in Health Economics among the Modelling Community: PharmacoEconomics. 2010 Aug;28(8):609–13.

42. Cohen JT, Neumann PJ, Wong JB. A Call for Open-Source Cost-Effectiveness Analysis. Ann Intern Med. 2017;167(6):432–3.

43. Cohen JT, Wong JB. Can Economic Model Transparency Improve Provider Interpretation of Cost-Effectiveness Analysis? A Response: Med Care. 2017 Nov;55(11):912–4.

44. Linthicum MT, Bright J, Shafrin J. Open-Source Tools for Value Assessment: A Promising Approach. J Clin Pathw. 2018 Jun 12;4(5):37–9.

45. 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 Nov 1;37(11):1313–20.

46. McQueen RB, Padula WV, Campbell JD. A Call for Open-Source Cost-Effectiveness Analysis. Ann Intern Med. 2018 Apr 3;168(7):528.

47. Jansen JP, Incerti D, Linthicum MT. An Open-Source Consensus-Based Approach To Value Assessment [Internet]. 2017. Available from: https://www.healthaffairs.org/action/showDoPubSecure?doi=10.1377%2Fhblog20171212.640960&format=full

48. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, et al. Transparency in Decision Modelling: What, Why, Who and How? PharmacoEconomics. 2019 Nov 1;37(11):1355–69.

49. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. J Clin Epidemiol. 2017 Nov 1;91:23–30.

50. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol. 2017 Nov 1;91:31–7.

51. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018. Available from: http://www.practicereproducibleresearch.org/

52. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019. Available from: https://www.nap.edu/catalog/25303

53. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. PharmacoEconomics. 2017 Jan 1;35(1):125–8.

54. Hatswell AJ, Chandler F. Sharing is Caring: The Case for Company-Level Collaboration in Pharmacoeconomic Modelling. PharmacoEconomics. 2017 Aug;35(8):755–7.

55. 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 Jun;37(6):829–43.

56. Wu EQ, Zhou Z-Y, Xie J, Metallo C, Thokala P. Transparency in Health Economic Modeling: Options, Issues and Potential Solutions. PharmacoEconomics. 2019 Nov;37(11):1349–54.

Chapter 2

Computing Strategies to Enhance the Reproducibility of Health Technology Assessments: A Systematized Review

Background

Health Technology Assessment (HTA) has been recognized as a continuous process requiring regular revision. However, updating previously commissioned work can be duplicative, leading to delayed decision-making. Experience in other fields has indicated that strategies for reproducible computing can reduce the cost of HTA iteration. This review identified computing strategies from the reproducible research literature which could be adopted in the HTA context.

Methods

Relevant literature was identified using snowball sampling and citation screening. The search strategy was initiated from two key reports that described the history, features, and challenges of reproducible research. Key quotes describing computing strategies were organized into specific domains known to influence reproducibility of a project. Qualitative content analysis was subsequently performed to generate the results.

Results

Twelve computing strategies were identified from 33 records representing a range of disciplines in the natural and social sciences. Collectively, they encouraged the preservation of data in a machine-readable format, which serve as a common reference point within an automated workflow. Recommendations included supplementing code and data with specific documentation and project organization strategies.

Conclusions

The reproducible research strategies identified in this review encourage the modification, rather than redevelopment of existing datasets and procedures. If implemented, the cost of updating HTA evaluations would be reduced, resulting in more effective use of resources to support health care decision-making. However, success will require individuals and groups to commit significant resources to develop the necessary computing skills.

2.1 Introduction

Health Technology Assessment (HTA), and the use of economic evaluation to support it, has been recognized as a continuous process of evidence gathering and decision-making (1–3). Economic evaluation is used to determine comparative cost-effectiveness and can also be used to determine whether any uncertainty associated with the evidence base and cost-effectiveness results warrant commissioning of additional research (1–8). Triggers for updating an HTA include the emergence of a new competitive technology or information likely to affect the decision (1–3,9). However, the process of updating a previously commissioned HTA has been consistently described as a duplication of the original effort (10–16). In addition to incurring research costs which could be spent elsewhere, delayed decisions risk increased health losses for patients and healthcare costs.

To improve the efficiency of HTA iteration, many have advocated for the release of computer files either on request or as open-source models (10–18). Given that these solutions seek to distribute rather than reduce labour costs, confidence in their effectiveness may be misplaced (19–21). For example, to update a previously commissioned HTA once new evidence becomes available, the original authors will need to commit resources to attract and retain new contributors (22,23). Such contributors will then familiarize themselves with the supplied computer files and choose whether to update them or redevelop the entire project – both time consuming tasks. In the event the new contributors abandon the update, the responsibility of completing the work will revert to the original authors.

Computing plays an essential and ubiquitous role in systematic reviews, meta-analyses, and economic evaluations. Yet current standards and training have yet to expand its use beyond the preparation or execution of specific simulations and analyses (6,24). Other fields have

resolved similar challenges through an emphasis on reproducibility. This refers to the ability to obtain consistent computational results, including tables and figures, given the same set of files and processes (25,26). It is viewed as a strategic approach to directly validate, repeat, improve, and potentially repurpose projects. As such, it is conceptualized as a spectrum from irreproducible to exactly reproducible (25–27).

In contrast, the HTA literature has treated reproducibility synonymously with replicability and as a means to promote or assess transparency (10–12,16–18,28–32). While they are strongly associated, transparency is about providing value to others while the primary beneficiary of reproducibility is one's future self (25,26). The experience with reproducible research in other scientific fields may offer a window into the use of computing beyond specific analytical techniques, thereby lowering the cost of iteration. The objective of the present study is to identify relevant computing strategies from the reproducible research literature which can be adopted to promote the effective development and maintenance of HTA.

2.2 Methods

2.2.1 Search Strategy and Eligibility

Relevant computing strategies from the reproducible research literature were identified using a systematized review. This method leveraged elements of the systematic review process to identify and organize included records. As a result, the systematized review is not meant to be as comprehensive as the systematic review (33). In the present study, record identification was facilitated by the development of a high-level conceptual framework using two key texts on reproducibility in science. These texts were selected because they both provided a comprehensive overview of the history, features, and impact of reproducible research (25,26). The purpose of the framework was to summarize the factors which can influence the

reproducibility of a project from the perspective of the original authors or an external audience. As detailed in Appendix A, the eight identified constructs included: (i) Provenance: Process; (ii) Provenance: Data; (iii) Provenance: Environment; (iv) Documentation; (v) Project Management; (vi) Open Reporting; (vii) Copyright; and (viii) Tool Selection (25,26).

Records which described at least one specific practice or strategy relevant to the conceptual framework were included in this review. Given the breadth of fields to which reproducibility may apply, potentially eligible records were identified using backwards and forwards citation screening of included records as part of a staged process. This bidirectional snowballing strategy offered a rigorous mechanism to identify a representative sample of the literature on reproducibility in science (33,34). A complementary database search was not considered due to the limited ability to define a search query with sufficient specificity and sensitivity. This is reflected by the findings of a review by McManus et al., which explored various definitions of replication, yet failed to identify its role as a form of reproducibility (30,33,34).

The search strategy began with the two records which informed the development of the conceptual framework (25,26). In each stage, candidate records were compiled from the reference lists (backward citation screening) and Google Scholar citations (forward citation screening) of items which satisfied the inclusion criteria from the preceding stage (33). Identified records were subsequently screened to exclude duplicates, items which had already been assessed, or those which were not relevant to this review. Record identification and eligibility assessment were conducted by a single investigator (D.W.) but were reviewed by three authors (D.W., G.H., and E.S.). Sampling continued until these same authors agreed that thematic saturation had been reached.

2.2.2 Data Extraction and Analysis

For each included record, every identified strategy was coded verbatim to one of the constructs listed in the analytic framework. To ensure consistency in the indexing process, specific definitions were created for each construct. Quotes that did not satisfy inclusion criteria were excluded.

A framework approach was used to conduct a qualitative content analysis of quotes indexed to a subset of factors influencing reproducibility (35). Construct selection for this activity was completed following data collection, but *prior* to the identification of key themes. Of the eight constructs available, qualitative analysis was restricted to quotes coded to *Provenance: Process, Provenance: Data, Provenance: Environment, Documentation, and Project Management.* The *Tool Selection, Open Reporting,* and *Copyright* constructs were beyond the scope of the present review.

The quotes associated with the remaining five constructs were analyzed inductively to identify emergent strategies relevant to the original authors of an HTA (35). For each construct, the indexed quotes were organized into groups according to their perceived thematic overlap. The main points of each group were then summarized into a single narrative to facilitate reporting. Results were validated by contrasting the emergent strategies against those used in the concurrent redevelopment of an HTA case study (details not reported here).

2.3 Results

A total of 33 records satisfied the inclusion criteria for this review (25,26,36–66). A summary of the number of records identified at each stage of the search strategy is presented in Figure 2-1. Thirty-two were published between 2009 and 2020 (25,26,36–66). The identified records represented a broad range of fields of study; however, the two most common fields were

statistics (37,40,47,52,62,66) and the biological sciences (39,43,45,46,49,53,54,57,58,61,63–65). Results from the qualitative synthesis yielded twelve strategies which can be used to enhance the reproducibility of HTAs as summarized in Figure 2-2. Complete descriptions of each strategy, along with supporting quotations, are included in the sub-sections below.

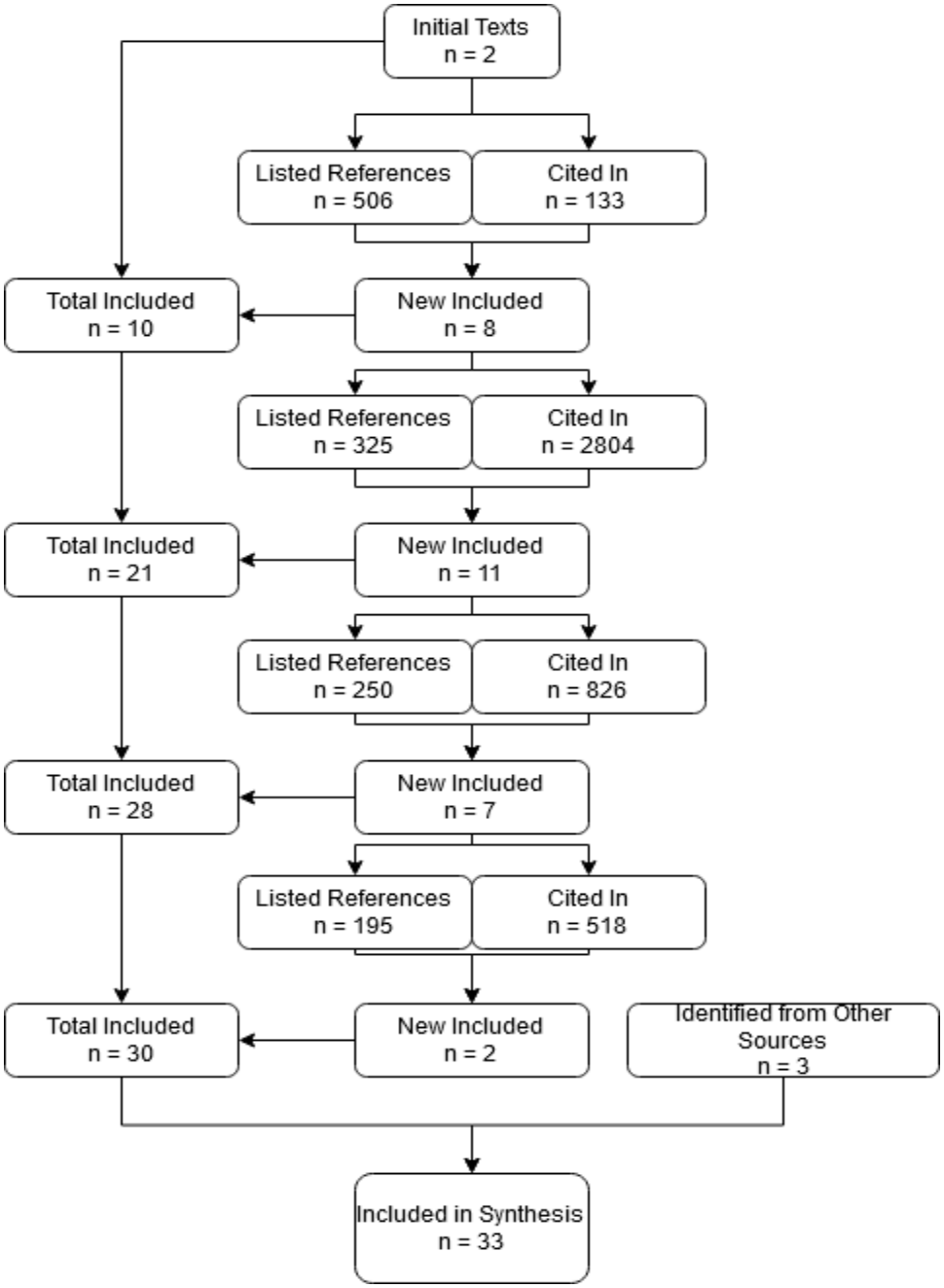


Figure 2-1. Details of identified reports and those selected for inclusion in this review.

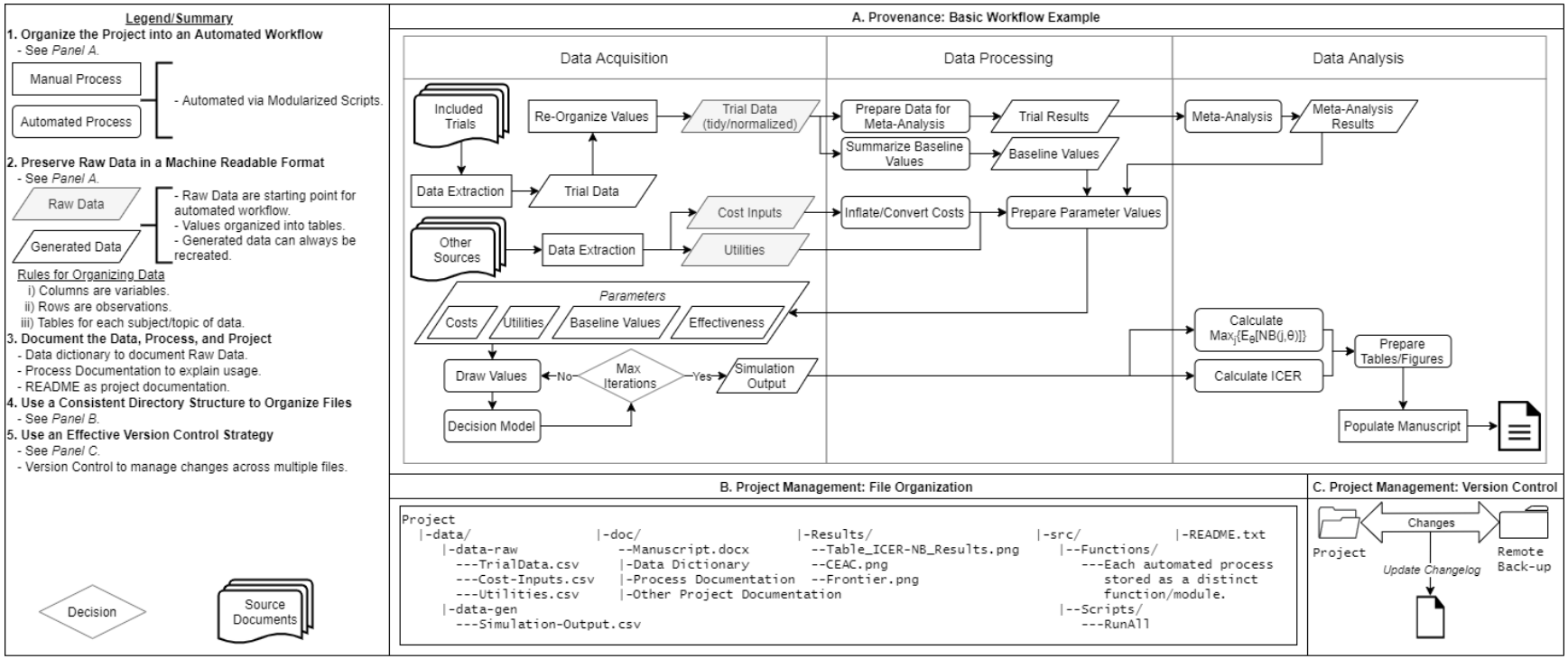


Figure 2-2. Summary of Identified Reproducible Research Strategies. Panel A: Schematic of a basic reproducible workflow for a systematic review and economic evaluation; Panel B: Recommended Project Directory Structure; Panel C: Version control recommendations.

2.3.1 Provenance: Process

2.3.1.1 *Organize the Project Process into a Workflow*

Reproducibility requires the re-execution of each procedure performed on a computer (i.e., "clicks" or commands) with minimal difficulty. This can be facilitated through the creation of a workflow, which refers to the sequence of steps required to obtain the desired outputs from a defined set of inputs (26,47). The intent of this strategy is to encourage the systematic, rather than *ad hoc*, record keeping of every operation performed to get from the starting point to the intended result (25,26,38,42,46,47,53,54,57–59,64).

One record recommended organizing each project task into a specific stage of a basic reproducible workflow (26). The data acquisition stage should capture each step needed to produce raw data, which could be acquired from primary or secondary sources or created from a simulation (26). The data processing stage should capture each step needed to prepare the raw data for simulation or analysis (26,41,48). Such wrangling or munging operations include, but are not limited to, sub-setting values, merging related tables, converting between "wide" and "long" layouts, and computing new columns (36,41,48,62). The analysis stage should capture all relevant statistical analyses and visualizations to prepare for another workflow operation or an answer to the research question (26) (Table 2-1).

2.3.1.2 *Write Scripts to Automate the Project Workflow*

To enable reliable execution of all or part of a workflow, each step should be encoded in one or more scripts (26,40,47,50,57). A script is a plain text file containing the instructions that a computer can read to complete a data acquisition/creation, processing, or analysis task. As scripts require the explicit statement of all assumptions and procedural steps for machine execution, they represent an accurate and complete record of the tasks performed by a computer. Therefore,

a workflow will also represent an editable and auditable record of the project methods (25,26,36,40,42,45,46,48,50,54,57,63,64). (Table 2-1).

2.3.1.3 *Develop Scripts Through Iterative Revision*

The benefits of scripts are maximized when they are written to consider two audiences: the computer and the human. For the computer, the syntax of the scripting language must be employed correctly so that the supplied instruction can be executed. For the human, the script must be written to allow a future collaborator who is familiar with the language, to read, navigate, and modify its contents. Following the development of an initial working script, two types of iterative revision were encouraged: *(i) Elimination of Duplication* and *(ii) Minimize Human Intervention* (39,41,48,54,63). (Table 2-1)

Elimination of Duplication

Duplication in scripts is often a strategy for iterative computation or as an attempt to expand the scope of an operation to include multiple datasets or variables (45). This redundancy can distort the intent of the operation and increases the risk of error as each occurrence must be managed independently (42). Modularization and abstraction were recommended to revise scripts and eliminate duplication. Modularization refers to the re-organization of code within scripts into single purpose programs called functions. Script clarity is improved by replacing multiple instances of the same code with a single command which accepts and returns specific input and output (26,39,40,42,54,57,58,63–66). Abstraction refers to the process of re-writing a script, even after modularization, to remove any duplicate instructions (64). Through the single specification of a task or sub-task, the code will be easier to read and modify should new requirements be identified in the future (36,42,45) (Table 2-1).

Minimize Human Intervention

Where possible, new scripts or functions should be created to automate a manually executed task (26,42,47,54,56,63–65). However, the decision to expand the scope of operations for a workflow should be determined by its state of reproducibility (26). For example, rather than downloading life expectancy data from a specific webpage, a national statistics agency may publish a protocol enabling researchers to write programs which can acquire the data directly from the server. In most circumstances, the benefits of further automation will quickly outweigh the costs of investing in re-useable commands. Furthermore, it will confer time savings on the project team since the computer will be able to execute the workflow much faster than a person could (42) (Table 2-1**)**.

Table 2-1. Supporting Quotes for Strategies Relevant to Provenance: Process

Organize the Project Process into a Workflow
• "Design a workflow as a sequence of small steps that are glued together with intermediate outputs from one step feeding into the next step as inputs. Met through overall workflow design especially a clear conceptualization of the different operations that need to occur sequentially and how they support each other." (26)
Write Scripts to Automate the Workflow
• "A script is a plain text file containing instructions composed in a programming language that direct a computer to accomplish a task. In a research context, researchers write scripts to do data ingest cleaning analysis visualization and reporting. By writing scripts a very high-resolution record of the research workflow is created and is preserved in a plain text file that can be reused and inspected by others." (50) • "Automation of the Research Process. Means that the main steps in the project (transformations of data processing and calculations visualizations) are encoded in software and documented in such a way that they can reliably and mechanically be replicated. In other words the conclusions and illustrations that appear in the article are the result of a set of computational routines or scripts that can be examined by others and re-run to reproduce the results." (26)
Develop Scripts through Iterative Revision
• "A happy medium often involves iterative improvement of scripts. An initial script is designed with minimal functionality and without the ability to restart in the middle of partially completed experiments. As the functionality of the script expands and the script is used more often it may need to be broken into several scripts." (54) • "A program should not require its readers to hold more than a handful of facts in memory at once. So programs should limit the total number of items to be remembered to accomplish

a task. The primary way to accomplish this is to break programs up into easily understood functions each of which conducts a single easily understood task. This serves to make each piece of the program easier to understand in the same way that breaking up a scientific paper using sections and paragraphs makes it easier to read." (63)

- "Turning the specific instances of something into a general-purpose tool. [Abstraction] is essential to writing good code for at least two reasons. First it eliminates redundancy which reduces the scope for error and increases the value you can get from the code you write. Second it makes code more readable. (42)
- "Pushing the boundaries of automation pays big dividends. The costs tend to be lower than they appear and the benefits bigger. A rule of research is that you will end up running every step more times than you think. And the costs of repeated manual steps quickly accumulate beyond the costs of investing once in a reusable tool." (42)
- "…automation pays big dividends. The costs tend to be lower than they appear and the benefits bigger. A rule of research is that you will end up running every step more times than you think. And the costs of repeated manual steps quickly accumulate beyond the costs of investing once in a reusable tool." (42)

2.3.2 Provenance: Data

2.3.2.1 *Preserve the Raw Data*

A dataset refers to a collection of values (numbers or characters) about a subject (62). In the included records, raw data was used as a relative designation to represent an unprocessed, unmodified dataset which contains all original values and observations. Raw data serve as a common reference point to initiate the workflow and act as a single authoritative source where the most up-to-date information about a topic can be accessed (40). In contrast, data returned from the other workflow stages can always be re-generated and only need to be preserved for the purposes of reporting (53,57,64). In recognition of the key role that raw data play in reproducibility, it is recommended that such datasets are preserved and treated as read-only (44,48,58,62,66).

While the read-only recommendation will protect the raw data from accidental changes, specific strategies will be necessary to satisfy the requirement that the raw data remain up to date. Deliberate changes to a dataset may be necessary for one of five reasons: (i) to correct

errors in formatting or values; (ii) to reconcile or replace missing values; (iii) to modify the classification system of a categorical variable; (iv) to record new observations; and (v) to incorporate emergent information requirements (44,48,58,62,66). Recommended procedures for the preservation of raw data differed according to the unique updating, validation, and storage requirements of each dataset (46,53,59). For example, data extracted from a specific observational study which is infrequently updated will require less maintenance than a dataset created from an ongoing systematic review of randomized controlled trials. Different considerations may be necessary for automated data acquisition/creation processes like simulation (40,46,53,57–59,61,64). (Table 2-2)

2.3.2.2 *Store Data in a Machine-Readable Format*

The integration of data into the workflow is enabled using text-based formats with open file specifications. Text-based formats are machine readable because the user can clearly describe the data and its formatting to a computer. In contrast, formats which supplement the data with information describing its display (colours, fonts, etc.) will make it more difficult to access and operate on potentially relevant information (25,26,46,51,58,61,64). Additionally, formats with open specifications (i.e., comma separated values; csv) are preferable since the data can be accessed without proprietary software or hardware. Therefore, storing data in text-based open formats can promote inter-operability across software and hardware systems. (Table 2-2).

2.3.2.3 *Create Datasets with Unambiguous Relationships Between Rows and Columns*

To promote the reliable storage and organization of tabular datasets, three rules were consistently promoted. First, each column should represent a single variable containing values which measure the same attribute. Second, each row should represent a single observation. Third, there should be one table for each subject/topic of data. In other words, the variables must

be attributes of the table's subject/topic. When multiple tables are required, an additional column is needed to uniquely identify each observation. Such values enable the creation intermediate tables for further processing or analysis. Examples which illustrate this approach to data organization are included in referenced works. The technical term for the application of these rules is "data normalization", and their application has received considerable attention as "tidy data" (40,42,46,61,62). (Table 2-2).

Table 2-2. Supporting Quotes for Strategies Relevant to Provenance: Data

Preserve the Raw Data
• "Consider the intrinsic value of the data. Observations of phenomena that cannot be repeated may need to be stored indefinitely. Data from easily repeatable experiments may only need to be stored for a short period. Simulations may only need to have the source code initial conditions and verification data stored." (53) • "Keep raw data raw: Since the analytical and data processing procedures improve or otherwise change over time having access to the "raw" (unprocessed) data can facilitate future re-analysis and analytical reproducibility." (46) • "Data errors and problems may include entry errors missing values duplicates outliers and data inconsistencies and discrepancies any of which may affect the validity reproducibility and thus the quality of studies." (66)
Store Datasets in a Machine-Readable Format
• "Always encode every piece of information about the observations using text. For example, if storing data in a spreadsheet and a form of coloured text or cell background formatting to indicate information about an observation then this information will not be exported (and will be lost!) when the data is exported as raw text. Every piece of data should be encoded as actual text that can be exported." (40) • "Data should be stored in a format that computers can use easily for processing. This is especially crucial as datasets become larger. Making data easily usable is best achieved by using standard data formats that have open specs or by using databases. These ensure interoperability facilitate re-use and reduce the chances of data loss or mistakes being introduced during conversion between formats." (46)
Create Datasets with Unambiguous Relationships between Rows and Columns
• "Tidy Data. A standard way of mapping the meaning of a dataset to its structure. A dataset is messy or tidy depending on how rows columns and tables are matched up with observations variables and types. In tidy data: 1) Each variable forms a column; 2) Each observation forms a row; 3) Each type of observable unit forms a table." (62) • "Useful for data to be structured in a way that makes use interpretation and analysis easy. Implemented through the use of "Codd's 3rd normal form" also known as tidy data. Duplication of information is reduced and it is easier to subset or summarize the dataset to include the variables or observations of interest." (46)

2.3.3 Documentation

2.3.3.1 *Document the Data, Process, and Project*

Documentation should communicate the purpose, behaviour, or content of the computer files created for a project. Intended exclusively for humans, it will be most useful when maintained in parallel with its machine-readable counterpart (26,39,42,43,49,53,54,63,64). The included records recommended comments for scripts and usage documentation for functions. A comment is a statement embedded in a script that the computer will ignore, which succinctly communicates the behaviour or purpose of a section of code (26,49,56,63). In contrast, usage documentation explains how a function can be used to return its output. It should describe what the function does, the inputs and their default values, the content and structure of the output, and provide a working example (49,54). A “data dictionary” was recommended for raw data not delivered by a program as such datasets will contain information which require further explanation (26,36,37,40,45,47,50,53,56,61,66). Lastly, the creation of a README file was recommended to serve as the project homepage. Suggested content included a summary of the project objectives, information to orient new collaborators, the expected level of reproducibility, and a list of active contributors (26,43,49,53,54,56,63). (Table 2-3).

2.3.3.2 *Use Short, But Meaningful Names for Files, Functions and Variables*

Naming is a form of documentation because it serves as a declarative statement of an object's meaning or function. Thus, choosing good names will make a script, program, dataset, or file easier to read and understand. The main principle is to choose names which are short, distinctive, and meaningful. This is reflected in four recommended practices. First, a name of an object should reflect its scope. For example, while naming a loop counter “i” or “j” is clear, the same cannot be said for a dataset or variable named "data1" or "x1". Second, special characters

should be avoided, except for hyphens and underscores which should be used as substitutes for spaces. Third, lengthy explanations are best suited for the relevant documentation. Lastly, a name should never include the word "final", as it will eventually be ignored (37,39,40,42,63,64). (Table 2-3).

Table 2-3. Supporting Quotes for Strategies Relevant to Documentation

Document the Project, Process, and Data
• "Don't write documentation that you will not maintain. Because you don't have to keep documentation up to date for code to work or for results to be right internal consistency becomes a problem. It is tempting to make improvements to the code without making parallel improvements to the comments only to find later that your comments are confusing or misleading. To avoid this, you will need to keep your comments up to date. But if its not worth maintaining a piece of documentation up to that standard it probably isn't worth writing it in the first place." (42) • "Data tables do not necessarily display all the variables needed to figure out what makes each row unique. For such information you sometimes need to look at the documentation of how the data were collected and what the variables mean." (36) • "Write Comments as you code. Comments are the single most important aspect of software documentation. Although it may be perfectly obvious to you what your code does without comments other readers will likely not be so fortunate. Indeed, you yourself may not even be able to understand your own code after you've moved on to another project." (49) • "Every script or program no matter how simple should be able to produce a fairly detailed usage statement that makes it clear what the inputs and outputs are and what options are available." (54) • "Your README file acts like a homepage for your project…A good rule of thumb is to assume that the information contained within the README will be the only documentation your users read. For this reason, your README should include how to install and configure your software where to find its full documentation under what license it is released how to test it to ensure functionality and acknowledgements. You should include your quick start guide in your README." (49)
Use Short, but Meaningful Names for Files, Functions and Variables
• "Give functions and variables meaningful names. Mechanism to document their purpose and to make the program easier to read. As a rule of thumb the greater the scope of a variable the more informative its name should be; while its acceptable to call the counter variable in a loop i or j things that are reused often like major data structures should not have 1-letter names." (64) • "As a general rule, do not use spaces either in variable names or in file names. They make programming harder. Where you might use spaces use underscores or perhaps hyphens - whatever you choose pick one and be consistent. Avoid special characters expect for underscores and hyphens." (37)

2.3.4 Project Management

2.3.4.1 *Organize Projects with a Consistent Directory Structure*

The included records recommended organizing all project files within a single project folder. Files within this folder should be accompanied by the README, as well as additional sub-directories to organize the files based on their content. Frequent examples included documents ("doc"), data ("data"), results ("results"), and scripts ("src") to store text files which define and apply each function (26,42,54,56,57,64) (Table 2-4).

2.3.4.2 *Use an Effective Version Control Strategy*

Unlike the "track changes" feature of a word processor, version control is meant to manage revisions across multiple related files. Five common characteristics were identified from the manual and software-based version control strategies promoted in the included records. First, a "version" should be conceptualized as a group of edits to a file or group of files which can be instantly reverted (42,64). For example, a meaningful change may involve correcting an error in the source code or data. Second, a version should always allow for the correct execution of the workflow. Any change has the potential to distort, or interrupt workflow behaviour and failing to revise dependent operations risks undermining reproducibility in the event the error goes unnoticed (26,39,42,47,57,65). Third, a changelog should be maintained to serve as a historical record of the project's development. Such documentation provides a centralized location to summarize the changes made for each version (48,49,64). Fourth, version control should always promote a single authoritative project directory. Revisions should be applied to existing files, rather than creating a copy of a file and then modifying its contents. In addition to populating a project directory with ambiguously named copies of the same file (i.e., "analysis_02" and "analysis_new") the copy-and-modify approach will distort the provenance of the workflow

(47,63). Lastly, it should include a form of back-up which mirrors the project directory to a remote location and synchronizes the changes on a regular basis. Examples include an external hard drive, cloud storage, or a version control service (46,54,64) (Table 2-4).

Table 2-4. Supporting Quotes for Strategies Relating to Project Organization

Organize Projects with a Consistent Directory Structure
• "Someone unfamiliar with your project should be able to look at your computer files and understand in detail what you did and why. This "someone" could be: i) someone who read your article and wants to try to reproduce your work ii) a collaborator who wants to understand the details of your experiments iii) a future student working in your lab who wants to extend your work after you have moved on to a new job; iv) your research advisor; v) Your future self!" (54)
Use an Effect Version Control Strategy
• "A general problem: anytime you have more than one representation of the same information you run the risk that the two will someday come into conflict. In the best-case scenario, you will need to do some work to untangle the mess. In the worst case scenario, your results will be wrong or internally inconsistent." (42) • "Use a Version Control System. A VCS stores snapshots of a project's files in a repository. Programmers can modify their working copy of the project at will then commit changes to the repository when they are satisfied with the results to share them with colleagues." (63)

2.4 Discussion

This review identified a series of computing strategies which can enhance the reproducibility of HTAs. Findings revealed that the duplicative nature of HTA iteration may be attributable to an over-reliance on human intervention for tasks which a computer could execute faster and more reliably. In contrast, the identified strategies encourage the preservation of raw data which serve as a common reference point in an automated workflow. Duplication is avoided by leveraging human intervention to modify rather than recreate each raw dataset or script.

The benefits can be illustrated using the example of adding a new trial of an existing comparator to the sequence of systematic review, meta-analysis, and decision model. As presented in Figure 2-2 (Panel A), a workflow could begin in the data acquisition stage where the

existing trial data are updated to include values extracted manually from the new study. In the data processing stage, a query would be executed to isolate the specific observations and columns from the trial data required for the meta-analysis. Updated effectiveness values would then be collected with the other parameters to be passed into the probabilistic simulation of the decision model. The process of passing output from one function into the next as input would continue until estimates of cost-effectiveness are returned. If one were to make procedural changes, revisions could be applied to the relevant modular scripts. Following methodological and behavioural validation, generating updated results could be as simple as executing a "run all" command.

In practice, it may not be feasible to capture the provenance of the results from the rawest form of data. For example, workflow design will be impacted by confidentiality requirements when raw data contain personally identifiable information. Likewise, the division of responsibility in a particular project may result in the creation of independent workflows for systematic review and economic evaluation. Additional coordination will be needed to develop a single integrated workflow or to provide the economic evaluation with read-only access to a persistent location storing the most up-to-date results. The absence of such coordination will narrow the scope of operations captured by the workflow and restrict the extent to which reproducibility can reduce duplication in future iterations.

Efficiency gains in HTA iteration will be further restricted to those who have read and write access to the files which constitute the reproducible workflow. Given that reporting standards have yet to mandate file sharing, it remains unlikely that independent investigators will be able to iterate a previous (reproducible) HTA more efficiently than the original authors (25,28,55). While investments in reproducibility will complement ongoing efforts to improve

transparency, file sharing may not be sufficient to extend the reproducible capability to a broader set of contributors with different interests, expertise, and skills (50,55,59). Strategies for reproducibility which apply to the *Open Reporting*, *Copyright*, and *Tool Selection* constructs of the conceptual framework remain an important area for future research.

Ultimately, enhancing the reproducibility of a project will depend on the development of computing skills which are not included in HTA training curricula (6,21,24). The contribution of the present review is therefore restricted to raising awareness towards a collection of strategies which could be adopted by interested readers (21,25,26,63,64,67). Self-directed skill development can be aided further using materials produced for Data Carpentry workshops (68). It is recommended that an incremental approach be applied to the adoption of each strategy, including the automation of a small but increasingly complex sequence of manually executed tasks. This exercise may help clarify the requirements considered in the selection of a computing tool to support HTA (21,69–71). Lastly, it would be unreasonable to expect all contributors to become experts in data management or programming. However, familiarity with the underlying concepts will foster more effective collaborations (45,67).

2.4.1 Limitations

The findings of this review should not be treated as an exhaustive or complete characterization of all practices which may influence reproducibility. Strategies relating to the computing environment were not reported due to their technical complexity (25,26,40,50,56,57). Moreover, the *Tool Selection*, *Open Reporting* and *Copyright* constructs represent factors which can influence the independent reproduction previously published research. In addition, the comprehensive nature of the texts used to create the conceptual framework does not eliminate the possibility that some factors which can impact reproducibility were overlooked (25,26). The

creation of a comprehensive framework to summarize and critically appraise reproducible research strategies presents a unique inter-disciplinary research opportunity which deserves pursuit. Lastly, the authors acknowledge that reliance on a single reviewer for record identification, data extraction and analysis may be a risk to the validity of the results. The impact of this methodological choice was mitigated using deductive coding of quotes to constructs identified *a priori* and the maintenance of an audit trial to ensure all results were data driven.

2.5 Conclusion

The process of HTA requires computing to generate and assemble evidence to inform health policy decisions. While the consideration of new information is an ongoing requirement, it often involves a duplication of the original effort. Findings from this review suggest that this may be a sign of poor digital record keeping by the original authors. A series of strategies were identified to better manage the provenance of the results through the deliberate preservation of data which serve as a common reference point in an automated workflow. Efficient iteration is enabled by encouraging the modification rather than redevelopment of existing datasets and procedures. While straightforward in principle, realization of the benefits from reproducibility will be restricted to individuals and groups who invest in the development of the requisite computing skills.

2.6 References

1. Wilson E, Abrams K. From evidence-based economics to economics-based evidence: using systematic review to inform the design of future research. In: Shemilt I, Donaldson C, Marsh K, Mugford M, Vale L, editors. Event-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated; 2010.

2. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

3. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

4. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Serv Res. 2013 Sep 30;13:371.

5. Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, Drummond M. Evidence synthesis, economics and public policy. Res Synth Methods. 2010 Apr 1;1(2):126–35.

6. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006. 237 p. (Briggs A, Gray A, editors. Oxford handbooks in health economic evaluation).

7. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research. PharmacoEconomics. 2006 Nov 1;24(11):1055–68.

8. Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. PLOS Med. 2015 Sep 29;12(9):e1001882.

9. National Institute for Health and Care Excellence. Guide to the Process of Technology Appraisal [Internet]. National Institute for Health and Care Excellence; 2018 May p. 101. Report No.: PMG19. Available from: https://www.nice.org.uk/process/pmg19/chapter/acknowledgements

10. Cohen JT, Neumann PJ, Wong JB. A Call for Open-Source Cost-Effectiveness Analysis. Ann Intern Med. 2017;167(6):432–3.

11. Cohen JT, Wong JB. Can Economic Model Transparency Improve Provider Interpretation of Cost-Effectiveness Analysis? A Response: Med Care. 2017 Nov;55(11):912–4.

12. 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 Nov 1;37(11):1313–20.

13. McQueen RB, Padula WV, Campbell JD. A Call for Open-Source Cost-Effectiveness Analysis. Ann Intern Med. 2018 Apr 3;168(7):528.

14. Linthicum MT, Bright J, Shafrin J. Open-Source Tools for Value Assessment: A Promising Approach. J Clin Pathw. 2018 Jun 12;4(5):37–9.

15. Jansen JP, Incerti D, Linthicum MT. An Open-Source Consensus-Based Approach To Value Assessment [Internet]. 2017. Available from: https://www.healthaffairs.org/action/showDoPubSecure?doi=10.1377%2Fhblog20171212.640960&format=full

16. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, et al. Transparency in Decision Modelling: What, Why, Who and How? PharmacoEconomics. 2019 Nov 1;37(11):1355–69.

17. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. PharmacoEconomics. 2017 Jan 1;35(1):125–8.

18. Wu EQ, Zhou Z-Y, Xie J, Metallo C, Thokala P. Transparency in Health Economic Modeling: Options, Issues and Potential Solutions. PharmacoEconomics. 2019 Nov;37(11):1349–54.

19. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada [Internet]. Canadian Agency for Drugs and Technologies in Health; 2017. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition

20. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013. 94 p. (NICE Process and Methods Guides). Available from: http://www.ncbi.nlm.nih.gov/books/NBK395867/

21. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YZ, Yang A, et al. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. PharmacoEconomics [Internet]. 2019 Sep 24; Available from: http://link.springer.com/10.1007/s40273-019-00837-x

22. Prlić A, Procter JB. Ten Simple Rules for the Open Development of Scientific Software. PLOS Comput Biol. 2012 Dec 6;8(12):e1002802.

23. Sholler D, Steinmacher I, Ford D, Averick M, Hoye M, Wilson G. Ten simple rules for helping newcomers become contributors to open projects. PLOS Comput Biol. 2019 Sep 12;15(9):e1007296.

24. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. 461 p.

25. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019. Available from: https://www.nap.edu/catalog/25303

26. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018. Available from: http://www.practicereproducibleresearch.org/

27. Peng RD. Reproducible Research in Computational Science. Science. 2011 Dec 2;334(6060):1226–7.

28. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and Facilitators to Model Replication Within Health Economics. Value Health. 2019 Sep 1;22(9):1018–25.

29. 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 Sep 1;32(5):733–43.

30. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. PharmacoEconomics. 2019 Sep 18;

31. Smith-Spangler CM. Transparency and Reproducible Research in Modeling: Why We Need It and How to Get There. Med Decis Making. 2012 Sep 1;32(5):663–6.

32. Catalá-López F, Caulley L, Ridao M, Hutton B, Husereau D, Drummond MF, et al. Reproducible research practices, openness and transparency in health economic evaluations: study protocol for a cross-sectional comparative analysis. BMJ Open. 2020 Feb 1;10(2):e034463.

33. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J. 2009;26(2):91–108.

34. Hinde S, Spackman E. Bidirectional Citation Searching to Completion: An Exploration of Literature Searching Methods. PharmacoEconomics. 2015 Jan 1;33(1):5–11.

35. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering [Internet]. New York, NY, USA: Association for Computing Machinery; 2014. p. 1–10. (EASE '14). Available from: https://doi.org/10.1145/2601248.2601268

36. Forman J, Damschroder L. Qualitative Content Analysis. In: Jacoby L, Siminoff LA, editors. Empirical methods for bioethics: a primer. 1. ed. Amsterdam: Elsevier JAI; 2008. (Advances in bioethics).

37. Baumer BS, Kaplan DT, Horton NJ. 5) Tidy Data and Iteration. In: Modern Data Science with R [Internet]. Boca Raton, UNITED STATES: CRC Press LLC; 2017. Available from: http://ebookcentral.proquest.com/lib/ucalgary-ebooks/detail.action?docID=5475748

38. Broman KW, Woo KH. Data organization in spreadsheets. Am Stat. 2018 Jan 2;72(1):2–10.

39. Buneman P, Khanna S, Tan W-C. Data Provenance: Some Basic Issues. In: Kapoor S, Prasad S, editors. FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science. Berlin, Heidelberg: Springer; 2000. p. 87–93. (Lecture Notes in Computer Science).

40. Dudley JT, Butte AJ. A Quick Guide for Developing Effective Bioinformatics Programming Skills. PLOS Comput Biol. 2009 Dec 24;5(12):e1000589.

41. Ellis SE, Leek JT. How to share data for collaboration. Am Stat. 2018 Jan 2;72(1):53–7.

42. Endel F, Piringer H. Data Wrangling: Making data useful again. IFAC-Pap. 2015 Jan 1;48(1):111–2.

43. Gentzkow M, Shapiro JM. Code and Data for the Social Sciences: A Practitioner's Guide [Internet]. University of Chicago mimeo; 2014. Available from: https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf

44. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten Simple Rules for the Care and Feeding of Scientific Data. PLOS Comput Biol. 2014 Apr 24;10(4):e1003542.

45. Green T. We Need Publishing Standards for Datasets and Data Tables. OECD Publ. 2009;

46. Hampton SE, Jones MB, Wasser LA, Schildhauer MP, Supp SR, Brun J, et al. Skills and Knowledge for Data-Intensive Environmental Research. BioScience. 2017 Jun 1;67(6):546–57.

47. Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten Simple Rules for Digital Data Storage. PLOS Comput Biol. 2016 Oct 20;12(10):e1005097.

48. Hejblum BP, Kunzmann K, Lavagnini E, Hutchinson A, Robertson DS, Jones SC, et al. Realistic and Robust Reproducible Research for Biostatistics. Preprints.org [Internet]. 2020 Jun 2; Available from: https://www.preprints.org/manuscript/202006.0002/v1

49. Kandel S, Heer J, Plaisant C, Kennedy J, van Ham F, Riche NH, et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data. Inf Vis. 2011 Oct 1;10(4):271–88.

50. Lee BD. Ten simple rules for documenting scientific software. PLOS Comput Biol. 2018 Dec 20;14(12):e1006561.

51. Marwick B. Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. J Archaeol Method Theory. 2017 Jun;24(2):424–50.

52. McCallum QE. Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work. O'Reilly Media, Inc.; 2012. 265 p.

53. McNamara A. Key Attributes of a Modern Statistical Computing Tool. Am Stat. 2019 Oct 2;73(4):375–84.

54. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLOS Comput Biol. 2015 Oct 22;11(10):e1004525.

55. Noble WS. A Quick Guide to Organizing Computational Biology Projects. PLOS Comput Biol. 2009 Jul 31;5(7):e1000424.

56. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. Science. 2015 Jun 26;348(6242):1422–5.

57. Rosenberg David E., Filion Yves, Teasley Rebecca, Sandoval-Solis Samuel, Hecht Jory S., van Zyl Jakobus E., et al. The Next Frontier: Making Research More Reproducible. J Water Resour Plan Manag. 2020 Jun 1;146(6):01820002.

58. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. PLOS Comput Biol. 2013 Oct 24;9(10):e1003285.

59. Shade A, Teal TK. Computing Workflows for Biologists: A Roadmap. PLOS Biol. 2015 Nov 24;13(11):e1002303.

60. Stodden V, Miguez S. Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research [Internet]. Rochester, NY: Social Science Research Network; 2013 Sep. Report No.: ID 2322276. Available from: https://papers.ssrn.com/abstract=2322276

61. Tuyl SV, Whitmire AL. Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. PLOS ONE. 2016 Feb 17;11(2):e0147942.

62. White EP, Baldridge E, Brym ZT, Locey KJ, McGlinn DJ, Supp SR. Nine simple ways to make it easier to (re)use your data. Ideas Ecol Evol [Internet]. 2013 Aug 30;6(2). Available from: https://ojs.library.queensu.ca/index.php/IEE/article/view/4608

63. Wickham H. Tidy Data. J Stat Softw. 2014 Sep 12;59(10):1–23.

64. Wilson G, Aruliah DA, Brown CT, Hong NPC, Davis M, Guy RT, et al. Best Practices for Scientific Computing. PLOS Biol. 2014 Jan 7;12(1):e1001745.

65. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. PLOS Comput Biol. 2017 Jun 22;13(6):e1005510.

66. Yenni GM, Christensen EM, Bledsoe EK, Supp SR, Diaz RM, White EP, et al. Developing a modern data workflow for regularly updated data. PLOS Biol. 2019 Jan 29;17(1):e3000125.

67. Zhu Y, Hernandez LM, Mueller P, Dong Y, Forman MR. Data Acquisition and Preprocessing in Studies on Humans: What is Not Taught in Statistics Classes? Am Stat. 2013 Nov 1;67(4):235–41.

68. Lowndes JSS, Best BD, Scarborough C, Afflerbach JC, Frazier MR, O'Hara CC, et al. Our path to better science in less time using open data science tools. Nat Ecol Evol. 2017 Jun;1(6):0160.

69. Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, et al. Data Carpentry: Workshops to Increase Data Literacy for Researchers. Int J Digit Curation. 2015 Feb 18;10(1):135–43.

70. Hollman C, Paulden M, Pechlivanoglou P, McCabe C. A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. PharmacoEconomics. 2017 Aug;35(8):817–30.

71. 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 May 1;22(5):575–9.

72. 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 Jan 6;37(7):735–46.

Chapter 3

Impacts from Enhancing the Reproducibility of Health Technology Assessments: A Case Study in Psoriatic Arthritis

Background

The process of updating a previously commissioned Health Technology Assessment is widely acknowledged as inefficient. However, experiences from other scientific fields suggest that this may be a sign of sub-optimal reproducibility. To enhance this, it is necessary to change how computing is used to organize and manage HTAs. The objective of this study was to explore the procedural and methodological impacts from these insights in the HTA context.

Methods

In this case study, computing strategies known to support reproducibility guided the redevelopment of a previously commissioned HTA for Psoriatic Arthritis. A collection of modularized functions were used to create an automated workflow, organized into three distinct activities: i) the preparation of model parameters; ii) simulation for an adoption or research decision; and iii) the post-processing of simulation results. Once validated, the entire workflow was executed using the same analysis strategy as the original HTA. This involved the stochastic evaluation of 15 decision models across four distinct scenarios. To measure the impact of the redevelopment, the timing of each activity was recorded and summarized.

Results

Simulation ready model parameters were generated from raw data in four seconds. Unlike the original HTA, computational intensity did not restrict the use of stochastic simulation methods. For the adoption decision, it took 128 hours to generate results for the 15 models, using Monte Carlo simulation with 20,000 iterations. Lastly, the post-processing of simulation results was completed in minutes.

Conclusions

Meaningful procedural and methodological consequences were realized from the application of computing strategies which enhanced the reproducibiltiy of the PsA HTA. The editable, auditable, and authoritative nature of the automated workflow offered a mechanism for greater quality control. In addition, the re-usable nature of the workflow and its components meant

duplicate effort in iteration could be avoided. Lastly, the redevelopment revealed how computational efficiency can affect the characterization of uncertainty in decision making. While the investment in the underlying computing skills was significant, payoffs will include greater transparency, better decision making, and more efficient iteration.

3.1 Introduction

In the process of Health Technology Assessment (HTA), economic evaluation is used to inform the efficient allocation of scarce resources (1–3). To determine if an intervention or programme represents value for money, decision modelling is used to generate estimates of cost-effectiveness. This single framework is used to conduct an evidence synthesis of data from multiple sources (4–6). For example, following model conceptualization relevant evidence must be synthesized to inform the relevant input parameters. As much of this evidence is uncertain, generating unbiased estimates of expected costs and effects will require the probabilistic evaluation of the decision model (2,3). Additionally, consensus exists that ongoing iteration of an HTA is necessary to ensure decisions are made with current information (7–9). Prior to a given deadline, a new estimate of effectiveness may need to be incorporated into a systematic review. Iteration may also be necessary following the emergence of a new technology (7–9).

Satisfying these requirements will depend on the nature of the decision model and the time constraints of the appraisal process. Fidelity to the probabilistic requirement will be influenced by the computation time of a Monte Carlo simulation. This involves the repeated evaluation of a decision model using randomly sampled parameter values (10,11). The time-consuming nature is reflected by the need to generate a distribution of costs and effects for each alternative. However, computation times can also be affected by the complexity of the model structure and the number of scenario and sub-group combinations (3,10,12). Meanwhile, satisfaction of the iterative requirement will be influenced by the efficiency with which a new decision can be made with current information. However, updating an existing HTA with new information has been consistently cited as a duplicative activity (13–18). This may be indicative of development strategies which lead to sub-optimal levels of reproducibility. Narrowly, this

refers to the ability to obtain consistent computational results using the same set of files and processes. More broadly, reproducibility represents a strategic approach to directly validate, repeat, and potentially repurpose scientific work (19,20).

A systematic review was recently conducted to identify computing strategies applied in other scientific fields to support reproducible research. Findings revealed that enhancements to reproducibility require changes to the way in which computing is used to organize and manage HTAs (21). Recommended strategies encouraged the preservation of raw data in a machine-readable format to serve as a common reference point in an automated workflow. Reductions in the marginal cost of HTA iteration are expected to be realized from the revision or expansion of the scripts and data which comprise the workflow (21). However, the magnitude of the expected efficiency gains remains unspecified. Furthermore, the impact of the identified strategies on the efficiency of a Monte-Carlo simulation are unknown, especially in the context of complex model structures which consider multiple scenarios and sub-groups. Given the considerable time requirements involved with developing the necessary computing skills, implementation will be constrained until the return on investment can be clearly understood and justified.

This paper sought to understand the procedural and methodological impacts of using previously identified strategies for reproducible research in the HTA context. The most recent NICE Multiple Technology Appraisal (MTA) of Biologic Treatment for Psoriatic Arthritis (PsA) was selected as the case study because i) it had been subject to iterative revision; and ii) it considered a complex model structure with multiple sub-groups and scenarios which had previously limited the use of probabilistic methods owing to computational intensity (22–24).

3.2 Methods

A previously published HTA, originally commissioned by NICE, was redeveloped using the identified computing strategies to support reproducibility (21,24). The methods of the reference project were applied to be consistent with what was originally reported and are summarized in relation to the present study. For complete details on the methods for economic modelling and the estimation for each parameter, see Corbett et al. (24). Attention is given to the implementation of the identified computing strategies to support reproducibility.

3.2.1 Case Study: Biologic Treatment for Psoriatic Arthritis

The case study for this redevelopment was a Multiple Technology Appraisal of Biologic Treatment for Psoriatic Arthritis published by Corbett et al. in 2017 (24). The aim was to determine the clinical- and cost-effectiveness of two new biologic therapies, Secukinumab (SEC) and Certolizumab Pegol (CZP), relative to existing products among adults with PsA following ineffective treatment with a conventional Disease Modifying Anti-Rheumatic Drug (cDMARD). To determine clinical effectiveness, a systematic review of randomized controlled trials was first conducted. Eligible studies involved the treatment of adult PsA with one of the following interventions: SEC, CZP, Etanercept (ETN), Infliximab (INF), Adalimumab (ADA), Golimumab (GOL), Ustekinumab (UST), Apremilast (APR), or Placebo. Estimates of clinical effectiveness were synthesized using Bayesian Network Meta-Analysis (NMA) of four outcomes: i) Psoriatic Arthritis Response Criteria (PsARC); ii) Change in Health Assessment Questionnaire – Disability Index (HAQ-DI) conditional on PsARC Response; iii) Psoriasis Area Severity Index (PASI) Response; and iv) American College of Rheumatology Improvement Criteria (24).

A decision analytic model was used to compare the cost-effectiveness of alternative sequences of biologic treatments which were licensed for use in adult PsA. To reflect the

different stages of the treatment pathway for adult PsA, three sub-populations were specified: i) biologic naïve, 1 prior cDMARD; ii) biologic naïve, at least 2 prior cDMARDs; and iii) biologic experienced. Additionally, three sub-groups were considered within each sub-population to explore the impact of baseline Psoriasis severity (None, Mild, and Severe). Each model adopted an NHS and Personal Social Services perspective to evaluate costs and outcomes, expressed as Quality Adjusted Life Years (QALYs), over a time horizon of 40 years using a cycle length of 3 months (13 weeks). A price year of 2016 was assumed, and a 3.5% discount rate was applied to both costs and QALYs. Parameters for the decision model were obtained from published literature, manufacturers' reported data, and the results from three NMAs (PsARC, HAQ, and PASI) (24).

The decision analytic model used a Semi-Markov structure to track a cohort with a homogeneous baseline population across each cycle and health state. Markov states were defined as a sequence of biologic treatments, followed by best-supportive care (BSC) and death. Three distinct model structures were conceptualized to accommodate differences in the length of the treatment sequence by sub-population (Figure 3-1). In addition to the treatment, the model tracked arthritis (as measured by HAQ-DI) and psoriasis (as measured by PASI) severity across each cycle and Markov state. These estimates of symptom severity were used to calculate health utilities (EQ5D) as well as the health service costs of arthritis and psoriasis care. Treatment costs were estimated as the sum of drug acquisition, administration, and monitoring costs (24).

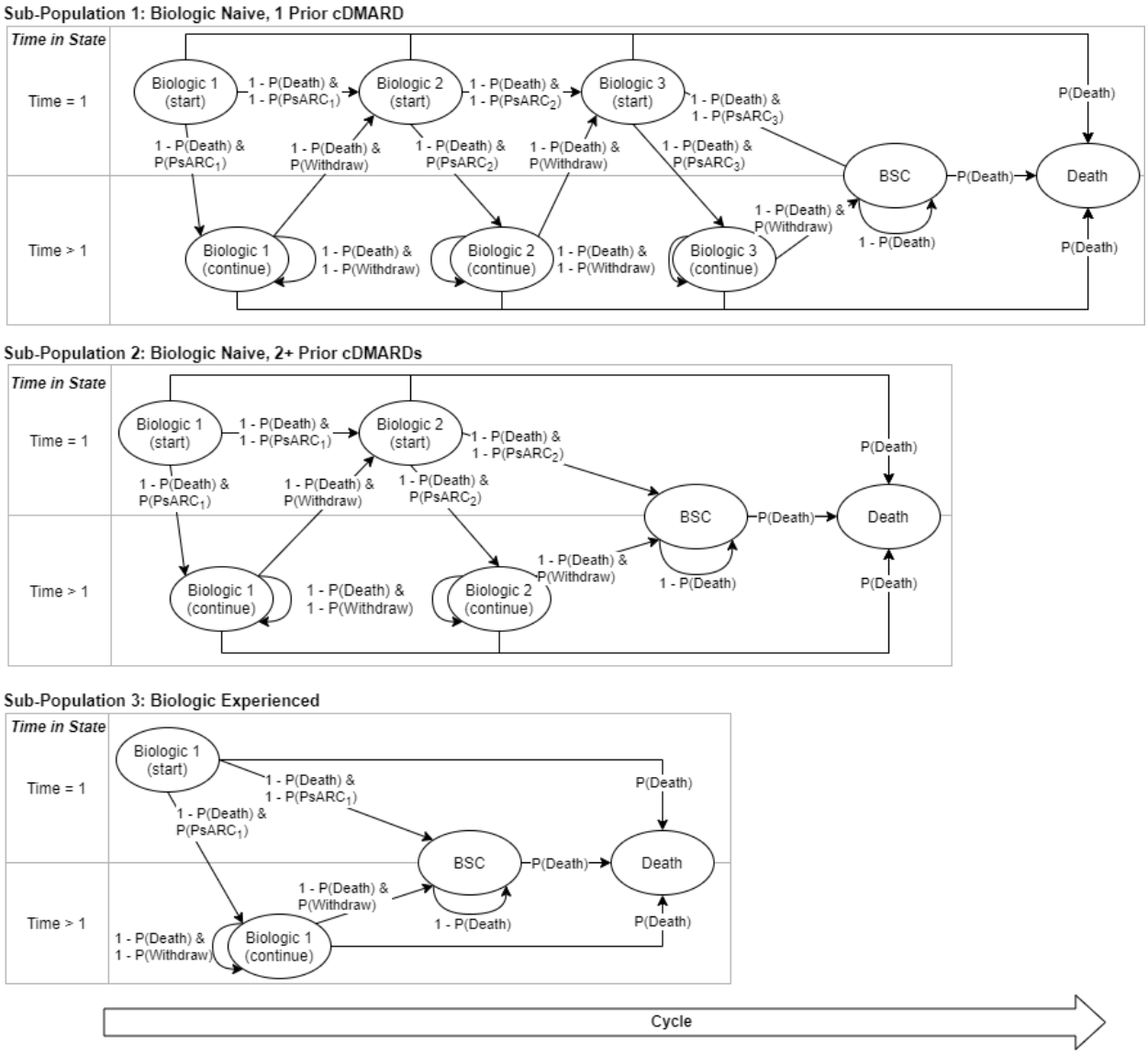


Figure 3-1. Semi-Markov structures required in the economic evaluation

3.2.2 Summary of Redevelopment

The Psoriatic Arthritis HTA was redeveloped as if it was commissioned *de novo*. The intent was to create a machine-executable record which captured the provenance of an adoption or research decision generated as part of the HTA process. To accomplish this, an automated workflow was designed to capture the procedures for the economic evaluation as well as those used to estimate each parameter of the decision model. Each identified computing strategy was used to achieve a level of reproducibility that would allow for the reliable regeneration of computational results, including intermediate datasets. Details summarizing the implementation of each strategy are reported under the following sub-headings, which match their description in the preceding systematic review (21).

3.2.2.1 *Provenance: Process/Code*

The R language and environment for statistical computing was used to program each step of the workflow (25). Before any code was written, pencil-and-paper were used to define the dependent relationships between the distinct methodological processes. The required tasks and datasets were subsequently organized into one of the acquisition, processing, and analysis stages of the basic reproducible workflow (19,21). An iterative development process was then used to create a collection of functions representing each distinct task of the workflow. The re-usable nature of each function was critical to this process for two reasons. First, separating function definitions from applications enabled safe prototyping of new tasks as well as instant propagation of changes to the source code. Second, the ability to call one function from the body of another saved the user from remembering the exact order of execution. Instead, higher level functions were created to capture and collect the output from each sequence of tasks. This led to the re-organization of the workflow into three distinct activities: i) the estimation of model

parameters directly from raw data; ii) simulation to inform the adoption or research decision; and iii) the estimation and preparation of simulation results.

The objective of the first activity was to generate the parameter values which could be fed directly into the simulation without additional modifications. To accomplish this, the relevant data acquisition, processing, or analysis steps were defined to prepare each parameter set (reflecting a declared currency and price year) from its raw data. Candidates for raw data were identified by tracing the provenance of each parameter in Corbett et al. to its publicly available origin (24). However, a candidate was only incorporated as raw data if the identified data processing or analysis steps could be reverse engineered and validated. When such attempts were unsuccessful, the output of the process of interest was used as raw data instead. The acquisition of raw data was primarily a manual process, apart from datasets containing life tables and values from the OECD used for inflation or currency conversions (26–28). Due to their serial nature, functionality was included to automatically update each dataset if the local version was more than one-year old.

Functions developed for the second activity were designed to return deterministic or probabilistic simulation output from the decision model. Modularization facilitated the satisfaction of the sub-group requirement by returning the model output for each sub-group from a single iteration of the decision model. This was permissible because the baseline PASI scores for each sub-group (None = 0; Mild = 7.3; Severe = 12.5) were not assumed to be uncertain parameters (24). Meanwhile, abstraction was used to iteratively revise the code for the decision model to accept a treatment sequence of any length (21). To achieve this, a multi-dimensional array was used to implement the Semi-Markov model in an approach similar to that described by Hawkins et al. (29). This data structure was well-suited to support this approach because it

offered the ability to independently subset the cycle, Markov state, and time-in-state dimensions. The resulting function leveraged this feature to define the Markov states dynamically to match the requirements defined in Figure 3-1.

For the third activity, functions were created to capture the standard post-processing of simulation output. This included the calculation of net-benefit statistics, expected values, incremental cost-effectiveness ratios, and Value-of-Information analyses. In addition, code was developed to summarize the results in tabular and graphical form (30).

3.2.2.2 *Provenance: Data*

Strategies for data management differed for raw and generated data. Raw data values were preserved in a machine-readable format (comma separated values) and treated as read-only by the automated workflow. This approach to data management eliminated the need to preserve datasets generated from downstream operations, as such values could always be re-generated. Therefore, the decision to preserve workflow generated datasets was left to user discretion.

All values were organized using the identified rules for creating “tidy” or “normalized” data. Each collection of related values was stored in a distinct table (and plain text file) where the rows and columns represented the observations and variables of the dataset (31). Fidelity to these rules offered a disciplined approach to data organization which made it easier to manage and access the data of interest. The most complex implementation was the data from the systematic review of the eligible treatments. Data from the 19 trials identified in the original systematic review were re-compiled from published sources following Cochrane guidance on data collection (32–72). Further details regarding the re-compiled data are presented in Appendix B. Complexities emerged from the need to organize data representing many different levels of observation across multiple tables. For example, information about risk of bias was represented

at the trial level while the trial results were collected for each outcome, event (week), sub-population, and arm within each trial. This exercise exposed the need for more sophisticated technologies to overcome the limitations of plain text formats and spreadsheets when managing complex datasets (73).

3.2.2.3 *Documentation*

Project level documentation was restricted to a minimal README which summarized the objective of the project and its ownership. Attempts were made to use meaningful names for all functions, files, and datasets. Duplicate names were avoided, as were terms such as “final” and “data”. Data documentation included a summary of the dataset, a description of each variable, and a reference stating its original source. Usage documentation was prepared for each function summarizing its purpose, argument definitions, methodology, and included example code. Comments were used in each script file to provide context to each code segment (21).

3.2.2.4 *Project Management*

Following identified guidance, all project files were organized under a common root directory (21). A consistent sub-directory structure was then used to separate scripts, data, documents, and results. Further sub-directories were created to separate the different types of files applicable to each parent directory. Changes across the project files were managed using the git version control system and mirrored to a private repository on GitHub (74–76).

3.2.3 Analysis Strategy

Following validation of the collection of functions and data created from the redevelopment, the workflow was executed in its entirety. To measure the impact of the reproducible strategies on the HTA process, the timing of each activity was recorded and summarized. A total of fifteen decision models were specified across each sub-population,

psoriasis sub-group, and eligible combination of effectiveness data. The treatment sequences and evidence synthesis combinations were unchanged from Corbett et al. and are presented in Table 3-1 (24).

Table 3-1. Available Comparators and Network Meta-Analyses for Each Decision Model

Treatment Sequences by Sub-Population			
Sub-Population 1: Biologic Naïve (1 Prior cDMARD)	**Sub-Population 2:** Biologic Naïve (2+ Prior cDMARDs)		**Sub-Population 3:** Biologic Experienced
SEC300-ETN-UST	SEC300-UST	GOL-UST	SEC300
SEC150-ETN-UST	SEC150-UST	ETN-UST	UST
CZP-ETN-UST	CZP-UST	INF-UST	BSC
BSC	ADA-UST	BSC	
Network Meta-Analysis Combinations			
Strategy	**PsARC Model[1]**	**HAQ Model[2]**	**PASI Model[3]**
Independent Analysis	A1	E1	F1
Meta-Regression[4]	D2	E2	G1

Notes:
1. See Table B-2
2. See Table B-5
3. See Table B-8
4. Restricted to biologic naïve sub-populations.

Each decision model was evaluated stochastically using Monte Carlo simulation. Normal distributions were used to characterize the uncertainty in most uncertain parameters, except for the HAQ change off treatment and excess mortality risk parameters which were assigned gamma and log-normal distributions (24). All simulations were executed in parallel on a high-performance compute cluster. For the adoption decision, simulations of 20,000 iterations were executed for the base case and three scenarios described in Corbett et al. (24). Optimal treatment sequences were identified at thresholds of £20,000 and £30,000, and incremental cost-effectiveness ratios were calculated from mean costs and QALYs. For the research decision, the expected value of perfect information (EVPI) was estimated from the adoption decision output. Additionally, expected value of partial parameter information (EVPPI) was estimated from a

nested Monte Carlo simulation of 1,000 inner and 1,000 outer-loop iterations. As this process was included for demonstration, EVPPI for the HAQ and PsARC NMA parameter sets was estimated for a single sub-population, using the independent analysis combination of NMA estimates. For cost-effectiveness results, see Corbett et al. (24).

3.3 Results

Execution of the entire workflow from a blank workspace is summarized in Figure 3-2. The process began by calling a single function to prepare the model parameters from raw data. The code within this function executed additional commands to carry out the relevant data processing or analysis steps needed to return the baseline trial characteristics, Gompertz coefficients, and inflated costing parameters. These values were then combined with the parameter sets designated as raw data to return the values reported in Table 3-2 in 4 seconds. Over 1,000 calls, identical results were returned in an average of 2.90 seconds (SD = 0.49).

Parameter sets which were treated as raw or generated data by the workflow are distinguished in the "Raw Data" column of Table 3-2. Unfortunately, attempts to reverse engineer the methodological process for every candidate parameter were unsuccessful. Time constraints prevented the re-use of previously reported code and data needed to estimate the treatment withdrawal parameter (23,24). Additionally, confidentially requirements prevented the replication of the original NMAs using the re-compiled trial data (Appendix B) (24). As a result, the output from these methodological processes were designated as raw data for the purposes of this study. Differences between the values reported in Table 3-2 and the original parameter values were attributed to data management, potentially undocumented procedures, and the serial nature of datasets needed to calculate inflation or currency statistics.

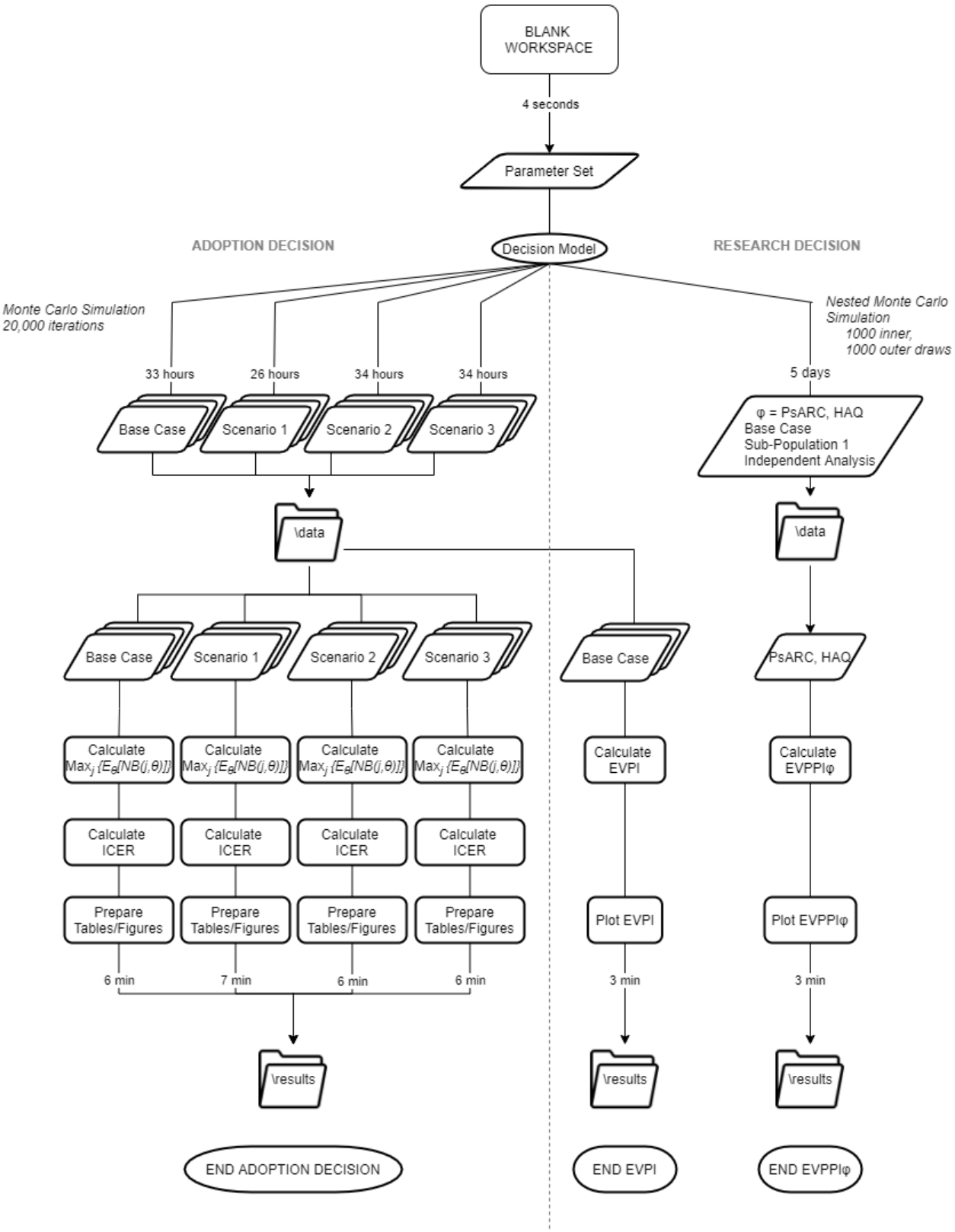


Figure 3-2. Time to event summary of the distinct stages in the reproducible workflow

Table 3-2. Workflow Generated Simulation Parameters

Parameters Estimated from Trial Data	**Mean**	**SE**	**Raw Data[1]**
Baseline Age	42.40	--[2]	✗
Baseline HAQ Score	1.22	--[2]	✗
Baseline Weight (kg)	86.96	--[2]	✗
Probability of PsARC Response	See Table B-2		✓
ΔHAQ Given PsARC Response Status	See Table B-5		✓
Probability of PASI50/75/90 Response	See Table B-8		✓
Life Tables	**Mean**	**SE**	**Raw Data[1]**
Age (Male, Female)	-10.6, -11.4	0.05, 0.0689	✗
Intercept (Male, Female)	0.096, 0.103	0.0007, 0.0009	
Excess Mortality Risk	**Hazard Ratio**	**95% CI**	**Raw Data[1]**
Overall	1.36	1.12 – 1.64	
Male	1.25	0.95 – 1.65	✓
Female	1.47	1.13 – 1.91	
Treatment Withdrawal	**Mean**	**SE**	**Raw Data[1]**
Log Annual Withdrawal Rate	-1.82	0.204	✓
Long Term HAQ Change	**Mean**	**SE**	**Raw Data[1]**
ΔHAQ on Treatment Per Cycle	0.00	0.002	✓
ΔHAQ after withdrawal (rebound)	0.00	0.5	✓
ΔHAQ off Treatment Per Cycle	0.0192	0.008	✗
EQ5D Utilities – Wyeth Coefficients	**Mean**	**SE**	**Raw Data[1]**
Intercept (β_0)	0.895	0.007	
HAQ (β_{HAQ})	-0.295	0.008	✓
PASI (β_{PASI})	-0.004	0.00	
HAQ x PASI ($\beta_{HAQ \times PASI}$)	0.00	0.00	
PsARC-PASI Correlation	0.4	0.1	✓
HAQ Costs – Kobelt Coefficients[2]	**Mean**	**SE**	**Raw Data[1]**
Intercept (β_0)	1862	655	✗
HAQ Mid-Point ($\beta_{Mid\text{-}Point}$)	563	364	
HAQ Costs – Poole Coefficients[2]	**Mean**	**SE**	**Raw Data[1]**
Intercept (β_0)	4.09	0.010	
HAQ (β_{HAQ})	2.37	0.006	✗
Age (β_{Age})	0.03	0.000	
HAQ x Age ($\beta_{HAQ \times Age}$)	-0.014	0.000	
Psoriasis Costs (Sub-Group)[2]	**Mean**	**SE**	**Raw Data[1]**
PASI75 Response/Non-Response (None)	£0.00, £0.00	£0.00, £0.00	
PASI75 Response/Non-Response (Mild)	£23.30, £423.98	£1.00, £9.00	✗
PASI75 Response/Non-Response (Severe)	£23.30, £423.98	£1.00, 9.00	
Total Treatment Costs (2016)	**Cycle 1**	**Cycle 2-160[3]**	**Raw Data[1]**
ETN	£2698	£2331	
INF	£7025	£3641	
ADA	£2836	£2292	
GOL	£3422	£2480	✗
UST	£4665	£2327	
SEC150/SEC300	£4029/£7685	£1988/£3969	
CZP	£3946	£2327	

Notes:
1. Raw data: values which were not modified for the economic evaluation.
2. Not applicable. Baseline values were not treated as uncertain parameters.
3. Mean estimated treatment costs per cycle from cycle 2 to 160.

Once the model parameters were loaded into memory, independent simulations were initiated to generate data to inform an adoption or research decision. Simulations were executed at the sub-population level and scripts were organized by scenario. A summary of the conditions and execution time for each simulation is presented in Table 3-3. Output from each adoption decision scenario were generated in 34 hours, with one scenario finishing in 24 hours. Meanwhile, the EVPPI simulation which considered one sub-population and one set of NMA values took 4.6 days to complete. Due to the time intensive nature of each simulation, the generated output was preserved within the project's data sub-directory.

The workflow concluded by reading the preserved simulation data into memory, generating estimates of cost-effectiveness, and depositing the relevant tables/figures into the results directory. For the adoption decision, each scenario was completed in 6 to 7 minutes. For the research decision, results were returned in 3 minutes. An example of the generated results is presented in Figure 3-3.

Table 3-3. Summary of the Execution Times for Each Decision Model Simulation

Adoption Decision					
Sub-Population	**Comparators**	**NMA Combinations**	**Iterations**	**Cores**	**Execution Time**
Base Case					
Naïve 1	4	2	500	4	2 hours[1]
Naïve 1	4	2	20,000	10	10.26 hours
Naïve 2	8	2	20,000	10	19.55 hours
Experienced	3	1	20,000	10	3.37 hours
Scenario 1					
Naïve 1	4	2	20,000	10	8.23 hours
Naïve 2	8	2	20,000	10	15.41 hours
Experienced	3	1	20,000	10	2.86 hours
Scenario 2					
Naïve 1	4	2	20,000	10	10.91 hours
Naïve 2	8	2	20,000	10	19.49 hours
Experienced	3	1	20,000	10	3.56 hours
Scenario 3					
Naïve 1	4	2	20,000	10	10.48 hours
Naïve 2	8	2	20,000	10	20.30 hours
Experienced	3	1	20,000	10	3.40 hours
Research Decision					
Sub-Population	**Comparators**	**NMA Combinations**	**Iterations**	**Cores**	**Execution Time**
Base Case					
Naïve 1	4	1	1,000,000	20	4.64 days

Notes:
1. Included for demonstration. Estimated that 20,000 iterations would take at least 80 hours.

A

Cost-Effectiveness Results - Biologic Naive (1 Prior cDMARD)[1]

Mild Psoriasis, Independent Analysis, Base Case

				λ = £20000/QALY[2]			λ = £30000/QALY[2]		
Initial Tx	**Cost**	**QALY**	**ICER**	**NMB**	**Probability**	**P(Error)**	**NMB**	**Probability**	**P(Error)**
bsc	£91,510.42	−0.21	—	−£95,674.08	0.08	—	−£97,755.91	0.00	—
CZP[3]	£170,298.35	5.06	D[4]	−£69,110.47	0.40	—	−£18,516.53	0.44	—
SEC150[3]	£165,083.19	5.09	13,890.50	−£63,314.32	0.53	0.47	−£12,429.88	0.55	0.45

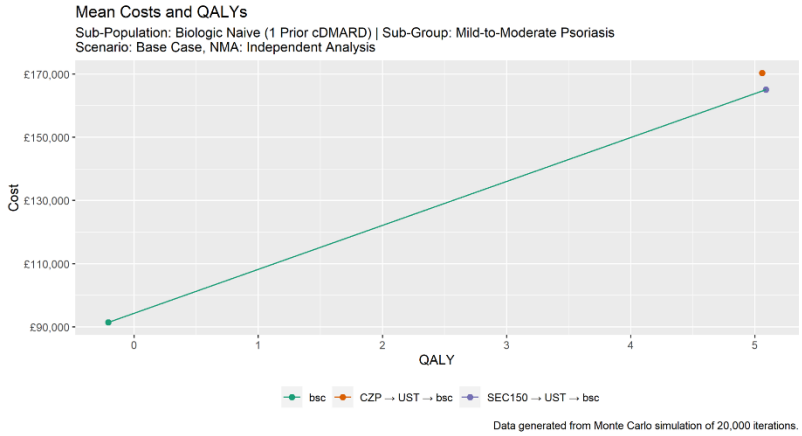
[1] Data generated from Monte Carlo simulation of 20,000 iterations

[2] λ = Threshold Ratio

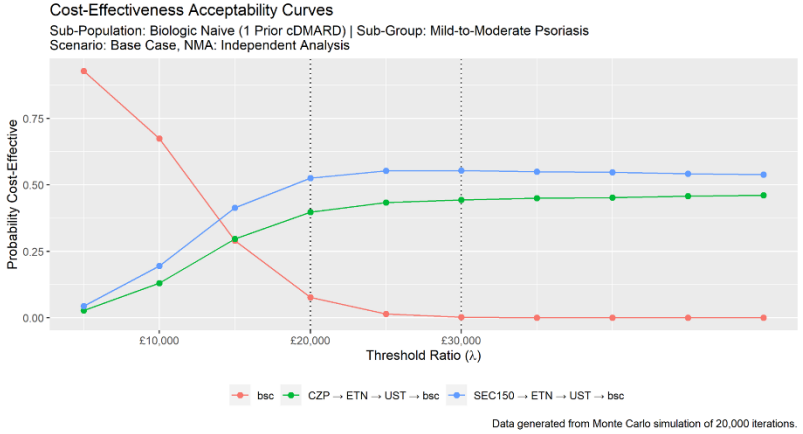
[3] Initial treatment followed by: ETN → UST → bsc

[4] D: Dominated

B



C



D

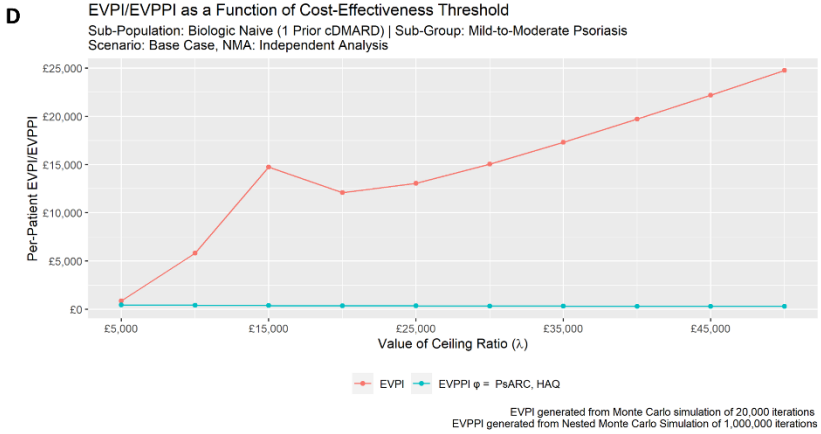


Figure 3-3. Base case results for a single sub-group, sub-population, and combination of effectiveness values.

3.4 Discussion

This paper summarized the execution of an automated workflow which was developed to enhance the reproducibility of a previously published HTA. Fidelity to the computing strategies for reproducibility may lead to greater transparency, better decision making, and population health gains. Broader realization of these outcomes will require a significant time commitment to learn the relevant computing skills along with new approaches to data and project management (21). In the present study, this investment was quickly recovered through the creation of robust code which can be used repeatedly at low cost. In other words, expanding the use of computing in HTA can have important procedural and methodological consequences.

3.4.1 Procedural Consequences

The primary difference between the reference project and the redevelopment was the approach used to track and manage the provenance of the results. *Ad hoc* processes reliant on independently managed computer files were replaced with an automated approach which mirrored the desired methodological sequence. One consequence from this process shift was superior quality control. Despite their potential to appear harmless, any programming or transcription error can undermine the objectives of the HTA process. For example, transcription errors in the parameter values can shift the simulated distributions of cost and effect, leading to the adoption of the wrong technology. Designing the project for reproducibility improved the feasibility of error detection as the behaviour and methodology of each function could be directly verified (77). This allowed each identified error to be traced to its origin and corrected. Lastly, the risk of transcription error was mitigated as human intervention was not required to pass output from one step as input to the next.

Another procedural consequence was a significant reduction in the marginal cost of iteration. To avoid duplication of effort, it was necessary to expand the scope of the workflow to include the generation of model parameters from raw data. In other words, efficiency gains were attributable to the implementation of a workflow design which minimized the accumulation of technical debt. This refers to the cost of additional work caused by choosing an easy solution to a problem rather than a better approach that would take longer (78). The exclusion of the code to execute each NMA is an example of technical debt which must be serviced in a future iteration. However, integrating the code for each NMA into the existing collection of functions will be much less expensive than redeveloping every parameter value from scratch. Once all the changes are incorporated one would be minutes away from initiating new simulations for an adoption or research decision (Figure 3-2). In the long run, integrating the generation of model parameters in the same automated workflow as the simulation is likely to yield significant savings for the research budget.

3.4.2 Methodological Consequences

Neither the complex model structure or the sub-population, sub-group, and NMA combinations constrained the use of Monte Carlo simulation in the present study. This deviation from the method used in Corbett et al. (24) was attributable to a relative improvement in computational efficiency – the speed with which a computer can execute a task given a particular piece of code (79). With any programming tool, there will be many ways to arrive at the same solution and some will be faster than others. In the present study, adherence to the identified computing strategies made the procedures easier to read and execute. For example, strategies for data organization encouraged the use of data structures in R which were fit-for-purpose and memory efficient. In addition to the use of multi-dimensional arrays to track the treatment and

symptom severity (29), parameter values were stored in a list to accommodate the heterogeneous dimensions of each dataset (Table 3-2). Furthermore, faster computations were achieved from eliminating duplication in code. In the process of developing each function, repeated operations in for-loops were replaced with vectorized commands which executed the same operation on a collection of values all at once. The relationship between memory usage and speed is illustrated in Table 3-3. Execution time was directly influenced by the number of Markov states (Figure 3-1), NMA strategies considered, and the number of comparators.

Further reductions in the execution time of each simulation were achieved using parallel computation. This refers to the simultaneous execution of a task across multiple Central Processing Units (CPU; the brain of a computer responsible for numeric calculations). To use this method, one requires access to a personal computer with multiple CPUs (cores) integrated on a single chip (i.e., most modern computers) or a network of computers maintained by a third party (i.e., high performance cluster). Monte Carlo simulation is well-suited for parallelization because each iteration represents an independent realization of the decision model (80). Compared to executing a simulation on a single CPU (the default), the performance gains can be substantial. However, constraints will emerge from the available compute power and memory. This is reflected by the simulation in Table 3-3 which was parallelized over four cores on a personal computer. It is estimated that it would have taken at least 80 hours to complete 20,000 iterations – which is inefficient relative to the 128 hours it took to execute all 15 simulations when parallelized over 10 cores on a high-performance cluster.

Lastly, computational efficiency will impact the characterization of uncertainty by making it feasible to consider a large number of comparators. In Corbett et al., the explicit acknowledgement of computational challenges suggest it may have played a small role in the

number of comparators considered (24,81). For example, the treatment sequences for Sub-Population 1 (Table 3-1) represent a small subset of the total possible sequence combinations. The performance improvements realized from the redevelopment would have allowed a larger subset of alternatives to be considered. Methods for optimizing computational efficiency and its impact on the characterization of uncertainty remain an important area for future research.

3.5 Conclusions

Economic evaluation plays a critical role in the HTA process by informing decisions regarding the efficient allocation of scarce resources. Fulfillment of this mandate will require the characterization of all relevant sources of uncertainty and the ability to incorporate new evidence efficiently. Findings from this study revealed that enhancing the reproducibility of an HTA can aid in satisfying both requirements. In the long run, development of the computing skills to create a reproducible workflow will lead to a drastic reduction in the marginal cost of iteration, greater transparency, and better decisions.

3.6 References

1. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Serv Res. 2013 Sep 30;13(1):371.

2. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada [Internet]. Ottawa, ON; 2017 Mar. Report No.: 4th ed. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition

3. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013 [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013 [cited 2018 May 30]. 94 p. (NICE Process and Methods Guides). Available from: http://www.ncbi.nlm.nih.gov/books/NBK395867/

4. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 9: Economic Evaluation Using Decision Analytic Modelling. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. p. 311–51.

5. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006;15(7):677–87.

6. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006. 237 p. (Briggs A, Gray A, editors. Oxford handbooks in health economic evaluation).

7. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

8. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessent: the role of iterative decision analytic modelling. 2000 May [cited 2021 Jan 11]; Available from: https://core.ac.uk/reader/6406440

9. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

10. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? Value Health. 2006;9(4):244–52.

11. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation: A Practical Approach. Med Decis Making. 1985 Jun;5(2):157–77.

12. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14(4):339–47.

13. Cohen JT, Neumann PJ, Wong JB. A Call for Open-Source Cost-Effectiveness Analysis. Ann Intern Med. 2017;167(6):432–3.

14. Cohen JT, Wong JB. Can Economic Model Transparency Improve Provider Interpretation of Cost-Effectiveness Analysis? A Response: Med Care. 2017 Nov;55(11):912–4.

15. 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 Nov 1;37(11):1313–20.

16. Linthicum MT, Bright J, Shafrin J. Open-Source Tools for Value Assessment: A Promising Approach. J Clin Pathw. 2018 Jun 12;4(5):37–9.

17. Wu EQ, Zhou Z-Y, Xie J, Metallo C, Thokala P. Transparency in Health Economic Modeling: Options, Issues and Potential Solutions. PharmacoEconomics. 2019 Nov;37(11):1349–54.

18. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, et al. Transparency in Decision Modelling: What, Why, Who and How? PharmacoEconomics. 2019 Nov 1;37(11):1355–69.

19. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018 [cited 2018 Jul 8]. Available from: http://www.practicereproducibleresearch.org/

20. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019 [cited 2020 Jan 16]. Available from: https://www.nap.edu/catalog/25303

21. Wagner DJ. Computing Strategies to Enhance the Reproducibility of Health Technology Assessments. University of Calgary; 2021.

22. Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess [Internet]. 2006 Sep [cited 2018 Apr 25];10(31). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta10310/

23. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess [Internet]. 2011 Feb [cited 2018 Apr 25];15(10). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta15100/

24. Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to

disease-modifying antirheumatic drugs: a systematic review and economic evaluation. Health Technol Assess. 2017 Oct;21(56):1–326.

25. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

26. United Kingdom Office for National Statistics. National Life Tables: England and Wales [Internet]. Office of National Statistics; 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesenglandandwalesreferencetables/current

27. OECD. Consumer prices - complete database. 2020 Oct 7;

28. OECD. Purchasing power parities (PPP) and exchange rates. 2020 Oct 7;

29. Hawkins N, Sculpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. Med Decis Making. 2016 Sep 5;25(5):511–9.

30. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. PharmacoEconomics. 2008 Sep 1;26(9):781–98.

31. Wickham H. Tidy Data. J Stat Softw. 2014 Sep 12;59(10):1–23.

32. Li T, Higgins JP, Deeks JJ, editors. Chapter 5: Collecting Data. In: Cochrane Handbook of Systematic Review of Interventions [Internet]. 6th ed. Cochrane; 2019. Available from: www.training.cochrane.org/handbook

33. Antoni CE, Kavanaugh A, Heijde D van der, Beutler A, Keenan G, Zhou B, et al. Two-year efficacy and safety of infliximab treatment in patients with active psoriatic arthritis: findings of the Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT). J Rheumatol. 2008 May 1;35(5):869–76.

34. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, et al. Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: Results from the infliximab multinational psoriatic arthritis controlled trial (IMPACT). Arthritis Rheum. 2005 Apr 1;52(4):1227–36.

35. Antoni CE, Krueger GG, de Vlam K, Birbara CA, Beutler A, Guzzo C, et al. Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. Ann Rheum Dis. 2005 Aug 1;64(8):1150–7.

36. Choy EHS, Gladman DD, Sasso EH. Efficacy of adalimumab by disease duration in psoriatic arthritis: Subanalysis of ADEPT. In: Journal of the American Academy of Dermatology [Internet]. Washington, DC: American Academy of Dermatology; 2007 [cited 2019 Oct 15]. p. AB186. (Poster Abstracts; vol. 56). Available from: http://www.sciencedirect.com/science/article/pii/S0190962206037352

37. Corbett M, Sideris E, Palmer S, Harden M, Woolacott N, Bojke L. Apremilast for treating active psoriatic arthritis. York, UK: CRD and CHE Technology Assessment Group, University of York; 2015 Mar p. 150. (Single Technology Appraisal). Report No.: 13/169/01.

38. Craig D, O'Connor J, Rodgers M, Rodriguez-Lopez R, Smith A, Woolacott N. Ustekinumab for treating active and progressive psoriatic arthritis [Internet]. York, UK: CRD and CHE Technology Assessment Group, University of York; 2013 Oct p. 187. (Single Technology Appraisal). Report No.: HTA 12/58. Available from: https://www.journalslibrary.nihr.ac.uk/programmes/hta/125801/#/

39. Cutolo M, Myerson GE, Fleischmann RM, Lioté F, Díaz-González F, Bosch FV den, et al. A Phase III, Randomized, Controlled Trial of Apremilast in Patients with Psoriatic Arthritis: Results of the PALACE 2 Trial. J Rheumatol. 2016 Sep 1;43(9):1724–34.

40. Edwards CJ, Blanco FJ, Crowley J, Birbara CA, Jaworski J, Aelion J, et al. Apremilast, an oral phosphodiesterase 4 inhibitor, in patients with psoriatic arthritis and current skin involvement: a phase III, randomised, controlled trial (PALACE 3). Ann Rheum Dis. 2016 Jun 1;75(6):1065–73.

41. Genovese MC, Mease PJ, Thomson GTD, Kivitz AJ, Perdok RJ, Weinberg MA, et al. Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying antirheumatic drug therapy. J Rheumatol. 2007 May 1;34(5):1040–50.

42. Gladman DD, Mease PJ, Cifaldi MA, Perdok RJ, Sasso EH, Medich J. Adalimumab improves joint-related and skin-related functional impairment in patients with psoriatic arthritis: patient-reported outcomes of the Adalimumab Effectiveness in Psoriatic Arthritis Trial. Ann Rheum Dis. 2007 Feb 1;66(2):163–8.

43. Gladman DD, Mease PJ, Ritchlin CT, Choy EHS, Sharp JT, Ory PA, et al. Adalimumab for long-term treatment of psoriatic arthritis: Forty-eight week data from the adalimumab effectiveness in psoriatic arthritis trial. Arthritis Rheum. 2007 Jan 30;56(2):476–88.

44. Gladman D, Mease P, Kavanaugh A, Weinberg M. Adalimumab is efficacious in treating skin disease in psoriatic arthritis: subanalysis of moderate versus severe psoriasis in the ADEPT trial. In: American Academy of Determatology 64th Annual Meeting. San Francisco, California: American Academy of Dermatology; 2006. p. AB10–1.

45. Gladman D, Mease P, Ritchlin C, Okun M. PASI-100 is associated with better dermatology-specific patient-reported outcomes compared with PASI-75–99 in adalimumab-treated patients with psoriatic arthritis: Subanalysis of ADEPT. In: Journal of the American Academy of Dermatology [Internet]. Washington, DC: American Academy of Dermatology; 2007 [cited 2019 Oct 15]. p. AB185. (Poster Abstracts; vol. 56). Available from: http://www.sciencedirect.com/science/article/pii/S0190962206037339

46. Gottlieb AB, Langley RG, Philipp S, Sigurgeirsson B, Blauvelt A, Martin R, et al. Secukinumab Improves Physical Function in Subjects With Plaque Psoriasis and Psoriatic

Arthritis: Results from Two Randomized, Phase 3 Trials. J Drugs Dermatol JDD. 2015 Aug;14(8):821–33.

47. Kavanaugh A, Antoni CE, Gladman DD, Wassenberg S, Zhou B, Beutler A, et al. The Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT): results of radiographic analyses after 1 year. Ann Rheum Dis. 2006 Aug 1;65(8):1038–43.

48. Kavanaugh A, Antoni CE, Krueger GG, Yan S, Bala M, Dooley LT, et al. Infliximab improves health related quality of life and physical function in patients with psoriatic arthritis. Ann Rheum Dis. 2006 Apr 1;65(4):471–7.

49. Kavanaugh A, Antoni C, Mease P, Gladman D, Yan S, Bala M, et al. Effect of infliximab therapy on employment, time lost from work, and productivity in patients with psoriatic arthritis. J Rheumatol. 2006 Nov 1;33(11):2254–9.

50. Kavanaugh A, Krueger GG, Beutler A, Guzzo C, Zhou B, Dooley LT, et al. Infliximab maintains a high degree of clinical response in patients with active psoriatic arthritis through 1 year of treatment: results from the IMPACT 2 trial. Ann Rheum Dis. 2007 Apr 1;66(4):498–505.

51. Kavanaugh A, McInnes I, Mease P, Krueger GG, Gladman D, Gomez-Reino J, et al. Golimumab, a new human tumor necrosis factor α antibody, administered every four weeks as a subcutaneous injection in psoriatic arthritis: Twenty-four–week efficacy and safety results of a randomized, placebo-controlled study. Arthritis Rheum. 2009 Apr 1;60(4):976–86.

52. Kavanaugh A, Mease PJ, Gomez-Reino JJ, Adebajo AO, Wollenhaupt J, Gladman DD, et al. Treatment of psoriatic arthritis in a phase 3 randomised, placebo-controlled trial with apremilast, an oral phosphodiesterase 4 inhibitor. Ann Rheum Dis. 2014 Jun 1;73(6):1020–6.

53. Langley RG, Elewski BE, Lebwohl M, Reich K, Griffiths CEM, Papp K, et al. Secukinumab in Plaque Psoriasis — Results of Two Phase 3 Trials. N Engl J Med. 2014 Jul 24;371(4):326–38.

54. Lebwohl M, Gottlieb AB, Goffe BS, Jahreis A. Etanercept in psoriatic arthritis: Sustained improvement in skin and joint disease and inhibition of radiographic progression at 2 years. In: Journal of the American Academy of Dermatology [Internet]. New Orleans, LA; 2005 [cited 2019 Oct 15]. p. AB8. (American Academy of Dermatology Intersociety Liasion Concurrent Session Abstracts; vol. 52). Available from: http://www.sciencedirect.com/science/article/pii/S0190962205003439

55. McInnes IB, Kavanaugh A, Gottlieb AB, Puig L, Rahman P, Ritchlin C, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. The Lancet. 2013 Aug 31;382(9894):780–9.

56. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic

arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3 trial. The Lancet. 2015 Sep 19;386(9999):1137–46.

57. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Bosch FV den, Wellborne F, et al. Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). Ann Rheum Dis. 2009 May 1;68(5):702–9.

58. Mease P, Gladman D, Ritchlin C, Mark W. Clinical efficacy, safety and inhibition of joint destruction of adalimumab in the treatment of moderate to severe psoriatic arthritis: 48-week results of the ADEPT trial. In: Journal of the American Academy of Dermatology [Internet]. San Francisco, California: American Academy of Dermatology; 2006 [cited 2019 Oct 15]. p. AB214–8. (American Academy of Dermatology: Poster Abstracts; vol. 54). Available from: http://www.sciencedirect.com/science/article/pii/S0190962205042519

59. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. The Lancet. 2000 Jul 29;356(9227):385–90.

60. Mease PJ, Fleischmann RM, Deodhar AA, Wollenhaupt J, Khraishi M, Kielar D, et al. Effect of certolizumab pegol on signs and symptoms in patients with psoriatic arthritis: 24-week results of a Phase 3 double-blind randomised placebo-controlled study (RAPID-PsA). Ann Rheum Dis. 2014 Jan 1;73(1):48–55.

61. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EHS, et al. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: Results of a double-blind, randomized, placebo-controlled trial. Arthritis Rheum. 2005 Oct 1;52(10):3279–89.

62. Mease PJ, Heijde D van der, Ritchlin CT, Okada M, Cuchacovich RS, Shuler CL, et al. Ixekizumab, an interleukin-17A specific monoclonal antibody, for the treatment of biologic-naive patients with active psoriatic arthritis: results from the 24-week randomised, double-blind, placebo-controlled and active (adalimumab)-controlled period of the phase III trial SPIRIT-P1. Ann Rheum Dis. 2017 Jan 1;76(1):79–87.

63. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, et al. Etanercept treatment of psoriatic arthritis: Safety, efficacy, and effect on disease progression. Arthritis Rheum. 2004 Jul 1;50(7):2264–72.

64. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, et al. Continued inhibition of radiographic progression in patients with psoriatic arthritis following 2 years of treatment with etanercept. J Rheumatol. 2006 Apr 1;33(4):712–21.

65. Mease PJ, van der Heijde D, Ritchlin CT, Cuchacovich R, Shuler CL, Lee CH, et al. A randomized, double-blind, active-and placebo-controlled phase 3 study of efficacy and safety of ixekizumab, adalimumab, and placebo therapy in patients naive to biologic disease modifying anti-rheumatic drugs with active psoriatic arthritis. In: Arthritis & Rheumatology [Internet]. San Francisco, California; 2015. p. 1. Available from: https://acrabstracts.org/abstract/a-randomized-double-blind-active-and-placebo-controlled-

phase-3-study-of-efficacy-and-safety-of-ixekizumab-adalimumab-and-placebo-therapy-in-patients-naive-to-biologic-disease-modifying-anti/

66. Mease P, Sharp J, Ory P, Gladman D, Ritchlin C, Choy E, et al. Adalimumab treatment effects on radiographic progression of joint disease in patients with psoriatic arthritis: results from ADEPT. In: Journal of the European Academy of Dermatology and Venereology [Internet]. London, UK; 2005 [cited 2019 Oct 15]. p. 1–411. Available from: http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-3083.2005.01311.x

67. National Institute for Health and Care Excellence. Psoriatic Arthritis (Active) – Apremilast (Post DMARDs) [Internet]. National Institute for Health and Care Excellence; 2015 Apr p. 229. (Single Technology Appraisal). Report No.: ID682. Available from: https://www.nice.org.uk/guidance/TA372/documents/psoriatic-arthritis-active-apremilast-post-dmards-id682-committee-papers-2

68. O'Connor J, Rice S, Smith A, Rodgers M, Lopez RR, Craig D, et al. The Clinical and Cost Effectiveness of Ustekinumab for the Treatment of Psoriatic Arthritis: A Critique of the Evidence. PharmacoEconomics. 2016 Jan 27;34(4):337–48.

69. Ritchlin C, Rahman P, Kavanaugh A, McInnes IB, Puig L, Li S, et al. Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional non-biological and biological anti-tumour necrosis factor therapy: 6-month and 1-year results of the phase 3, multicentre, double-blind, placebo-controlled, randomised PSUMMIT 2 trial. Ann Rheum Dis. 2014 Jun 1;73(6):990–9.

70. Thaçi D, Blauvelt A, Reich K, Tsai T-F, Vanaclocha F, Kingo K, et al. Secukinumab is superior to ustekinumab in clearing skin of subjects with moderate to severe plaque psoriasis: CLEAR, a randomized controlled trial. J Am Acad Dermatol. 2015 Sep 1;73(3):400–9.

71. van der Heijde D, Kavanaugh A, Gladman DD, Antoni C, Krueger GG, Guzzo C, et al. Infliximab inhibits progression of radiographic damage in patients with active psoriatic arthritis through one year of treatment: Results from the induction and maintenance psoriatic arthritis clinical trial 2. Arthritis Rheum. 2007 Jul 30;56(8):2698–707.

72. Yang H, Epstein D, Bojke L, Craig D, Light K, Bruce I, et al. Golimumab for the treatment of psoriatic arthritis. Health Technol Assess. 2011 Jun 15;15(Suppl. 1):87–96.

73. Murrell P. Introduction to Data Technologies [Internet]. CRC Press; 2009. 445 p. Available from: https://www.stat.auckland.ac.nz/~paul/ItDT/

74. Blischak JD, Davenport ER, Wilson G. A Quick Introduction to Version Control with Git and GitHub. PLOS Comput Biol. 2016 Jan 19;12(1):e1004668.

75. Bryan J. Excuse Me, Do You Have a Moment to Talk About Version Control? Am Stat. 2018 Jan 2;72(1):20–7.

76. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F da V, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLOS Comput Biol. 2016 Jul 14;12(7):e1004947.

77. 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 Sep 1;32(5):733–43.

78. Kruchten P, Nord RL, Ozkaya I. Technical Debt: From Metaphor to Theory and Practice. IEEE Softw. 2012 Nov;29(6):18–21.

79. Visser MD, McMahon SM, Merow C, Dixon PM, Record S, Jongejans E. Speeding Up Ecological and Evolutionary Computations in R; Essentials of High Performance Computing for Biologists. PLOS Comput Biol. 2015 Mar 26;11(3):e1004140.

80. McCallum E, Weston S. Parallel R. O'Reilly Media, Inc.; 2011. 123 p.

81. Mauskopf J, Samuel M, McBride D, Mallya UG, Feldman SR. Treatment Sequencing After Failure of the First Biologic in Cost-Effectiveness Models of Psoriasis: A Systematic Review of Published Models and Clinical Practice Guidelines. Pharmacoeconomics. 2014;32(4):395–409.

Chapter 4

Overcoming Barriers to the Characterization of Uncertainty Through Investments in Reproducibility

The characterization of all relevant sources of uncertainty in a decision model necessitates the use of probabilistic methods. Despite the widespread use of computing to facilitate this task, probabilistic analysis is often abandoned when simulations take too long to complete. While this computing challenge is often attributed to the nature of a model's structure, it may instead be an outcome of the model's programming. This paper sought to describe how these computing challenges may be addressed as a by-product of good computing practices focused on improved reproducibility. To achieve this, the paper summarizes how an emphasis on reproducibility influenced the programming of a complex decision model in a redevelopment of an existing technology appraisal. Strategies which made it simple to reproduce the results also enabled the joint characterization of structural and parameter uncertainty. In addition, the model was programmed to make effective use of the available computational resources. Ultimately, the paper argues how, in the context of reproducible computing practices, the performance of a Monte Carlo simulation becomes more constrained by the capacity limits of the computer hardware rather than the model's structure.

4.1 Introduction

Health Technology Assessment (HTA) leverages well established methods of evidence synthesis to support policy making on many factors, including value for money (1,2). When an economic evaluation is commissioned, methods for decision modelling are used to combine evidence from a range of sources to generate information on relative cost-effectiveness (1,3). Given that such evidence is uncertain, methodological literature and guidance documents unambiguously endorse the probabilistic evaluation of decision models (4–6). In this context, the decision model can be used as a vehicle to inform two decisions: i) adoption of a technology; and ii) the commissioning of additional research. The general use of a Bayesian framework reflects a consensus that HTA should represent a continuous process of evidence gathering and decision making (7–9). However, there is a serious risk that computing challenges can interfere with efforts to fully characterize decision uncertainty.

As with other areas of science, computing has become essential for the economic evaluation of health technologies. Unfortunately, the widespread adoption of computing has yet to be accompanied by sufficient training in the effective use of this technology. One way this is reflected in practice is when the inability to stochastically evaluate a decision model is attributed to computational intensity, burden, or expense (5,6,10,11). While these terms have not been explicitly defined in the HTA literature, their usage implies that it is appropraite to abandon probabilistic methods when some threshold level of "intensity" is reached. This is inconsistent with the consensus that the characterization of uncertainty is necessary to explore its consequences and to return unbiased estimates of cost-effectiveness (6,10,11).

Regardless of how computational intensity is defined or measured, it will always reflect the way the decision model was programmed. The relative nature of this designation therefore

challenges the notion that decision models with complex structures will always be computationally expensive. This is evidenced by a case study which re-developed a NICE commissioned economic evaluation of biologic treatment for Psoriatic Arthritis (PsA) (12,13). The goal of this effort was to enhance the reproducibility of the original study using a series of computing strategies synthesized from the broader scientific literature (12,14). Both projects used the R programming language to compare sequences of biologic treatments with a Semi-Markov structure. Further complexity was reflected by the need to consider multiple sub-groups, sub-populations, and scenarios. Despite these similarities, efforts to characterize uncertainty in the case study were not burdened by any of the challenges documented in the original appraisal by Corbett et al. (12,13).

Critically, the scope of the case study was restricted to summarizing the development and execution of a reproducible workflow (12). This left little room to explain how the emphasis on reproducibility influenced the programming of the economic evaluation. While this led to an improvement in computational efficiency, the impact on different components of the computer hardware was only briefly described (12). As a result, a detailed overview of the techniques used to overcome common computing challenges in economic evaluation is required.

The objective of this paper is to describe how an emphasis on reproducibility can enable more consistent and thorough explorations of uncertainty in economic evaluations of health technologies. Following a brief summary of the case study, the paper will begin with an overview of the reproducible design of the decision model. Afterwards, it will illustrate how this design can be leveraged to overcome two limitations to the characterization of uncertainty in the case study. The paper will then conclude with a summary of the techniques used to overcome computing challenges in the characterization of uncertainty.

4.2 Summary: Reproducible Case Study

A previously published HTA was redeveloped with the intent to achieve a level of reproducibility that would allow for the reliable re-generation of results (12,13). To guide this effort, the case study implemented a series of computing strategies known to support reproducible research efforts in other scientific fields (12,14). This involved automating each step in the methodological sequence using the R programming language (15). An iterative development strategy was used to modularize the code for each distinct task into single purpose programs called functions (12). These re-useable commands could be executed independently or within the body of another, higher-level function. This allowed the project process to be organized into three distinct activities: i) the estimation of simulation ready model parameters; ii) evaluation of the decision model(s) for an adoption or research decision; and iii) preparation of adoption/research decision results (12). The focus of this paper is on the design of the code used for the second activity.

4.2.1 Economic Evaluation

The economic evaluation in Corbett et al., sought to determine the cost-effectiveness of two biologic therapies, Secukinumab (SEC) and Certolizumab Pegol (CZP) in treating adult PsA (12,13). Alternative treatments were restricted to those licensed to treat PsA at the time of the evaluation: Etanercept (ETN), Infliximab (INF), Adalimumab (ADA), Golimumab (GOL), Ustekinumab (UST), Apremilast (APR), and Placebo (12,13). A key feature in this economic model was the ability for patients to switch between treatments. As a result, estimates of cost-effectiveness were generated for alternative sequences of biologics whose length differed by sub-population. Patients were eligible to receive a maximum of 3 biologics (Biologic Naïve, 1 Prior cDMARD), 2 biologics (Biologic Naïve, at least 2 prior cDMARDs), or 1 biologic (Biologic

Experienced) (12,13). To explore the impact of baseline psoriasis severity, three sub-groups were considered within each sub-population (None, Mild, Severe). In addition, two distinct combinations of model inputs from three distinct evidence syntheses were considered (12,13).

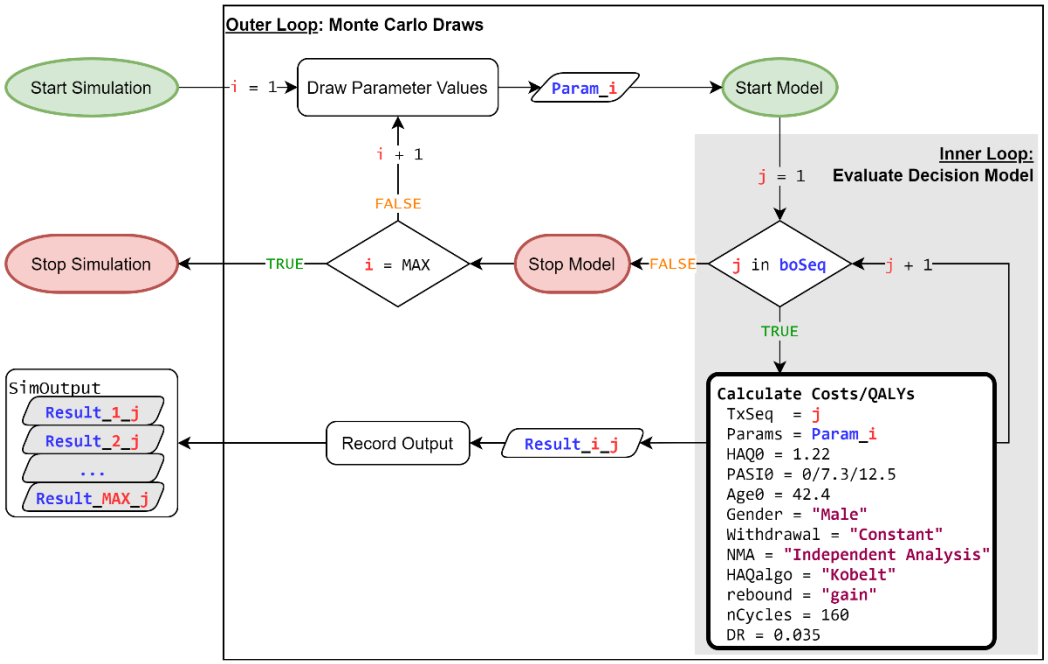
Each decision analytic model adopted a Semi-Markov structure to track a cohort with a homogeneous baseline population. Markov states were defined as the sequence of biologic treatments, followed by best-supportive care (BSC) and Death. In addition to treatment, the model also tracked arthritis and psoriasis severity across each cycle and Markov state. These estimates of symptom severity were used to calculate health utilities (EQ5D) as well as the health service costs of arthritis and psoriasis treatment. Alternate assumptions to the base case were explored using three scenario analyses. Every specified decision model was evaluated using Monte Carlo simulations of 20,000 iterations. Incremental cost-effectiveness ratios were calculated from mean costs and QALYs, and optimal treatment sequences were identified at thresholds of £20,000 and £30,000 (12).

4.3 Programming the Decision Models

A diagrammatic representation of the code used for each Monte Carlo simulation in the case study is presented in Figure 4-1. Distributions of costs and QALYs (SimOutput) were generated from the repeated execution of an inner loop (which evaluated a decision model) within an outer loop (which performed each Monte Carlo draw). Upon initiation of a simulation, the outer loop used a single function (Draw Parameter Values) to return a random draw of the model parameters (Param_i) according to their assigned distributions (12). This triggered the initiation of the inner loop, which involved the repeated execution of a function that calculated the costs and QALYs for a single comparator (Result_i_j). Each pass of the inner loop supplied a different value to the TxSeq argument, while holding every other argument constant. The total

number of inner loop repetitions was determined by comparing the value of the loop counter, j, with the collection of comparators specified for a given sub-population (boSeq). In other words, the inner loop for sub-populations 1, 2, and 3 would stop when the value of j was equal to 5, 9, and 4. This reflected a value of j which was a single increment greater than the total number of treatment sequences specified for each sub-population. Upon return to the outer loop, a new Monte Carlo draw would be performed for all parameters, and the process would be repeated until the target number of outer loop repetitions was reached (MAX = 20,000).

The ability to repeat a single function call under slightly different conditions played a critical role in the realization of the benefits from reproducibility. Most importantly, it offered a reliable and transparent mechanism to define and evaluate each Semi-Markov model for each sub-population. This was facilitated by the function arguments which were responsible for i) supplying the data required for the various computations inside the function; or ii) controlling the details of how the costs and QALYs were calculated. The only input responsible for both tasks was the treatment sequence (TxSeq). This was necessary to accommodate sub-population specific differences in the assumed sequence length without introducing any duplicate code. As described elsewhere, duplication in code can undermine reproducibility by increasing the risk of error and distorting the provenance of the results (14). A technique called abstraction was used to program each relevant task via generalized patterns rather than literal statements (14). This allowed for the reliable and consistent implementation of two key tasks, regardless of sequence length. First, each function call began by using this input to subset and re-arrange the treatment specific values within the sampled parameters (Params). Second, it also allowed the Markov states to be defined dynamically by combining the supplied values with “BSC” and “Death”.



boSeq:
Collection of Comparator Sequences for each sub-population.
Only one collection can be evaluated at a time.

Sub-Population 1	
j	
1	SEC150, ETN, UST
2	SEC300, ETN, UST
3	CZP, ETN, UST
4	BSC

Sub-Population 2			
j		j	
1	SEC150, UST	5	GOL, UST
2	SEC300, UST	6	ETN, UST
3	CZP, UST	7	INF, UST
4	ADA, UST	8	BSC

Sub-Population 3	
j	
1	SEC300
2	UST
3	BSC

Structural Assumptions: Permissible Values

Input	Accepted Values
Gender	"Male" "Female"
Withdrawal	"Constant", "Half SEC" "Half All"
NMA	"Independent Analysis" "Meta-Regression"
HAQalgo	"Kobelt" "Poole"
rebound	"gain"

Figure 4-1. Programming Logic for the Monte Carlo simulation, using the Base Case configuration. The function in the inner loop was designed to calculate costs and QALYs for each supplied parameter set (Param_i) according to a declared set of logical pathways.

To illustrate how the code responded to the value of the treatment sequence, the process to define and evaluate a Semi-Markov model is summarized in Figure 4-2. Each row depicts this process under conditions specific to the first treatment sequence from each sub-population included in the case study (see boSeq in Figure 4-1). The process began by populating the transition matrices using the pre-defined Markov States and treatment specific parameter values. Consistent with Hawkins et al., time-dependent transition probabilities were stored in an array with three dimensions: current state, future state, and time (16). In Figure 4-2, this data structure is represented as a cube to reflect the fact that the third dimension was used to distinguish a series of matrices. Subsequent function calls returned separate arrays which tracked the cohort's treatment, disease severity, costs, and utilities across the cycle, Markov state, and time-in-state dimensions. While differences in the length of each treatment sequence affected the size of each array in each sub-population, it had no impact on the methods used to calculate the results. Once the arrays storing costs and utilities were reduced to matrices, discounted values were summed across columns and rows to return the costs and QALYs corresponding to Result_i_j in Figure 4-1.

Returning to Figure 4-1, the conditions used to build and evaluate the arrays depicted in Figure 4-2 were controlled by twelve distinct inputs. Unlike the treatment sequence, the remaining eleven inputs served a single purpose. The sampled parameters (Params), baseline values (HAQ0, PASI0, Age0), cycle count (nCycles), and discount rate (DR) inputs were required for the series of computations depicted in Figure 4-2. Meanwhile the inputs representing each independent structural assumption (Gender, Withdrawal, HAQalgo, NMA, and rebound) were used to control the flow of execution. This capability was implemented by placing each relevant code segment inside of an if/else statement. To avoid the introduction of duplicate code,

these conditional statements were inserted within the relevant functions used to build one of the arrays in Figure 4-2. The explicit parameterization of each structural assumption offered a simple mechanism to switch between alternative scenarios. For example, switching from the base case to a scenario analysis in Figure 4-1 involved substituting the value supplied to the relevant input with a permissible alternative. The parameterized structural assumptions were also used to communicate meaningful information about the methods. For example, the rebound input only considered one assumption regarding the change in arthritis severity immediately following withdrawal. However, its inclusion helped avoid methodological ambiguity from assumptions used in prior evaluations (13,17).

To summarize, an emphasis on reproducibility led to the development of a generalized function which calculated costs and QALYs according to a defined set of logical pathways. However, the function itself did not reflect any specific model from the case study. Instead, it represented an abstraction of the instructions to build and evaluate a Semi-Markov model according to the conditions specified by its inputs. This implementation might correspond to the idea of a common Meta-Model described by Claxton (10).

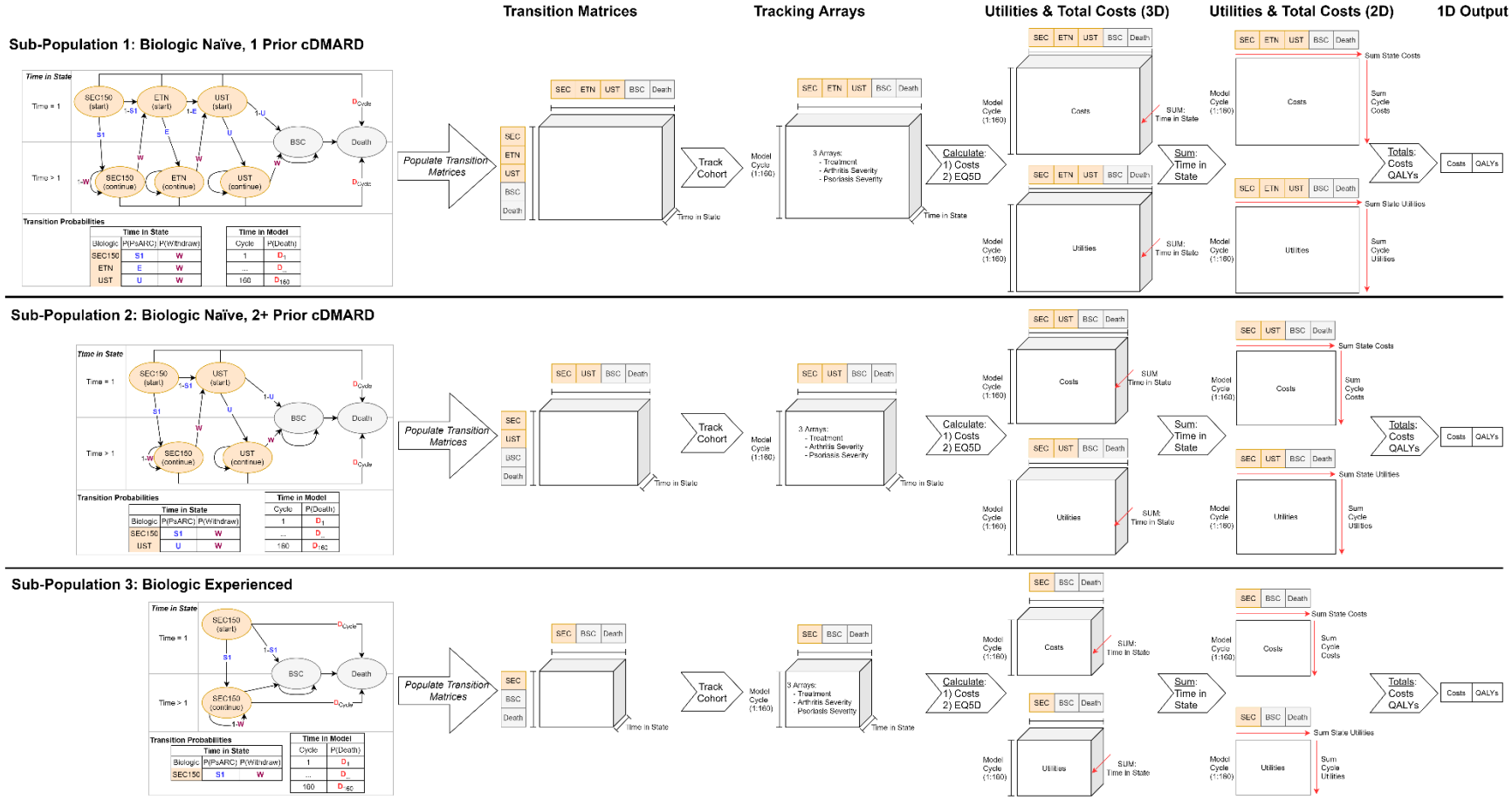


Figure 4-2. Illustration of sub-population specific conditions to calculate costs and QALYs for a single treatment sequence.

4.3.1 A Thorough Characterization of Uncertainty

The characterization of uncertainty in the case study was constrained by two distinct factors. First, assumptions relating to treatment sequence composition may have restricted the assessment of parameter uncertainty. While treatment switching was a key feature of the model, the alternatives considered in some sub-populations (Figure 4-1) reflected a small subset of the possible combinations (10). Second, the analysis strategy did not incorporate all relevant sources of uncertainty. This would have required the joint characterization of parameter and structural uncertainty. In this context, structural uncertainty refers to the impact that different assumptions or scientific judgements in the decision model can have on expected costs and effects (10,18,19). Reliance on probabilistic scenario analyses meant that the characterization of uncertainty across scenarios would be left to the implicit judgement of the decision maker (10,18,19).

The design of the function to calculate costs and QALYs offered the ability to overcome both limitations, enabling a more thorough characterization of uncertainty. As illustrated in Figure 4-1, parameterization of the treatment sequence enabled the evaluation of any number of alternatives (of varying length). Therefore, concerns regarding the characterization of parameter uncertainty could be resolved by specifying a more comprehensive set of alternatives. Meanwhile, the independent parameterization of each structural assumption could be used to jointly characterize parameter and structural uncertainty. One way to do this would involve conducting an independent scenario analysis for all 24 possible scenario configurations. Under this approach, expected values would need to be calculated as a weighted average across the merged data from each simulation (10,18,19). An equivalent alternative would be to incorporate the scenario configuration into a single Monte Carlo simulation. Illustrated in Figure 4-3, the

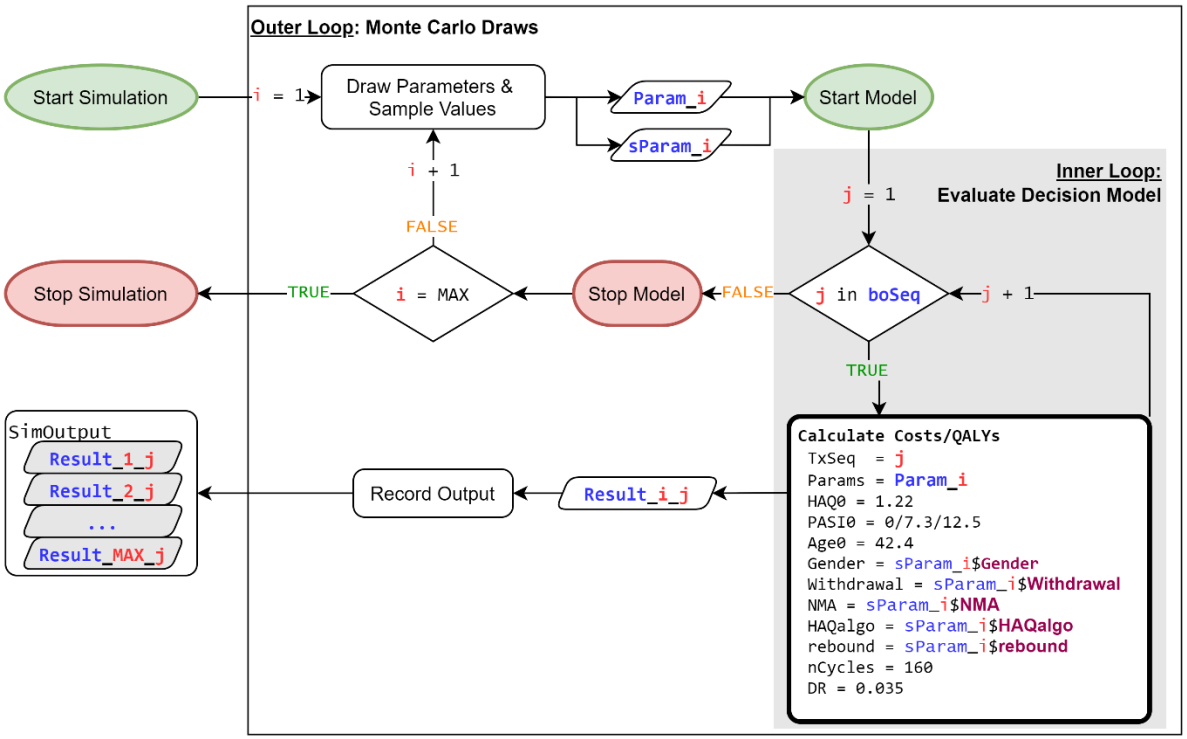
permissible values for each assumption could be sampled with replacement (according to defined probabilities) at each iteration of the simulation.

To demonstrate this capability, a Monte Carlo simulation was used to evaluate a modified decision model for Sub-Population 1. This sub-population was selected because it had the longest treatment sequence (3 biologics) and considered the smallest subset of sequence combinations in the case study. The purpose of this exercise was to highlight how an emphasis on reproducibility led to programmed behaviours which could enable a more thorough characterization of uncertainty. While this necessitated a larger number of comparator sequences from the case study, the identification of an optimal sequence was beyond the scope of this paper. Such a task would require the incorporation of additional evidence on degradation effects from switching between biologics (13).

A conservative rule set was used to expand the number of alternatives from four to twenty-one sequences of biologic treatments (Appendix C). The joint characterization of parameter and structural uncertainty was achieved using the procedures outlined in Figure 4-3. As with the case study, the model assumed a price year of 2016 and a discount rate of 3.5%. Applying this approach in a formal decision-context will require a meaningful number of Monte Carlo draws for each scenario configuration. As this exercise was conducted as a demonstration, only 20,000 iterations were considered – equivalent to approximately 833 iterations for each of the 24 scenario configurations.

The Cost-Effectiveness Acceptability Curves (CEAC) in Figure 4-4 offer a comparison of the characterization of uncertainty with the original case study. In every CEAC, the coloured lines represent the probability that a treatment sequence is cost-effective across a range of values for the cost-effectiveness threshold (λ). Panel A represents the characterization of uncertainty

from the case study. Each column represents a scenario, whereas rows represent the psoriasis sub-group stratified by the evidence synthesis strategy. Meanwhile, the CEACs presented in Panel B reflect the results from the joint characterization of uncertainty using the expanded set of treatment sequences for each psoriasis sub-group. While 21 alternatives were specified, many had to be excluded ($P_{\lambda} = 0.00$) due to sub-group specific dosing restrictions. Nevertheless, the flexibility of the code used to generate this information further supports the notion that the characterization of uncertainty will be much more dependent on the decision model's programming than its structure.



boSeq:
Collection of Comparator Sequences for each sub-population.

Sub-Population 1	
j	
1	SEC150,
2	SEC300,
3	CZP, ETN,
...	...
∞	BSC

Structual Assumptions: Permissible Values

Input	Accepted Values	Probability
Gender	"Male" "Female"	0.50 0.50
Withdrawal	"Constant", "Half SEC" "Half All"	0.33 0.33 0.33
NMA	"Independent Analysis"	0.50 0.50
HAQalgo	"Kobelt" "Poole"	0.50 0.50
rebound	"gain"	1

Figure 4-3. Programming logic for a Monte Carlo simulation which jointly characterizes parameter and structural uncertainty. At each iteration of the outer loop, a random sample of parameter values are drawn from assigned distributions (Param_i) and values for each structural assumption (sParam_i) are sampled with replacement.

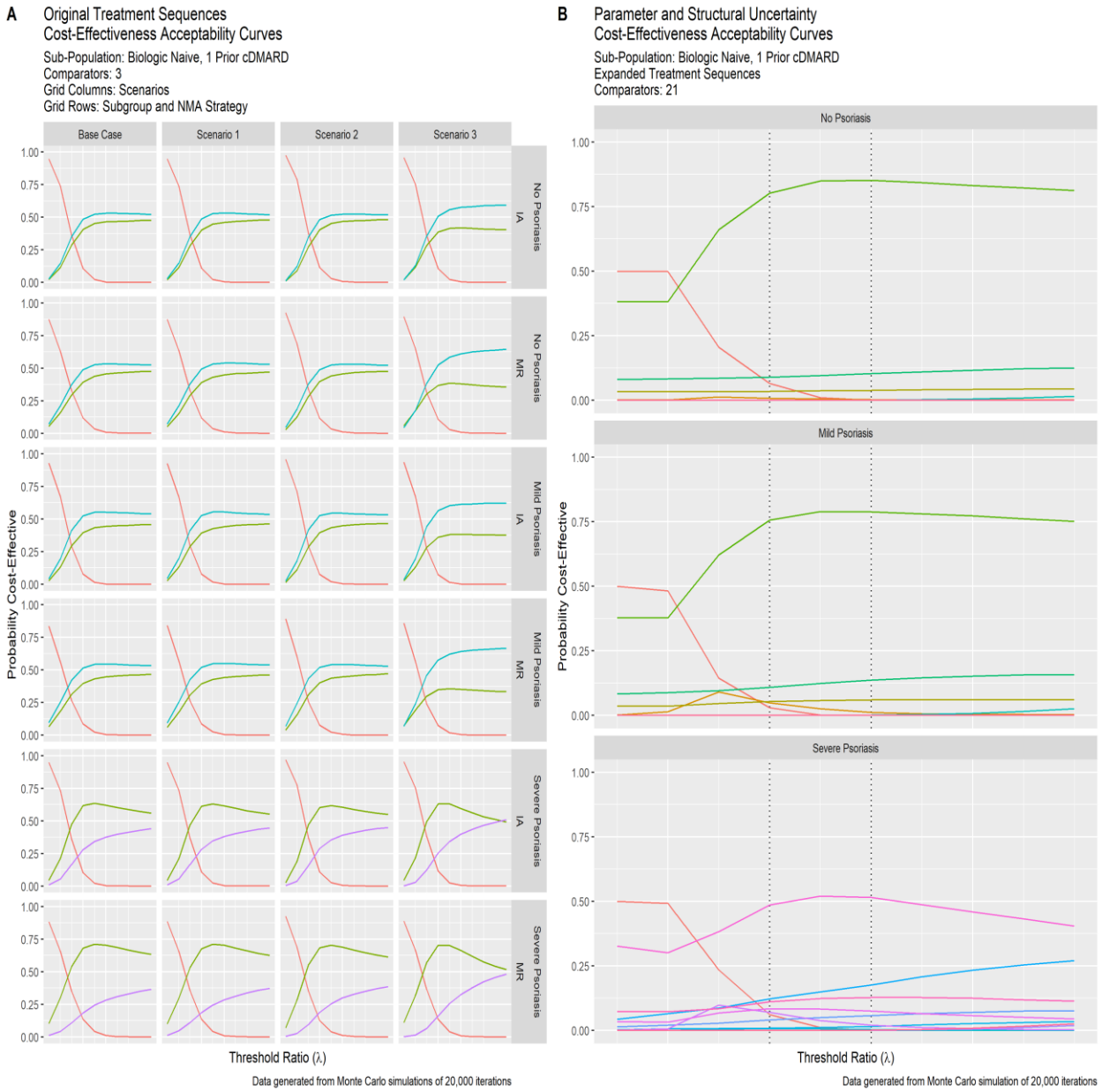


Figure 4-4. Comparison of CEACs for the Biologic Naïve Sub-Population with 1 Prior cDMARD. Panel A represents the CEACs generated from the case study (Chapter 3) with columns reflecting each scenario and rows reflecting the Psoriasis sub-groups stratified by NMA combination (IA: Independent Analysis; MR: Meta-Regression). Panel B represents the CEACs using an expanded set of comparators generated from a Monte Carlo simulation which incorporates structural uncertainty. Additional iterations are required to fully characterize the uncertainty between the alternatives in Panel B.

4.4 Overcoming Computing Challenges

Monte Carlo simulation is the most common method used to characterize decision uncertainty or explore its consequences. As highlighted above, it relies on the repeated evaluation of a decision model using randomly sampled parameters to generate distributions of cost and effect for each specified alternative (10,19,20). Many repetitions will be required for this output to reflect the range of values each parameter is likely to take. Given the considerable amount of computation involved, using Monte Carlo simulation can be a time-consuming effort for even the simplest of decision models (10,19,20). However, the exact amount of time will depend on the complexity of the model, the number of repetitions involved, how the methods are programmed (some approaches are faster than others), and the specific computer used to complete the simulation (some computers are faster than others). This suggests that the programming of a decision model must satisfy two distinct objectives. First, the instructions must reflect a correct implementation of the methods to calculate costs and QALYs. Second, the implementation must also be *fast enough* to allow a computer to generate the dataset of interest in a reasonable amount of time.

Achieving reasonable execution times for each Monte Carlo simulation required the effective use of available computational resources. Given that this project was developed in R, this involved the component responsible for performing calculations (Central Processing Unit; CPU); and the memory component used to store information (Random Access Memory; RAM) (21). As described elsewhere, meaningful improvements in computational efficiency were realized from programming choices which affected calculation speed (vectorization) and memory efficiency (data structures) (12,22,23). Another factor which affected memory efficiency was modularization. Organizing code into functions meant that an intermediate

collection of values was only preserved if it was required. For example, the collections of values included in Figure 4-2 were discarded as soon as the function to calculate costs and QALYs (Figure 4-1 and Figure 4-3) returned its output.

Ultimately, the ability to complete a simulation and the resulting execution time will depend on the physical computer hardware. It is critical to ensure one will have enough memory (RAM) to complete a simulation. While the incremental generation of results will slow the simulation down, one must ensure there will be enough memory to store the generated data as well as any intermediate values. Eventually, any task will fail when the amount of available memory is exceeded. One can anticipate the size of the output dataset as a function of the number of comparators and Monte Carlo draws. For example, each pass of the inner loop in Figure 4-1 returned costs and QALYs for all three psoriasis sub-groups. At 20,000 iterations, the model with four comparators required 80,000 function calls to generate 240,000 rows of output. Memory requirements will be even larger when conducting simulations for Value of Information analysis. In the case study, this effort required 1,000,000 function calls to generate 12,000,000 rows of output (12). This was managed by reformatting the output data frame requiring 1.2 Gigabytes of memory to an array which required 187 Megabytes to store the same information.

Another way to improve execution time without making the code more efficient is to use parallel computation. Instead of executing a Monte Carlo simulation sequentially on one CPU (the default), each independent iteration (the outer loop in Figure 4-1) can be distributed across multiple CPUs of a multi-cored personal computer or a high-performance cluster (22,24). Due to the reproducible design, implementation only required an appropriate looping construct, like those offered by the parallel and foreach packages (15,25).

4.4.1 Recommendations

Given the many factors which can affect the execution time of a Monte Carlo simulation, it can be difficult to know which to prioritize. The most productive approach will be to start with the development of a reproducible decision model. Concerns about the generalizability of the strategies to do so should be mitigated from their use across many scientific fields (12,14,26,27). More advanced techniques, like performance optimization and parallel computation should be considered following the validation of a working prototype (22). Understanding how to use these techniques across different model structures, such as patient level simulation, remains an important area for future research.

4.5 Conclusion

To provide unbiased estimates of relative cost-effectiveness, the stochastic evaluation of a decision model must incorporate all relevant sources of uncertainty. Unfortunately, such efforts are often abandoned when simulations take too long to complete. This barrier is a result of the way the methods were communicated to the computer. Enhancing the reproducibility of an economic evaluation led to the elimination of this barrier in three unique ways. First, the programming techniques intended to make it simple for a human to reproduce the results also enabled the joint characterization of parameter and structural uncertainty. Second, the design of the code encouraged a more efficient use of computational resources. Third, it revealed the impact that the hardware itself can have on completing a simulation in a reasonable timeframe. As such techniques are model agnostic, reproducibility offers a pathway to the consistent and thorough characterization of uncertainty in HTA.

4.6 References

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. 461 p.

2. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Serv Res. 2013 Sep 30;13(1):371.

3. Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, Drummond M. Evidence synthesis, economics and public policy. Res Synth Methods. 2010 Apr 1;1(2):126–35.

4. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada [Internet]. Ottawa, ON; 2017 Mar. Report No.: 4th ed. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition

5. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013. 94 p. (NICE Process and Methods Guides). Available from: http://www.ncbi.nlm.nih.gov/books/NBK395867/

6. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? Value Health. 2006;9(4):244–52.

7. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

8. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessent: the role of iterative decision analytic modelling. 2000 May; Available from: https://core.ac.uk/reader/6406440

9. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

10. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. PharmacoEconomics. 2008 Sep 1;26(9):781–98.

11. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14(4):339–47.

12. Wagner DJ. Impacts from Enhancing the Reproducibility of Health Technology Assessments: A Case Study in Psoriatic Arthritis. University of Calgary; 2021.

13. Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. Health Technol Assess. 2017 Oct;21(56):1–326.

14. Wagner DJ. Computing Strategies to Enhance the Reproducibility of Health Technology Assessments. University of Calgary; 2021.

15. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

16. Hawkins N, Sculpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. Med Decis Making. 2016 Sep 5;25(5):511–9.

17. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess [Internet]. 2011 Feb;15(10). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta15100/

18. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. Value Health. 2009 Jul 1;12(5):739–49.

19. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 11: Characterizing, Reporting, and Interpreting Uncertainty. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. p. 353–88.

20. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation: A Practical Approach. Med Decis Making. 1985 Jun;5(2):157–77.

21. Murrell P. Introduction to Data Technologies [Internet]. CRC Press; 2009. 445 p. Available from: https://www.stat.auckland.ac.nz/~paul/ItDT/

22. Visser MD, McMahon SM, Merow C, Dixon PM, Record S, Jongejans E. Speeding Up Ecological and Evolutionary Computations in R; Essentials of High Performance Computing for Biologists. PLOS Comput Biol. 2015 Mar 26;11(3):e1004140.

23. 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 Apr 1;38(3):400–22.

24. McCallum E, Weston S. Parallel R. O'Reilly Media, Inc.; 2011. 123 p.

25. Wallig M, Microsoft, Weston S. foreach: Provides Foreach Looping Construct [Internet]. 2020. Available from: https://CRAN.R-project.org/package=foreach

26. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019. Available from: https://www.nap.edu/catalog/25303

27. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018. Available from: http://www.practicereproducibleresearch.org/

Chapter 5 Discussion

This thesis explored how an emphasis on reproducibility can support the effective development and maintenance of Health Technology Assessments. At the time of this writing, the use of economic evaluation within a continuous process of evidence gathering and decision making has been constrained by two distinct issues. First, the cost of generating evidence identified from a Value of Information analysis (within a reasonable timeframe for decision making) can often be prohibitive. Second, commissioned appraisals have been developed at sub-optimal levels of reproducibility. As a result, authors will be limited in their ability to update previous efforts with new information which may affect a model's structure or parameters. This latter issue will limit any decision-making process, including those which do not formally require VoI methods, probabilistic analyses for adoption decisions, or those which have explicit timelines for considering new information. Implementation has been constrained by the widespread, yet ineffective, use of computing in the research process. As a result, most projects lack a complete record of the specific steps taken to collect, transform, and analyze data in each part of the methodological sequence. Without this information, the risk of error will increase, an author's ability to detect errors will fall, and HTA iteration will require a duplication of the original effort. These outcomes can undermine the efficient allocation of health system resources in two distinct ways. First, the high marginal cost of iteration will commit resources which could be allocated to other appraisals or health system activities. Second, delayed and incorrect decisions may impose a health loss on current and future patients as well as increased health system costs.

Findings from this thesis revealed that these opportunity costs can be avoided by enhancing the standard level of reproducibility for HTAs. However, there was no single source

of guidance to direct how a computer could be used to achieve this goal. To address this gap, Chapter 2 used factors known to affect reproducibility to identify and organize strategies that had proven successful in other scientific fields (1,2). The results reflected two broad requirements for reproducibility. First, one must maintain an accurate and complete record of the process used to create each reportable result. The automated workflow fulfills this role as an editable, auditable, and authoritative record of the project methods. Second, additional contextual information should supplement the code and data to make it meaningful over the long run. Strategies for project management and documentation allow authors to return and re-use projects after a period of dormancy.

Results from the systematic review in Chapter 2 revealed that a significant investment in computing skills would be required to enhance the reproducibility of HTAs. The payoffs from this investment were explored in a case study which redeveloped a NICE commissioned systematic review and economic evaluation of biologic treatment for psoriatic arthritis (3). This appraisal was selected because it had been subjected to iterative revision and it considered a complex model structure which limited efforts to characterize uncertainty (3). Each identified computing strategy informed the creation of an automated workflow, with the modularized commands organized into three distinct activities: i) estimation of model parameters from raw data; ii) simulation for an adoption or research decision; and iii) post-processing of simulation results. Due to the multi-level effects from efforts to enhance the reproducibility of this HTA, findings were separated into two distinct chapters.

Chapter 3 explored the impacts of the reproducible computing strategies at the project level. This paper documented the development and execution of the workflow using the same decision models and parameter inputs from the original technology appraisal. Procedural

outcomes reflected the benefits of using a computer, rather than a human, to execute a collection of tasks. The modularized nature of the automated workflow offered a reliable mechanism to detect for and mitigate the risk of error. In addition, capturing the estimation of the input parameters within the scope of the workflow served to drastically reduce the marginal cost of iteration. Meanwhile, the redevelopment did not face any of the original appraisal's constraints in the characterization of uncertainty. This methodological outcome was primarily attributable to improvements in computational efficiency.

Chapter 4 explored the impact of the reproducible computing strategies on the programming of the decision models in the case study. Adherence to the reproducible research strategies necessitated the elimination of all duplicate code across fifteen decision models. This led to the development of a function which could define and evaluate a cohort model according to the conditions specified by its inputs. The mechanics of this function were subsequently used to argue that most barriers to the characterization of uncertainty reflect a model's programming rather than structure. This point was illustrated in two distinct ways. First, the flexible behaviour of the function was used to demonstrate the ease with which parameter and structural uncertainty could be jointly characterized. Second, a discussion of the different factors that can affect the execution time of a simulation highlighted how the function consumed different computational resources.

The research presented in this thesis confirmed that enhancements to current reproducibility standards can support the policy objectives of HTA. However, it would be unrealistic to expect the practice of reproducible research to be instantly embraced by researchers involved in the preparation of technology appraisals. Widespread implementation will require significant behaviour changes to the ways computers are used to organize, manage, and execute

research projects. While the primary responsibility for reproducibility lies with individual researchers, external parties such as research institutions and HTA agencies will also have a role to play (1,2). Future research leveraging knowledge from implementation science may prove useful in understanding the barriers and facilitators to this objective (4). It remains to be seen if the required programming and data management skills will be embraced by individuals or groups of collaborators. In the long run, these efforts will inform the way projects are managed and direct opportunities for future research.

5.1 Project Management

One topic which was under-represented in the thesis was the impact the project management strategy had on reproducibility. The initial plan was to place every file created for the case study in a single directory which was structured in accordance with the approach described in Chapter 2. This meant that executing a specific task would require the user to know each specific dependency for that process. However, it became difficult to keep track of the inter-dependent files which not only defined the workflow but also the various scripts and outputs created from its execution. Furthermore, this approach to file management made the project challenging to maintain. The implementation of simple changes would often cause a task to fail and would require hours of work to identify and resolve the issue. As a result, the project was at risk of becoming too difficult to re-use.

In response, the collection of computer files which defined the workflow were converted into an R package (5,6). Details on the acquisition and use of the code and data for this package are reported in Appendix D. Separating the specification and execution of the workflow into distinct directories made the package re-usable, while also protecting the source code and raw data from accidental changes. This strategy proved useful for several additional reasons. First,

the quality control mechanisms in R package creation helped to ensure that the iterative development of the workflow always produced code that worked. Second, development of the package allowed for safe prototyping of new functions or features. Changes to the source code or data would not affect the working code until a new package version was "installed". Third, the R package structure allowed for functions and datasets to be bundled together. This allowed the raw data for the workflow to be included in the package, saving configuration time for the user. Fourth, project versioning promoted the ability to re-use code from any point in the project's development. Fifth, the code to execute the workflow was no longer confined to a specific folder on a specific computer. While the package was intended for personal use, it could be shared with any R user on any computer (5,6).

While the creation of this R package was omitted from Chapter 3, much of the methods section described its development. This was an attempt to avoid the impression that the production of reproducible research requires the use of the R language or the creation of an R package. As indicated in Chapter 2, enhancing the reproducibility of an HTA will require the creation of a human and machine-readable record which captures the provenance of the results. Apart from spreadsheets, there are lots of alternative scripting tools and programming languages which could be used to achieve this goal (2,7,8). Emphasizing reproducibility will force researchers to consider a tool which can accommodate the procedural and methodological requirements of every distinct task in a project workflow (2,9).

A commitment to reproducibility will also promote tool specific conventions for the development of HTAs. One idea which will be explored further is the production of project-specific R packages, like the one created for this case study. Given the high degree of procedural overlap between projects, the ability to re-use code or module designs might improve the

efficiency of new project development. For example, the parameterization of the treatment sequence highlighted in Chapter 4 is a design feature which could be applied to other projects which require treatment switching. In addition, the code developed to support the post-processing of simulation results could also be re-used with little modification. To eliminate duplication between projects, a separate package could be created to bundle functions which calculate ICERs, Net-Benefit and Value of Information (VoI) statistics, as well as tables and figures for reporting. This approach would require the consistent formatting of datasets containing distributions of costs and effects for each mutually exclusive alternative in any decision model. In a group context, creation of a style guide could serve as a centralized location to document these standards as they evolve. Therefore, a commitment to reproducibility may also aid in promoting a community of practice and eventually more open science.

5.2 Future Research

Enhancing the reproducibility of Health Technology Assessments will also present new opportunities for future research. One topic which emerged from the development of this thesis was the absence of a decision rule for updating an evidence synthesis. In a context of uncertainty, it may be necessary to revise an adoption or research decision using information acquired from a primary study or an updated evidence synthesis (10–13). Both forms of research will impose independent opportunity costs on the health budget, reducing the pot of resources available to other activities. Given that market access is often sought when an evidence base is least mature, the timing of a literature search (and the resulting decision uncertainty) may affect the outcome of a research decision. This suggests it might be a good idea to update a literature search prior to committing resources to a primary study (13). However, the research decision does not explicitly consider this possibility. In other words, there is no mechanism to determine if updating an

evidence synthesis alone will be an efficient use of research resources. As a result, the research decision fails to consider the possibility that updating an evidence synthesis may be a more efficient use of resources compared to a primary study.

The above research gap was exposed from the efficiency gains realized from investments in reproducibility. Development of an automated workflow offered the capability to incorporate new information at a much lower cost compared to current standards. With enough investment, further cost reductions could be achieved by automating additional project tasks. Therefore, a decision to update an evidence synthesis could be justified if the marginal cost of iteration is less than the value of the remaining project budget. Like a traditional research decision, this approach would capture the opportunity cost imposed by this type of research. However, it might imply that an adoption decision should be revised every time relevant information is released into the public domain. To avoid the risk of decision reversal, standards for pre-registration could be used to delay a decision until the anticipated release of new information. Careful consideration will need to be given to the trade-offs faced by current and future patients from such deferrals. Unlike waiting for the emergence of an "ideal" evidence base, planning for the release of relevant information may be an appropriate approach to managing the entry of new technologies (14,15).

5.3 References

1. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019. Available from: https://www.nap.edu/catalog/25303

2. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018. Available from: http://www.practicereproducibleresearch.org/

3. Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. Health Technol Assess. 2017 Oct;21(56):1–326.

4. Damschroder L, Hall C, Gillon L, Reardon C, Kelley C, Sparks J, et al. The Consolidated Framework for Implementation Research (CFIR): progress to date, tools and resources, and plans for the future. Implement Sci. 2015 Aug 20;10(Suppl 1):A12.

5. R Core Team. Writing R Extensions (Version 4.1.1) [Internet]. 2021. Available from: https://cran.r-project.org/doc/manuals/r-release/R-exts.pdf

6. Wickham H, Bryan J. R Packages [Internet]. 2019. Available from: https://r-pkgs.org/

7. 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 May 1;22(5):575–9.

8. Baio G, Heath A. When Simple Becomes Complicated: Why Excel Should Lose its Place at the Top Table. Glob Reg Health Technol Assess. 2017 Jan 1;4(1):grhta.5000247.

9. Carey MA, Papin JA. Ten simple rules for biologists learning to program. PLOS Comput Biol. 2018 Jan 4;14(1):e1005871.

10. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

11. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessent: the role of iterative decision analytic modelling. 2000 May; Available from: https://core.ac.uk/reader/6406440

12. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

13. Wilson E, Abrams K. From evidence-based economics to economics-based evidence: using systematic review to inform the design of future research. In: Shemilt I, Donaldson C, Marsh

K, Mugford M, Vale L, editors. Event-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated; 2010.

14. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, et al. A Comprehensive Algorithm for Approval of Health Technologies With, Without, or Only in Research: The Key Principles for Informing Coverage Decisions. Value Health. 2016 Sep 1;19(6):885–91.

15. Sculpher M, Claxton K. Establishing the Cost-Effectiveness of New Pharmaceuticals under Conditions of Uncertainty—When Is There Sufficient Evidence? Value Health. 2005 Jul;8(4):433–46.

Conclusion

This thesis identified and implemented a series of computing strategies to enhance the reproducibility of Health Technology Assessments. In principle, this will require the provenance of the results to be captured in human and machine-readable formats. In practice, it involved significant changes to the way computers are used to organize, manage, and execute systematic reviews and economic evaluations. Although the case study presented in this thesis is restricted to a single decision problem, the identified strategies for reproducibility will be generalizable to other contexts. However, developing the necessary programming and data management skills will be an exercise in patience and deliberate practice. Committing to a practice of reproducible research will allow researchers (or teams) to improve their productivity, mitigate the risk of error, and consider more ambitious analyses. This will translate to better decision making, improved transparency, more efficient updating, and ultimately population health gains.

Bibliography

1. Ades AE, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. Med Decis Making. 2004 Mar 1;24(2):207–27.

2. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, Jalal H, Kao S-YZ, Yang A, et al. A Need for Change! A Coding Framework for Improving Transparency in Decision Modeling. PharmacoEconomics [Internet]. 2019 Sep 24; Available from: http://link.springer.com/10.1007/s40273-019-00837-x

3. Antoni CE, Kavanaugh A, Heijde D van der, Beutler A, Keenan G, Zhou B, et al. Two-year efficacy and safety of infliximab treatment in patients with active psoriatic arthritis: findings of the Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT). J Rheumatol. 2008 May 1;35(5):869–76.

4. Antoni CE, Kavanaugh A, Kirkham B, Tutuncu Z, Burmester GR, Schneider U, et al. Sustained benefits of infliximab therapy for dermatologic and articular manifestations of psoriatic arthritis: Results from the infliximab multinational psoriatic arthritis controlled trial (IMPACT). Arthritis Rheum. 2005 Apr 1;52(4):1227–36.

5. Antoni CE, Krueger GG, de Vlam K, Birbara CA, Beutler A, Guzzo C, et al. Infliximab improves signs and symptoms of psoriatic arthritis: results of the IMPACT 2 trial. Ann Rheum Dis. 2005 Aug 1;64(8):1150–7.

6. Arnold RJG, Ekins S. Time for Cooperation in Health Economics among the Modelling Community: PharmacoEconomics. 2010 Aug;28(8):609–13.

7. Baio G, Heath A. When Simple Becomes Complicated: Why Excel Should Lose its Place at the Top Table. Global & Regional Health Technology Assessment. 2017 Jan 1;4(1):grhta.5000247.

8. Banta D. The development of health technology assessment. Health Policy. 2003 Feb 1;63(2):121–32.

9. Baumer BS, Kaplan DT, Horton NJ. 5) Tidy Data and Iteration. In: Modern Data Science with R [Internet]. Boca Raton, UNITED STATES: CRC Press LLC; 2017. Available from: http://ebookcentral.proquest.com/lib/ucalgary-ebooks/detail.action?docID=5475748

10. Bleichrodt H. QALYs and HYEs: Under what conditions are they equivalent? Journal of Health Economics. 1995 May 1;14(1):17–37.

11. Blischak JD, Davenport ER, Wilson G. A Quick Introduction to Version Control with Git and GitHub. PLOS Computational Biology. 2016 Jan 19;12(1):e1004668.

12. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. Value in Health. 2009 Jul 1;12(5):739–49.

13. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006. 237 p. (Briggs A, Gray A, editors. Oxford handbooks in health economic evaluation).

14. Broman KW, Woo KH. Data organization in spreadsheets. The American Statistician. 2018 Jan 2;72(1):2–10.

15. Bryan J. Excuse Me, Do You Have a Moment to Talk About Version Control? The American Statistician. 2018 Jan 2;72(1):20–7.

16. Buneman P, Khanna S, Tan W-C. Data Provenance: Some Basic Issues. In: Kapoor S, Prasad S, editors. FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science. Berlin, Heidelberg: Springer; 2000. p. 87–93. (Lecture Notes in Computer Science).

17. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada [Internet]. Ottawa, ON; 2017 Mar. Report No.: 4th ed. Available from: https://www.cadth.ca/guidelines-economic-evaluation-health-technologies-canada-4th-edition

18. Carey MA, Papin JA. Ten simple rules for biologists learning to program. PLOS Computational Biology. 2018 Jan 4;14(1):e1005871.

19. Catalá-López F, Caulley L, Ridao M, Hutton B, Husereau D, Drummond MF, et al. Reproducible research practices, openness and transparency in health economic evaluations: study protocol for a cross-sectional comparative analysis. BMJ Open. 2020 Feb 1;10(2):e034463.

20. Choy EHS, Gladman DD, Sasso EH. Efficacy of adalimumab by disease duration in psoriatic arthritis: Subanalysis of ADEPT. In: Journal of the American Academy of Dermatology [Internet]. Washington, DC: American Academy of Dermatology; 2007. p. AB186. (Poster Abstracts; vol. 56). Available from: http://www.sciencedirect.com/science/article/pii/S0190962206037352

21. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics. 1999 Jun 1;18(3):341–64.

22. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. Pharmacoeconomics. 2008 Sep 1;26(9):781–98.

23. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research. Pharmacoeconomics. 2006 Nov 1;24(11):1055–68.

24. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, Soares M, et al. A Comprehensive Algorithm for Approval of Health Technologies With, Without, or Only in Research: The Key Principles for Informing Coverage Decisions. Value in Health. 2016 Sep 1;19(6):885–91.

25. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Economics. 2005;14(4):339–47.

26. Cohen JT, Neumann PJ, Wong JB. A Call for Open-Source Cost-Effectiveness Analysis. Annals of Internal Medicine. 2017;167(6):432–3.

27. Cohen JT, Wong JB. Can Economic Model Transparency Improve Provider Interpretation of Cost-Effectiveness Analysis? A Response: Medical Care. 2017 Nov;55(11):912–4.

28. Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. Health Technology Assessment. 2017 Oct;21(56):1–326.

29. Corbett M, Sideris E, Palmer S, Harden M, Woolacott N, Bojke L. Apremilast for treating active psoriatic arthritis. York, UK: CRD and CHE Technology Assessment Group, University of York; 2015 Mar p. 150. (Single Technology Appraisal). Report No.: 13/169/01.

30. Craig D, O'Connor J, Rodgers M, Rodriguez-Lopez R, Smith A, Woolacott N. Ustekinumab for treating active and progressive psoriatic arthritis [Internet]. York, UK: CRD and CHE Technology Assessment Group, University of York; 2013 Oct p. 187. (Single Technology Appraisal). Report No.: HTA 12/58. Available from: https://www.journalslibrary.nihr.ac.uk/programmes/hta/125801/#/

31. Cutolo M, Myerson GE, Fleischmann RM, Lioté F, Díaz-González F, Bosch FV den, et al. A Phase III, Randomized, Controlled Trial of Apremilast in Patients with Psoriatic Arthritis: Results of the PALACE 2 Trial. J Rheumatol. 2016 Sep 1;43(9):1724–34.

32. Damschroder L, Hall C, Gillon L, Reardon C, Kelley C, Sparks J, et al. The Consolidated Framework for Implementation Research (CFIR): progress to date, tools and resources, and plans for the future. Implementation Science. 2015 Aug 20;10(Suppl 1):A12.

33. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation: A Practical Approach. Med Decis Making. 1985 Jun;5(2):157–77.

34. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 9: Economic Evaluation Using Decision Analytic Modelling. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. p. 311–51.

35. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Chapter 11: Characterizing, Reporting, and Interpreting Uncertainty. In: Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. p. 353–88.

36. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015. 461 p.

37. Dudley JT, Butte AJ. A Quick Guide for Developing Effective Bioinformatics Programming Skills. PLOS Computational Biology. 2009 Dec 24;5(12):e1000589.

38. Dunlop WCN, Mason N, Kenworthy J, Akehurst RL. Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. PharmacoEconomics. 2017 Jan 1;35(1):125–8.

39. 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 Sep 1;32(5):733–43.

40. Edwards CJ, Blanco FJ, Crowley J, Birbara CA, Jaworski J, Aelion J, et al. Apremilast, an oral phosphodiesterase 4 inhibitor, in patients with psoriatic arthritis and current skin involvement: a phase III, randomised, controlled trial (PALACE 3). Ann Rheum Dis. 2016 Jun 1;75(6):1065–73.

41. Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. Journal of Clinical Epidemiology. 2017 Nov 1;91:23–30.

42. Ellis SE, Leek JT. How to share data for collaboration. The American Statistician. 2018 Jan 2;72(1):53–7.

43. Endel F, Piringer H. Data Wrangling: Making data useful again. IFAC-PapersOnLine. 2015 Jan 1;48(1):111–2.

44. Fenwick E, Claxton K, Sculpher M, Briggs A. Improving the efficiency and relevance of health technology assessent: the role of iterative decision analytic modelling. 2000 May; Available from: https://core.ac.uk/reader/6406440

45. Fenwick E, Palmer S, Claxton K, Sculpher M, Abrams K, Sutton A. An Iterative Bayesian Approach to Health Technology Assessment: Application to a Policy of Preoperative Optimization for Patients Undergoing Major Elective Surgery. Med Decis Making. 2006 Sep 1;26(5):480–96.

46. Fenwick E, Steuten L, Knies S, Ghabri S, Basu A, Murray JF, et al. Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. Value in Health. 2020 Feb 1;23(2):139–50.

47. Forman J, Damschroder L. Qualitative Content Analysis. In: Jacoby L, Siminoff LA, editors. Empirical methods for bioethics: a primer. 1. ed. Amsterdam: Elsevier JAI; 2008. (Advances in bioethics).

48. Genovese MC, Mease PJ, Thomson GTD, Kivitz AJ, Perdok RJ, Weinberg MA, et al. Safety and efficacy of adalimumab in treatment of patients with psoriatic arthritis who had failed disease modifying antirheumatic drug therapy. J Rheumatol. 2007 May 1;34(5):1040–50.

49. Gentzkow M, Shapiro JM. Code and Data for the Social Sciences: A Practitioner's Guide [Internet]. University of Chicago mimeo; 2014. Available from: https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf

50. Gladman DD, Mease PJ, Cifaldi MA, Perdok RJ, Sasso EH, Medich J. Adalimumab improves joint-related and skin-related functional impairment in patients with psoriatic arthritis: patient-reported outcomes of the Adalimumab Effectiveness in Psoriatic Arthritis Trial. Ann Rheum Dis. 2007 Feb 1;66(2):163–8.

51. Gladman DD, Mease PJ, Ritchlin CT, Choy EHS, Sharp JT, Ory PA, et al. Adalimumab for long-term treatment of psoriatic arthritis: Forty-eight week data from the adalimumab effectiveness in psoriatic arthritis trial. Arthritis Rheum. 2007 Jan 30;56(2):476–88.

52. Gladman D, Mease P, Kavanaugh A, Weinberg M. Adalimumab is efficacious in treating skin disease in psoriatic arthritis: subanalysis of moderate versus severe psoriasis in the ADEPT trial. In: American Academy of Determatology 64th Annual Meeting. San Francisco, California: American Academy of Dermatology; 2006. p. AB10–1.

53. Gladman D, Mease P, Ritchlin C, Okun M. PASI-100 is associated with better dermatology-specific patient-reported outcomes compared with PASI-75–99 in adalimumab-treated patients with psoriatic arthritis: Subanalysis of ADEPT. In: Journal of the American Academy of Dermatology [Internet]. Washington, DC: American Academy of Dermatology; 2007. p. AB185. (Poster Abstracts; vol. 56). Available from: http://www.sciencedirect.com/science/article/pii/S0190962206037339

54. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten Simple Rules for the Care and Feeding of Scientific Data. PLOS Computational Biology. 2014 Apr 24;10(4):e1003542.

55. Gottlieb AB, Langley RG, Philipp S, Sigurgeirsson B, Blauvelt A, Martin R, et al. Secukinumab Improves Physical Function in Subjects With Plaque Psoriasis and Psoriatic Arthritis: Results from Two Randomized, Phase 3 Trials. J Drugs Dermatol. 2015 Aug;14(8):821–33.

56. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Information & Libraries Journal. 2009;26(2):91–108.

57. Green T. We Need Publishing Standards for Datasets and Data Tables. OECD Publishing. 2009;

58. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? Value in Health. 2006;9(4):244–52.

59. Hampton SE, Jones MB, Wasser LA, Schildhauer MP, Supp SR, Brun J, et al. Skills and Knowledge for Data-Intensive Environmental Research. BioScience. 2017 Jun 1;67(6):546–57.

60. Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten Simple Rules for Digital Data Storage. PLOS Computational Biology. 2016 Oct 20;12(10):e1005097.

61. Hatswell AJ, Chandler F. Sharing is Caring: The Case for Company-Level Collaboration in Pharmacoeconomic Modelling. PharmacoEconomics. 2017 Aug;35(8):755–7.

62. Hawkins N, Sculpher M, Epstein D. Cost-Effectiveness Analysis of Treatments for Chronic Disease: Using R to Incorporate Time Dependency of Treatment Response. Medical Decision Making. 2016 Sep 5;25(5):511–9.

63. Heath A, Baio G. Calculating the Expected Value of Sample Information Using Efficient Nested Monte Carlo: A Tutorial. Value in Health. 2018 Nov 1;21(11):1299–304.

64. Hejblum BP, Kunzmann K, Lavagnini E, Hutchinson A, Robertson DS, Jones SC, et al. Realistic and Robust Reproducible Research for Biostatistics. Preprints.org [Internet]. 2020 Jun 2; Available from: https://www.preprints.org/manuscript/202006.0002/v1

65. Hinde S, Spackman E. Bidirectional Citation Searching to Completion: An Exploration of Literature Searching Methods. PharmacoEconomics. 2015 Jan 1;33(1):5–11.

66. Hollman C, Paulden M, Pechlivanoglou P, McCabe C. A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. PharmacoEconomics. 2017 Aug;35(8):817–30.

67. 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 Jun;37(6):829–43.

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 in Health. 2019 May 1;22(5):575–9.

69. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. Medical Decision Making. 2017 Jan 6;37(7):735–46.

70. Jansen JP, Incerti D, Linthicum MT. An Open-Source Consensus-Based Approach To Value Assessment [Internet]. 2017. Available from: https://www.healthaffairs.org/action/showDoPubSecure?doi=10.1377%2Fhblog20171212.640960&format=full

71. 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 Nov 1;37(11):1313–20.

72. Kandel S, Heer J, Plaisant C, Kennedy J, van Ham F, Riche NH, et al. Research directions in data wrangling: Visualizations and transformations for usable and credible data.

Information Visualization. 2011 Oct 1;10(4):271–88.

73. Kavanaugh A, Antoni CE, Gladman DD, Wassenberg S, Zhou B, Beutler A, et al. The Infliximab Multinational Psoriatic Arthritis Controlled Trial (IMPACT): results of radiographic analyses after 1 year. Ann Rheum Dis. 2006 Aug 1;65(8):1038–43.

74. Kavanaugh A, Antoni CE, Krueger GG, Yan S, Bala M, Dooley LT, et al. Infliximab improves health related quality of life and physical function in patients with psoriatic arthritis. Ann Rheum Dis. 2006 Apr 1;65(4):471–7.

75. Kavanaugh A, Antoni C, Mease P, Gladman D, Yan S, Bala M, et al. Effect of infliximab therapy on employment, time lost from work, and productivity in patients with psoriatic arthritis. J Rheumatol. 2006 Nov 1;33(11):2254–9.

76. Kavanaugh A, Krueger GG, Beutler A, Guzzo C, Zhou B, Dooley LT, et al. Infliximab maintains a high degree of clinical response in patients with active psoriatic arthritis through 1 year of treatment: results from the IMPACT 2 trial. Ann Rheum Dis. 2007 Apr 1;66(4):498–505.

77. Kavanaugh A, McInnes I, Mease P, Krueger GG, Gladman D, Gomez-Reino J, et al. Golimumab, a new human tumor necrosis factor α antibody, administered every four weeks as a subcutaneous injection in psoriatic arthritis: Twenty-four–week efficacy and safety results of a randomized, placebo-controlled study. Arthritis Rheum. 2009 Apr 1;60(4):976–86.

78. Kavanaugh A, Mease PJ, Gomez-Reino JJ, Adebajo AO, Wollenhaupt J, Gladman DD, et al. Treatment of psoriatic arthritis in a phase 3 randomised, placebo-controlled trial with apremilast, an oral phosphodiesterase 4 inhibitor. Ann Rheum Dis. 2014 Jun 1;73(6):1020–6.

79. Kitzes J, Turek D, Deniz F, editors. The Practice of Reproducible Research [Internet]. 1.2.1. Oakland, CA: University of California Press; 2018. Available from: http://www.practicereproducibleresearch.org/

80. Kruchten P, Nord RL, Ozkaya I. Technical Debt: From Metaphor to Theory and Practice. IEEE Software. 2012 Nov;29(6):18–21.

81. Kunst N, Wilson ECF, Glynn D, Alarid-Escudero F, Baio G, Brennan A, et al. Computing the Expected Value of Sample Information Efficiently: Practical Guidance and Recommendations for Four Model-Based Methods. Value in Health. 2020 Jun 1;23(6):734–42.

82. Langley RG, Elewski BE, Lebwohl M, Reich K, Griffiths CEM, Papp K, et al. Secukinumab in Plaque Psoriasis — Results of Two Phase 3 Trials. N Engl J Med. 2014 Jul 24;371(4):326–38.

83. Lebwohl M, Gottlieb AB, Goffe BS, Jahreis A. Etanercept in psoriatic arthritis: Sustained improvement in skin and joint disease and inhibition of radiographic progression at 2

years. In: Journal of the American Academy of Dermatology [Internet]. New Orleans, LA; 2005. p. AB8. (American Academy of Dermatology Intersociety Liasion Concurrent Session Abstracts; vol. 52). Available from: http://www.sciencedirect.com/science/article/pii/S0190962205003439

84. Lee BD. Ten simple rules for documenting scientific software. PLOS Computational Biology. 2018 Dec 20;14(12):e1006561.

85. Li T, Higgins JP, Deeks JJ, editors. Chapter 5: Collecting Data. In: Cochrane Handbook of Systematic Review of Interventions [Internet]. 6th ed. Cochrane; 2019. Available from: www.training.cochrane.org/handbook

86. Linthicum MT, Bright J, Shafrin J. Open-Source Tools for Value Assessment: A Promising Approach. Journal of Clinical Pathways. 2018 Jun 12;4(5):37–9.

87. Lowndes JSS, Best BD, Scarborough C, Afflerbach JC, Frazier MR, O'Hara CC, et al. Our path to better science in less time using open data science tools. Nature Ecology & Evolution. 2017 Jun;1(6):0160.

88. Luce BR, Drummond M, Jönsson B, Neumann PJ, Schwartz JS, Siebert U, et al. EBM, HTA, and CER: Clearing the Confusion. The Milbank Quarterly. 2010;88(2):256–76.

89. Marwick B. Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. J Archaeol Method Theory. 2017 Jun;24(2):424–50.

90. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. BMC Health Services Research. 2013 Sep 30;13(1):371.

91. Mauskopf J, Samuel M, McBride D, Mallya UG, Feldman SR. Treatment Sequencing After Failure of the First Biologic in Cost-Effectiveness Models of Psoriasis: A Systematic Review of Published Models and Clinical Practice Guidelines. Pharmacoeconomics. 2014;32(4):395–409.

92. McCallum E, Weston S. Parallel R. O'Reilly Media, Inc.; 2011. 123 p.

93. McCallum QE. Bad Data Handbook: Cleaning Up The Data So You Can Get Back To Work. O'Reilly Media, Inc.; 2012. 265 p.

94. McInnes IB, Kavanaugh A, Gottlieb AB, Puig L, Rahman P, Ritchlin C, et al. Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial. The Lancet. 2013 Aug 31;382(9894):780–9.

95. McInnes IB, Mease PJ, Kirkham B, Kavanaugh A, Ritchlin CT, Rahman P, et al. Secukinumab, a human anti-interleukin-17A monoclonal antibody, in patients with psoriatic arthritis (FUTURE 2): a randomised, double-blind, placebo-controlled, phase 3

trial. The Lancet. 2015 Sep 19;386(9999):1137–46.

96. McIntosh E. Using Discrete Choice Experiments within a Cost-Benefit Analysis Framework. Pharmacoeconomics. 2006 Sep 1;24(9):855–68.

97. McIntosh E, Clarke P, Frew EJ, Louviere J, editors. Applied Methods of Cost-Benefit Analysis in Health Care [Internet]. Oxford, UNITED KINGDOM: Oxford University Press USA - OSO; 2010. Available from: http://ebookcentral.proquest.com/lib/ucalgary-ebooks/detail.action?docID=975570

98. McKenna C, Soares M, Claxton K, Bojke L, Griffin S, Palmer S, et al. Unifying Research and Reimbursement Decisions: Case Studies Demonstrating the Sequence of Assessment and Judgments Required. Value in Health. 2015 Sep 1;18(6):865–75.

99. McManus E, Turner D, Gray E, Khawar H, Okoli T, Sach T. Barriers and Facilitators to Model Replication Within Health Economics. Value in Health. 2019 Sep 1;22(9):1018–25.

100. McManus E, Turner D, Sach T. Can You Repeat That? Exploring the Definition of a Successful Model Replication in Health Economics. Pharmacoeconomics. 2019 Sep 18;

101. McNamara A. Key Attributes of a Modern Statistical Computing Tool. The American Statistician. 2019 Oct 2;73(4):375–84.

102. McQueen RB, Padula WV, Campbell JD. A Call for Open-Source Cost-Effectiveness Analysis. Annals of Internal Medicine. 2018 Apr 3;168(7):528.

103. Mease PJ, Ory P, Sharp JT, Ritchlin CT, Bosch FV den, Wellborne F, et al. Adalimumab for long-term treatment of psoriatic arthritis: 2-year data from the Adalimumab Effectiveness in Psoriatic Arthritis Trial (ADEPT). Ann Rheum Dis. 2009 May 1;68(5):702–9.

104. Mease P, Gladman D, Ritchlin C, Mark W. Clinical efficacy, safety and inhibition of joint destruction of adalimumab in the treatment of moderate to severe psoriatic arthritis: 48-week results of the ADEPT trial. In: Journal of the American Academy of Dermatology [Internet]. San Francisco, California: American Academy of Dermatology; 2006. p. AB214–8. (American Academy of Dermatology: Poster Abstracts; vol. 54). Available from: http://www.sciencedirect.com/science/article/pii/S0190962205042519

105. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. The Lancet. 2000 Jul 29;356(9227):385–90.

106. Mease PJ, Fleischmann RM, Deodhar AA, Wollenhaupt J, Khraishi M, Kielar D, et al. Effect of certolizumab pegol on signs and symptoms in patients with psoriatic arthritis: 24-week results of a Phase 3 double-blind randomised placebo-controlled study (RAPID-PsA). Ann Rheum Dis. 2014 Jan 1;73(1):48–55.

107. Mease PJ, Gladman DD, Ritchlin CT, Ruderman EM, Steinfeld SD, Choy EHS, et al. Adalimumab for the treatment of patients with moderately to severely active psoriatic arthritis: Results of a double-blind, randomized, placebo-controlled trial. Arthritis Rheum. 2005 Oct 1;52(10):3279–89.

108. Mease PJ, Heijde D van der, Ritchlin CT, Okada M, Cuchacovich RS, Shuler CL, et al. Ixekizumab, an interleukin-17A specific monoclonal antibody, for the treatment of biologic-naive patients with active psoriatic arthritis: results from the 24-week randomised, double-blind, placebo-controlled and active (adalimumab)-controlled period of the phase III trial SPIRIT-P1. Ann Rheum Dis. 2017 Jan 1;76(1):79–87.

109. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, et al. Etanercept treatment of psoriatic arthritis: Safety, efficacy, and effect on disease progression. Arthritis Rheum. 2004 Jul 1;50(7):2264–72.

110. Mease PJ, Kivitz AJ, Burch FX, Siegel EL, Cohen SB, Ory P, et al. Continued inhibition of radiographic progression in patients with psoriatic arthritis following 2 years of treatment with etanercept. J Rheumatol. 2006 Apr 1;33(4):712–21.

111. Mease PJ, van der Heijde D, Ritchlin CT, Cuchacovich R, Shuler CL, Lee CH, et al. A randomized, double-blind, active-and placebo-controlled phase 3 study of efficacy and safety of ixekizumab, adalimumab, and placebo therapy in patients naive to biologic disease modifying anti-rheumatic drugs with active psoriatic arthritis. In: Arthritis & Rheumatology [Internet]. San Francisco, California; 2015. p. 1. Available from: https://acrabstracts.org/abstract/a-randomized-double-blind-active-and-placebo-controlled-phase-3-study-of-efficacy-and-safety-of-ixekizumab-adalimumab-and-placebo-therapy-in-patients-naive-to-biologic-disease-modifying-anti/

112. Mease P, Sharp J, Ory P, Gladman D, Ritchlin C, Choy E, et al. Adalimumab treatment effects on radiographic progression of joint disease in patients with psoriatic arthritis: results from ADEPT. In: Journal of the European Academy of Dermatology and Venereology [Internet]. London, UK; 2005. p. 1–411. Available from: http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-3083.2005.01311.x

113. Medical Services Advisory Committee. Technical Guidelines for Preparing Assessment Reports for the Medical Services Advisory Committee - Service Type: Investigative [Internet]. Canberra, Australia: Department of Health, Government of Australia; 2017 Jul. Available from: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/MSAC-Guidelines

114. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLOS Computational Biology. 2015 Oct 22;11(10):e1004525.

115. Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. PLOS Medicine. 2015 Sep 29;12(9):e1001882.

116. Murrell P. Introduction to Data Technologies [Internet]. CRC Press; 2009. 445 p. Available from: https://www.stat.auckland.ac.nz/~paul/ItDT/

117. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science [Internet]. Washington, D.C.: National Academies Press; 2019. Available from: https://www.nap.edu/catalog/25303

118. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal [Internet]. London: National Institute for Health and Care Excellence (NICE); 2013. 94 p. (NICE Process and Methods Guides). Available from: http://www.ncbi.nlm.nih.gov/books/NBK395867/

119. National Institute for Health and Care Excellence. Psoriatic Arthritis (Active) – Apremilast (Post DMARDs) [Internet]. National Institute for Health and Care Excellence; 2015 Apr p. 229. (Single Technology Appraisal). Report No.: ID682. Available from: https://www.nice.org.uk/guidance/TA372/documents/psoriatic-arthritis-active-apremilast-post-dmards-id682-committee-papers-2

120. National Institute for Health and Care Excellence. Guide to the Process of Technology Appraisal [Internet]. National Institute for Health and Care Excellence; 2018 May p. 101. Report No.: PMG19. Available from: https://www.nice.org.uk/process/pmg19/chapter/acknowledgements

121. Noble WS. A Quick Guide to Organizing Computational Biology Projects. PLOS Computational Biology. 2009 Jul 31;5(7):e1000424.

122. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open research culture. Science. 2015 Jun 26;348(6242):1422–5.

123. O'Donnell JC, Pham SV, Pashos CL, Miller DW, Smith MD. Health Technology Assessment: Lessons Learned from Around the World—An Overview. Value in Health. 2009 Jun;12(s2):S1–5.

124. O'Connor J, Rice S, Smith A, Rodgers M, Lopez RR, Craig D, et al. The Clinical and Cost Effectiveness of Ustekinumab for the Treatment of Psoriatic Arthritis: A Critique of the Evidence. PharmacoEconomics. 2016 Jan 27;34(4):337–48.

125. OECD. Consumer prices - complete database. 2020 Oct 7;

126. OECD. Purchasing power parities (PPP) and exchange rates. 2020 Oct 7;

127. Peng RD. Reproducible Research in Computational Science. Science. 2011 Dec 2;334(6060):1226–7.

128. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F da V, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLOS Computational Biology. 2016 Jul 14;12(7):e1004947.

129. Philips Z, Claxton K, Palmer S. The Half-Life of Truth: What Are Appropriate Time Horizons for Research Decisions? Med Decis Making. 2008 May 1;28(3):287–99.

130. Prlić A, Procter JB. Ten Simple Rules for the Open Development of Scientific Software. PLOS Computational Biology. 2012 Dec 6;8(12):e1002802.

131. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

132. R Core Team. Writing R Extensions (Version 4.1.1) [Internet]. 2021. Available from: https://cran.r-project.org/doc/manuals/r-release/R-exts.pdf

133. Ritchlin C, Rahman P, Kavanaugh A, McInnes IB, Puig L, Li S, et al. Efficacy and safety of the anti-IL-12/23 p40 monoclonal antibody, ustekinumab, in patients with active psoriatic arthritis despite conventional non-biological and biological anti-tumour necrosis factor therapy: 6-month and 1-year results of the phase 3, multicentre, double-blind, placebo-controlled, randomised PSUMMIT 2 trial. Ann Rheum Dis. 2014 Jun 1;73(6):990–9.

134. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technology Assessment [Internet]. 2011 Feb;15(10). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta15100/

135. Rosenberg David E., Filion Yves, Teasley Rebecca, Sandoval-Solis Samuel, Hecht Jory S., van Zyl Jakobus E., et al. The Next Frontier: Making Research More Reproducible. Journal of Water Resources Planning and Management. 2020 Jun 1;146(6):01820002.

136. Sampson CJ, Arnold R, Bryan S, Clarke P, Ekins S, Hatswell A, et al. Transparency in Decision Modelling: What, Why, Who and How? PharmacoEconomics. 2019 Nov 1;37(11):1355–69.

137. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. PLOS Computational Biology. 2013 Oct 24;9(10):e1003285.

138. Sculpher M, Claxton K. Establishing the Cost-Effectiveness of New Pharmaceuticals under Conditions of Uncertainty—When Is There Sufficient Evidence? Value in Health. 2005 Jul;8(4):433–46.

139. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. J Health Serv Res Policy. 1997 Jan 1;2(1):26–30.

140. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Economics. 2006;15(7):677–87.

141. Shade A, Teal TK. Computing Workflows for Biologists: A Roadmap. PLOS Biology. 2015 Nov 24;13(11):e1002303.

142. Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, Drummond M. Evidence

synthesis, economics and public policy. Research Synthesis Methods. 2010 Apr 1;1(2):126–35.

143. Sholler D, Steinmacher I, Ford D, Averick M, Hoye M, Wilson G. Ten simple rules for helping newcomers become contributors to open projects. PLOS Computational Biology. 2019 Sep 12;15(9):e1007296.

144. Smith-Spangler CM. Transparency and Reproducible Research in Modeling: Why We Need It and How to Get There. Med Decis Making. 2012 Sep 1;32(5):663–6.

145. Stahl JE. Modelling Methods for Pharmacoeconomics and Health Technology Assessment. Pharmacoeconomics. 2008 Feb 1;26(2):131–48.

146. Stodden V, Miguez S. Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research [Internet]. Rochester, NY: Social Science Research Network; 2013 Sep. Report No.: ID 2322276. Available from: https://papers.ssrn.com/abstract=2322276

147. Strong M, Breeze P, Thomas C, Brennan A. SAVI - Sheffield Accelerated Value of Information [Internet]. 2021 [cited 2021 Dec 6]. Available from: https://github.com/Sheffield-Accelerated-VoI/SAVI

148. Strong M, Oakley JE, Brennan A. Estimating Multiparameter Partial Expected Value of Perfect Information from a Probabilistic Sensitivity Analysis Sample: A Nonparametric Regression Approach. Med Decis Making. 2014 Apr 1;34(3):311–26.

149. Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, et al. Data Carpentry: Workshops to Increase Data Literacy for Researchers. International Journal of Digital Curation. 2015 Feb 18;10(1):135–43.

150. Thaçi D, Blauvelt A, Reich K, Tsai T-F, Vanaclocha F, Kingo K, et al. Secukinumab is superior to ustekinumab in clearing skin of subjects with moderate to severe plaque psoriasis: CLEAR, a randomized controlled trial. J Am Acad Dermatol. 2015 Sep 1;73(3):400–9.

151. Thomas J, Noel-Storr A, Marshall I, Wallace B, McDonald S, Mavergames C, et al. Living systematic reviews: 2. Combining human and machine effort. Journal of Clinical Epidemiology. 2017 Nov 1;91:31–7.

152. Tuffaha H. Value of Information Analysis: Are We There Yet? Pharmacoecon Open. 2020 Aug 11;5(2):139–41.

153. Tuyl SV, Whitmire AL. Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. PLOS ONE. 2016 Feb 17;11(2):e0147942.

154. United Kingdom Office for National Statistics. National Life Tables: England and Wales [Internet]. Office of National Statistics; 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexp

ectancies/datasets/nationallifetablesenglandandwalesreferencetables/current

155. van der Heijde D, Kavanaugh A, Gladman DD, Antoni C, Krueger GG, Guzzo C, et al. Infliximab inhibits progression of radiographic damage in patients with active psoriatic arthritis through one year of treatment: Results from the induction and maintenance psoriatic arthritis clinical trial 2. Arthritis Rheum. 2007 Jul 30;56(8):2698–707.

156. Visser MD, McMahon SM, Merow C, Dixon PM, Record S, Jongejans E. Speeding Up Ecological and Evolutionary Computations in R; Essentials of High Performance Computing for Biologists. PLOS Computational Biology. 2015 Mar 26;11(3):e1004140.

157. Wallig M, Microsoft, Weston S. foreach: Provides Foreach Looping Construct [Internet]. 2020. Available from: https://CRAN.R-project.org/package=foreach

158. Weinstein MC, Torrance G, McGuire A. QALYs: The Basics. Value in Health. 2009 Mar;12:S5–9.

159. Welton NJ, Thom HHZ. Value of Information: We've Got Speed, What More Do We Need? Medical Decision Making. 2015 Jul;35(5):564–6.

160. White EP, Baldridge E, Brym ZT, Locey KJ, McGlinn DJ, Supp SR. Nine simple ways to make it easier to (re)use your data. Ideas in Ecology and Evolution [Internet]. 2013 Aug 30;6(2). Available from: https://ojs.library.queensu.ca/index.php/IEE/article/view/4608

161. Wickham H. Tidy Data. Journal of Statistical Software. 2014 Sep 12;59(10):1–23.

162. Wickham H, Bryan J. R Packages [Internet]. 2019. Available from: https://r-pkgs.org/

163. Wilson E, Abrams K. From evidence-based economics to economics-based evidence: using systematic review to inform the design of future research. In: Shemilt I, Donaldson C, Marsh K, Mugford M, Vale L, editors. Event-Based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice. Hoboken, UNITED KINGDOM: John Wiley & Sons, Incorporated; 2010.

164. Wilson ECF. A Practical Guide to Value of Information Analysis. PharmacoEconomics. 2015 Feb;33(2):105–21.

165. Wilson G, Aruliah DA, Brown CT, Hong NPC, Davis M, Guy RT, et al. Best Practices for Scientific Computing. PLOS Biology. 2014 Jan 7;12(1):e1001745.

166. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. PLOS Computational Biology. 2017 Jun 22;13(6):e1005510.

167. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering [Internet]. New York, NY, USA: Association for Computing Machinery; 2014. p. 1–10. (EASE '14). Available from: https://doi.org/10.1145/2601248.2601268

168. Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technology Assessment [Internet]. 2006 Sep;10(31). Available from: https://www.journalslibrary.nihr.ac.uk/hta/hta10310/

169. Wu EQ, Zhou Z-Y, Xie J, Metallo C, Thokala P. Transparency in Health Economic Modeling: Options, Issues and Potential Solutions. PharmacoEconomics. 2019 Nov;37(11):1349–54.

170. Yang H, Epstein D, Bojke L, Craig D, Light K, Bruce I, et al. Golimumab for the treatment of psoriatic arthritis. Health Technol Assess. 2011 Jun 15;15(Suppl. 1):87–96.

171. Yenni GM, Christensen EM, Bledsoe EK, Supp SR, Diaz RM, White EP, et al. Developing a modern data workflow for regularly updated data. PLOS Biology. 2019 Jan 29;17(1):e3000125.

172. Zhu Y, Hernandez LM, Mueller P, Dong Y, Forman MR. Data Acquisition and Preprocessing in Studies on Humans: What is Not Taught in Statistics Classes? The American Statistician. 2013 Nov 1;67(4):235–41.

Appendix A

Analytic Framework for Review of Reproducible Research Practices

General Classification Domains
Recode
Definition: Statements which should be re-coded to a new or unlisted domain. Inclusion Criteria: Include statements which do not satisfy the inclusion criteria for any other domain but are deemed important to be included in the qualitative synthesis. Exclusion Criteria: Exclude statements which satisfy the inclusion criteria for at least one of the alternative domains in the framework.
Definition
Definition: A statement which describes the meaning of a specific term. Inclusion Criteria: Include statements which define a specific term in the context of reproducibility or scientific computing. Exclusion Criteria: Exclude statements which do not offer an explicit definition of a specific term in the context of reproducibility or scientific computing.
Background
Definition: A statement which provides contextual information for reproducible research including the origins of the movements, recent advances, and rationales for its adoption. Inclusion Criteria: Include statements which offer background information on reproducibility, scientific computing, or open science. Exclusion Criteria: Exclude statements which do not offer contextual information for reproducible research, scientific computing or open science.
Domains for Qualitative Synthesis
Open Reporting
Definition: Refers to the availability of information about a project to those who were not involved in developing the original research. This may include the publication of results in journal articles or conference abstracts. It may also include references to the availability of additional artifacts like reports, lab notebooks, and repositories for the project or specific collections of data. Inclusion Criteria: Include statements which provide explicit recommendations regarding strategies to promote reproducibility via open reporting. Exclusion Criteria: Exclude statements which do not provide recommendations to support open reporting.
Copyright
Definition: Refers to rules regarding the protection of intellectual property. Relates to reproducibility as creative work, such as research, is protected by copyright laws which can interfere with the distribution and verification of work. This domain therefore refers to the licensing of research, as such works are protected even if no license is applied. Inclusion Criteria: Include statements which offer guidance or descriptions regarding the impact of specific licensing strategies on reproducibility.

Exclusion Criteria: Exclude statements which do not offer guidance or descriptions of the impact of licensing on reproducibility.
Provenance: Data
Definition: Refers to the original input data used to execute the methods of a project as well as the corresponding output. If some steps cannot be reproduced, it can be helpful to incorporate intermediate data as well. Inclusion Criteria: Include statements which offer guidance or descriptions of data considerations in the reproducibility of a scientific effort. This mainly involves data organization/structure, data management practices, and factors to influence preservation decisions. Exclusion Criteria: Exclude statements which do not offer guidance to support managing the provenance of data to promote the reproducibility of a scientific effort. Further, statements regarding the process of creating, acquiring, or generating a dataset should be recorded in the process domain.
Provenance: Process
Definition: Refers to the specific steps used to execute the methods and procedures of a given project. These steps can be captured in computer files (code or spreadsheets) or in a describable form. Inclusion Criteria: Include statements which offer guidance or descriptions to manage the provenance of the process (or series of steps) used to execute the methods and procedures of a research project. Exclusion Criteria: Exclude statements which do not offer guidance or descriptions to manage the provenance of the process used to execute the methods and procedures of a research project.
Provenance: Environment
Definition: Refers to the assets belonging to the computational environment where the project's methods and procedures were executed. Covers information about the operating system, hardware architecture and software dependencies. Inclusion Criteria: Include statements which offer guidance or strategies relevant to managing the provenance of the computing environment and its impact on reproducibility. Exclusion Criteria: Exclude statements which do not describe practices to manage the provenance of the computing environment.
Documentation
Definition: Refers to descriptions of the project or computational artifact which serves to communicate to another person the detail one needs to know to reproduce or interpret the research. Inclusion Criteria: Include statements which offer guidance regarding the documentation of a project, process, dataset, or environment. This is capturing the describable provenance of a project. Exclusion Criteria: Exclude statements which do not offer specific recommendations regarding how one should leverage documentation to support the reproducibiltiy of a given project.
Project Management
Definition: Refers to the practices or strategies which are applied to organize and manage the various files and processes applied on one's computer. Inclusion Criteria: Include statements which offer strategies to support the organization and management of computer files on one's machine to promote reproducibility. Exclusion Criteria:

Exclude statements which offer no specific guidance to support the management and organization of the various files created as part of a research project.
Tool Selection
Definition: Refers to statements which discuss the factors or considerations in choosing a computing tool (software or hardware) to support the pursuit of reproducible research. Inclusion Criteria: Include statements which offer insight into the types of considerations or arguments which are considered in the selection of computing tools to promote reproducible research. Exclusion Criteria: Exclude statements which do not describe the assessment of requirements to inform tool selection. Additionally, exclude statements which detail how a specific computing tool should be used to support reproducibility.

Appendix B

Network Meta-Analysis Data Availability

The redevelopment of the Psoriatic Arthritis HTA included attempts to generate model parameters from their original sources, including three Network Meta-Analyses (NMA). The first NMA estimated the probability of PsARC response for each treatment. In the economic model, it was used to determine treatment continuation following the first cycle on any biologic. The second NMA estimated the expected change in HAQ-DI conditional on PsARC response. Estimated values from this analysis were used to model changes in arthritis severity. The third NMA estimated the probability of a PASI50, PASI75, and a PASI90 response for each treatment. Generated estimates were supplied to the decision model to capture the changes in psoriasis severity.

As described in the methods of the corresponding paper to this document (Chapter 3), data from the 19 trials identified in the original systematic review were re-compiled from publicly available sources. Data extraction followed the guidance available in the Cochrane Handbook. Information of interest included article meta-data, funding information, trial design, and results. With respect to the results, data were collected for each outcome, event (week), sub-population, and arm within each trial. To support reproducibility, the extracted data were organized using the rules for creating "tidy" or "normalized" data. This approach ensured that the collection of tables which was created could serve as a single source where the most up-to-date information about the included trials could be accessed.

Inclusion of each NMA in the redevelopment was dependent on the successful replication of the methods as originally implemented in the reference project. While the code used to

estimate each NMA was included in the appendices of Corbett et al., the decision to implement was dependent on the availability of each dataset compiled from the systematic review.

The objective of this appendix is to summarize the data availability for the evidence synthesis of all three outcomes. The re-compiled trial data were queried to return the available values needed to implement each meta-analysis. This information was supplemented with a visualization of data availability by week of each trial. Each section concludes with a description of how the NMA values were incorporated into the decision model.

Psoriatic Arthritis Response Criteria

Figure B-1 illustrates the availability of PsARC response data from Weeks 10 to 24 in each trial stratified by sub-population. At the time of this writing, many values included in the 2017 MTA but were redacted had yet to be publicly released or identified. The available PsARC response data for Network Meta-Analysis are reported in Table B-1. Incomplete reporting prevented the implementation of the relevant Network Meta-Analyses as described in the original HTA. While it was possible to consider implementing the NMAs using the available data, this was not pursued as it would have limited the ability to implement the same comparators in the decision models. To incorporate each PsARC NMA in the decision model, the original model results for the Naïve and Experienced sub-populations were transcribed and treated as the data source for this parameter. These values are reproduced in Table B-2. In the event the unpublished data are made available in the future, then it may be possible to implement the original NMA code.

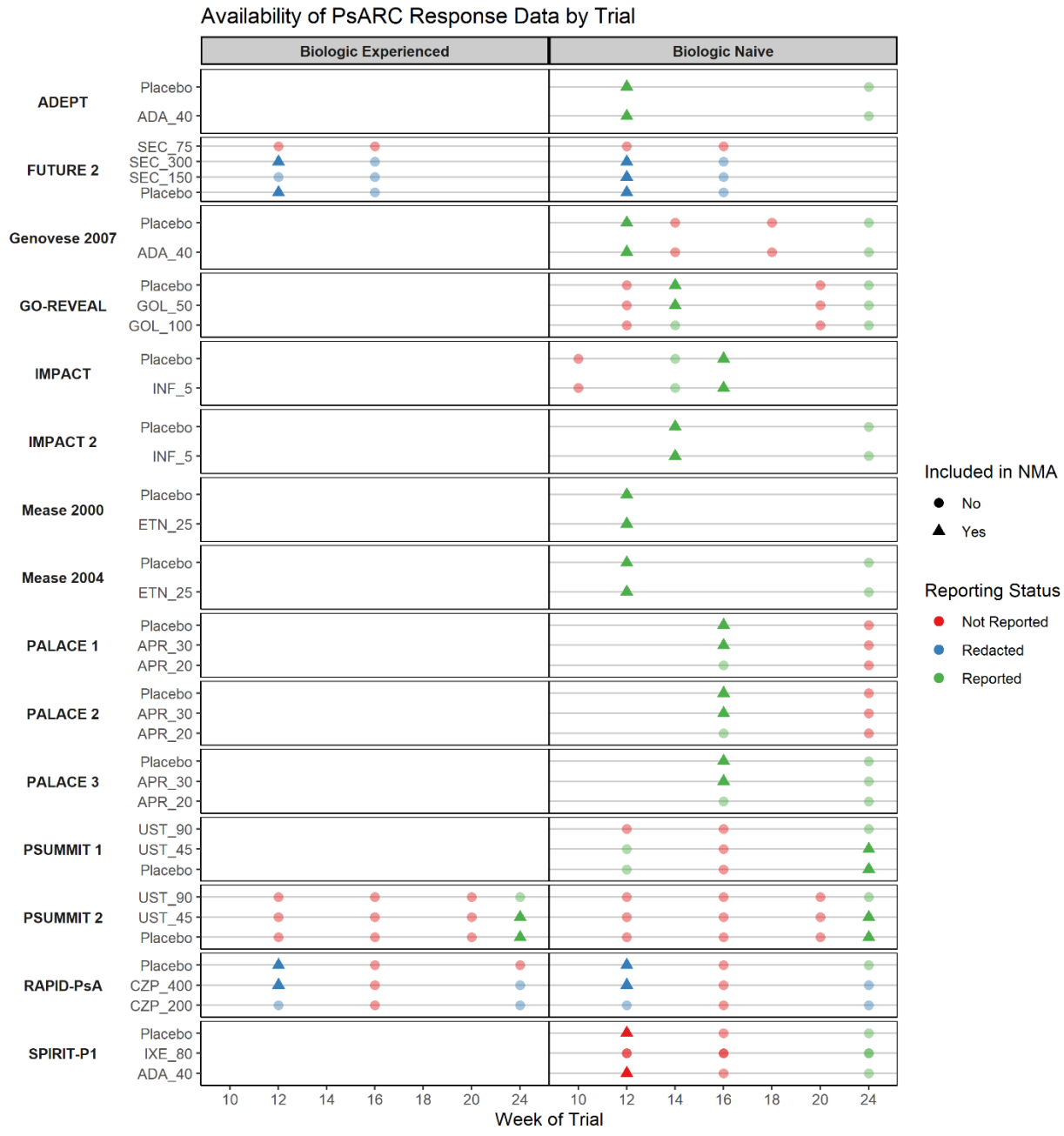


Figure B-1. Data Availability for PsARC Response

Table B-1. Inputs to Network Meta-Analysis of PsARC Response

		Biologic Naive			Biologic Experienced		
	Arm (Dose)	**Week**	**n[1]**	**R[2]**	**Week**	**n[1]**	**R[2]**
ADEPT	ADA (40)	12	151	94	–	–	–
	Placebo	12	162	42	–	–	–
FUTURE 2	SEC (150)	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
	SEC (300)	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
	Placebo	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
Genovese 2007	ADA (40)	12	51	26	–	–	–
	Placebo	12	49	12	–	–	–
GO-REVEAL	GOL (50)	14	146	107	–	–	–
	GOL (150)	14	146	105	–	–	–
	Placebo	14	113	24	–	–	–
IMPACT	INF (5)	16	52	39	–	–	–
	Placebo	16	52	11	–	–	–
IMPACT 2	INF (5)	14	100	77	–	–	–
	Placebo	14	100	27	–	–	–
Mease 2000	ETN (25)	12	30	26	–	–	–
	Placebo	12	30	7	–	–	–
Mease 2004	ETN (25)	12	101	73	–	–	–
	Placebo	12	104	32	–	–	–
PALACE 1	APR (20)	16	168	65	–	–	–
	APR (30)	16	168	78	–	–	–
	Placebo	16	168	50	–	–	–
PALACE 2	APR (20)	16	163	78	–	–	–
	APR (30)	16	162	78	–	–	–
	Placebo	16	159	53	–	–	–
PALACE 3	APR (20)	16	169	64	–	–	–
	APR (30)	16	167	88	–	–	–
	Placebo	16	169	46	–	–	–
PSUMMIT 1	UST (45)	24	205	115	–	–	–
	UST (90)	24	204	132	–	–	–
	Placebo	24	205	77	–	–	–
PSUMMIT 2	UST (45)	24	43	24	24	60	33
	UST (90)	24	47	27	24	58	27
	Placebo	24	42	16	24	62	16
RAPID-PsA	CZP (200)	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
	CZP (400)	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
	Placebo	12	Redacted[3]	Redacted[3]	12	Redacted[3]	Redacted[3]
SPIRIT-P1	ADA (40)	12	–	–	–	–	–
	Placebo	12	–	–	–	–	–

1. Number Randomized
2. Number of PsARC Responders
3. Redacted values remain unpublished. Trial registry data updated in June 2020 with week 24 values.

Table B-2. Network Meta-Analysis Results - PsARC Response

Placebo Response Models[1]			
Sub-Population	**Mean**	**Median**	**95% CrI**
Naïve[3]	-0.81	-0.81	-1.02 – -0.61
Experienced[4]	-1.00	-1.01	-1.48 – -0.58
Treatment Effects Models[2]			
Tx	**Mean**	**Median**	**95% CrI**
Model A1, Biologic Naïve[5]			
SEC300	1.18	1.18	0.44 – 1.93
SEC150	1.18	1.18	0.43 – 1.94
UST	0.76	0.76	0.40 – 1.12
CZP	1.10	1.09	0.63 – 1.57
GOL	2.34	2.34	1.77 – 2.94
ADA	1.40	1.40	0.99 – 1.83
INF	2.30	2.30	1.78 – 2.84
ETN	2.05	2.04	1.51 – 2.61
APR	0.81	0.81	0.55 – 1.08
Model A1, Biologic Experienced[6]			
SEC300	1.73	1.80	0.77 – 2.91
UST	1.25	1.28	0.53 – 2.07
Model D2, Biologic Naïve[7]			
SEC300	1.84	1.83	1.15 – 2.59
SEC150	1.83	1.82	1.13 – 1.59
UST	1.18	1.17	0.81 – 1.58
CZP	1.72	1.72	1.28 – 2.21
GOL	1.71	1.71	1.17 – 2.20
ADA	1.20	1.20	0.83 – 1.55
INF	1.87	1.88	1.43 – 2.31
ETN	1.87	1.87	1.48 – 2.29
APR	0.77	0.77	0.57 – 0.97

Notes:

1. Values reported on log-odds scale.
2. Values reported as log-odds ratios.
3. Data acquired from Table 126 in Corbett et al. (24).
4. Data acquired from Table 45 in Corbett et al. (24).
5. Data acquired from Table 127 in Corbett et al. (24).
6. Data acquired from Table 45 in Corbett et al. (24).
7. Data acquired from Table 133 in Corbett et al. (24).

HAQ-DI Change Conditional on PsARC Response

Figure B-2 and Figure B-3 illustrate the availability of trial data relevant to the HAQ NMA from Weeks 10 to 24. The data are stratified by PsARC response status. Values included in the 2017 MTA, but redacted in the published report, remain unpublished at the time of this writing (FUTURE 2, RAPID-PsA). Furthermore, the provenance of many values included in the 2017 MTA could not be established. For example, while the 2017 MTA reported the combined results from the PSUMMIT 1 and 2 trials at week 24 – the individual results from each trial was not reported. The aim of the figures is to view the reporting inventory as represented in the underlying dataset. The available data which could have been used in an NMA is reported in Table B-3 and Table B-4. As with PsARC, implementing the NMAs with the available data was not pursued as it would have constrained the evaluation of comparators in the decision model. Instead, the original NMA results for the Naïve and Experienced sub-populations, which are presented in Table B-5, were treated as the data source for this parameter set.

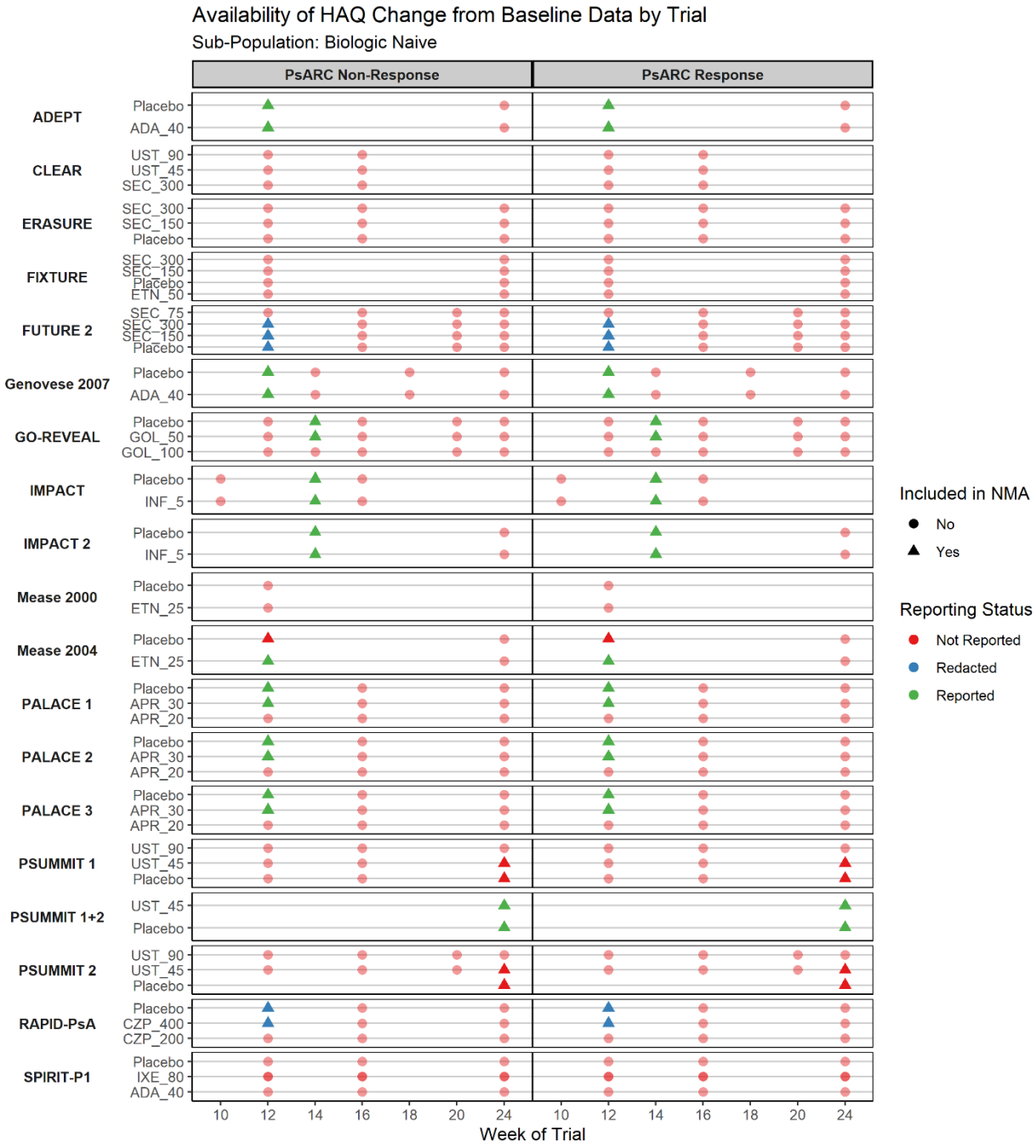


Figure B-2. Data Availability for ΔHAQ for Biologic Naive Sub-Population

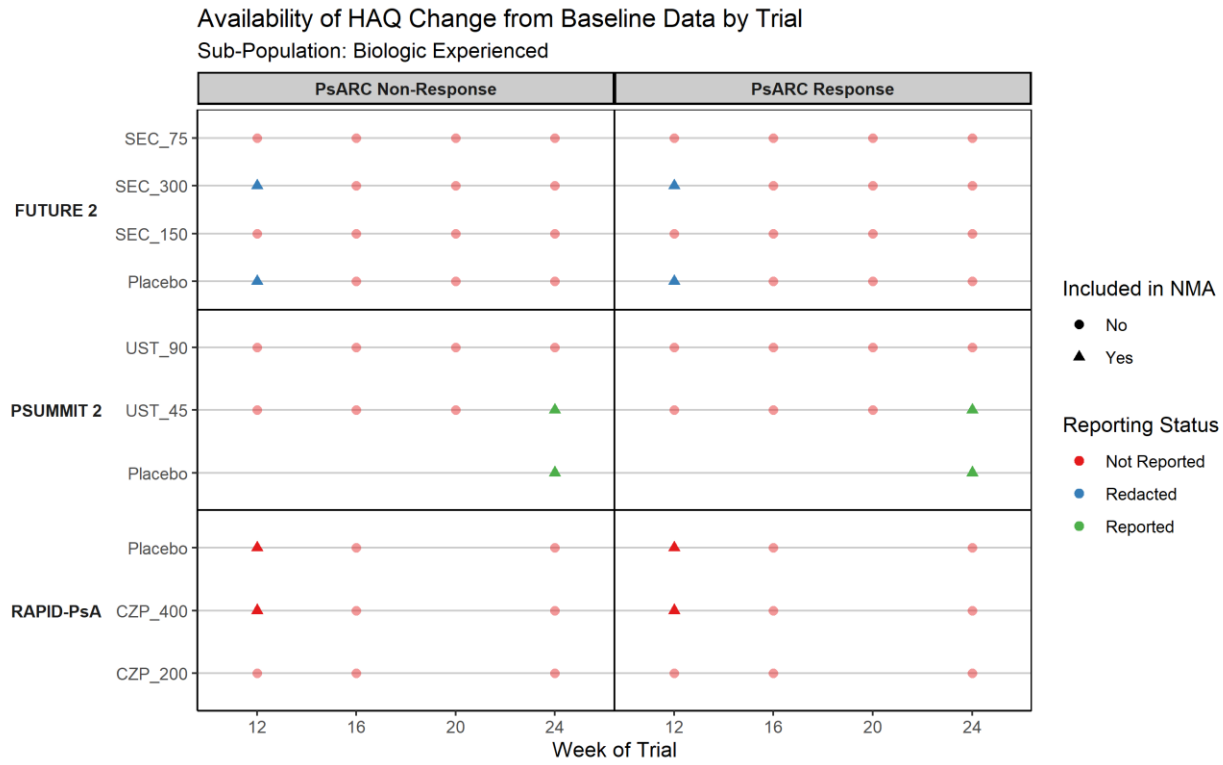


Figure B-3. Data Availability for ΔHAQ for the Biologic Experienced Sub-Population

Table B-3. Inputs to Network Meta-Analysis of ΔHAQ Conditional on PsARC Response (Biologic Naive)

			PsARC Response		PsARC Non-Response	
Trial	**Arm (Dose)**	**Week**	**Mean**	**SE**	**Mean**	**SE**
ADEPT	ADA (40)	12	-0.500	0.050	-0.120	0.050
	Placebo	12	-0.313	0.080	0.026	0.040
FUTURE 2	SEC (150)	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
	SEC (300)	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
	Placebo	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
Genovese 2007	ADA (4)	12	-0.423	0.080	-0.150	0.090
	Placebo	12	-0.177	0.060	-0.057	0.050
GO-REVEAL	GOL (50)	14	-0.424	0.070	-0.049	0.060
	Placebo	14	-0.286	0.050	0.023	0.020
IMPACT	INF (5)	14	-0.650[2]	0.090[2]	-0.200[2]	0.090[2]
	Placebo	14	-0.270[2]	0.140[2]	0.020[2]	0.050[2]
IMPACT 2	INF (5)	14	-0.580	0.060	-0.110	0.060
	Placebo	14	-0.160	0.100	0.070	0.040
Mease 2004	ETN (25)	12	-0.635	0.060	-0.196	0.070
	Placebo	12	-0.258	0.010	-0.002	0.040
PALACE 1	APR (20)	12	–	–	–	–
	APR (30)	12	-0.460[2]	0.050[2]	-0.070[2]	0.050[2]
	Placebo	12	-0.320[2]	0.070[2]	0.000[2]	0.040[2]
PALACE 2	APR (30)	12	-0.330[2]	0.060[2]	-0.120[2]	0.050[2]
	Placebo	12	-0.220[2]	0.070[2]	0.010[2]	0.040[2]
PALACE 3	APR (30)	12	-0.290[2]	0.050[2]	-0.080[2]	0.050[2]
	Placebo	12	-0.250[2]	0.060[2]	0.000[2]	0.030[2]
PSUMMIT 1	UST (45)	24	_[3]	_[3]	_[3]	_[3]
	UST (90)	24	_[3]	_[3]	_[3]	_[3]
	Placebo	24	_[3]	_[3]	_[3]	_[3]
PSUMMIT 2	UST (45)	24	_[3]	_[3]	_[3]	_[3]
	UST (90)	24	_[3]	_[3]	_[3]	_[3]
	Placebo	24	_[3]	_[3]	_[3]	_[3]
PSUMMIT 1+2	UST (45)	24	-0.487	0.050	-0.097	0.050
	Placebo	24	-0.260	0.040	-0.001	0.030
RAPID-PsA	CZP (200)	12	–	–	–	–
	CZP (400)	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
	Placebo	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]

1. Redacted values remain unpublished. Trial registry data updated in June 2020 with week 24 values.
2. Values reported in Corbett et al. (2017). Original source unclear.
3. Data from individual trials not available.

Table B-4. Inputs to Network Meta-Analysis of ΔHAQ Conditional on PsARC Response (Biologic Experienced)

			PsARC Response		PsARC Non-Response	
Trial	**Arm (Dose)**	**Week**	**Mean**	**SE**	**Mean**	**SE**
FUTURE 2	SEC (150)	12	–	–	–	–
	SEC (300)	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
	Placebo	12	Redacted[1]	Redacted[1]	Redacted[1]	Redacted[1]
PSUMMIT 2	UST (45)	24	-0.315	0.110	0.007	0.130
	UST (90)	24	–	–	–	–
	Placebo	24	-0.146	0.090	0.010	0.050
RAPID-PsA	CZP (200)	12	–	–	–	–
	CZP (400)	12	–	–	–	–
	Placebo	12	–	–	–	–

1. Redacted values remain unpublished. Trial registry data updated in June 2020 with week 24 values.

Table B-5. Network Meta-Analysis Results for ΔHAQ Conditional on PsARC Response

	PsARC Response			*PsARC Non-Response*		
Tx	**Mean**	**Median**	**95% CrI**	**Mean**	**Median**	**95% CrI**
Model E1, Biologic Naive[1]						
Placebo	-0.26	-0.26	-0.30 – -0.22	---	---	---
SEC150	-0.39	-0.40	-0.55 – -0.24	-0.08	-0.08	-0.39 – 0.22
SEC300	-0.55	-0.55	-0.72 – -0.37	-0.05	-0.05	-0.29 – 0.18
UST	-0.49	-0.49	-0.60 – -0.38	-0.10	-0.10	-0.21 – 0.01
CZP	-0.43	-0.43	-0.53 – -0.33	-0.07	-0.07	-0.19 – 0.06
GOL	-0.44	-0.44	-0.59 – -0.29	-0.06	-0.06	-0.18 – 0.06
ADA	-0.49	-0.49	-0.58 – -0.40	-0.14	-0.13	-0.24 – -0.03
INF	-0.66	-0.66	-0.77 – -0.55	-0.20	-0.20	-0.31 – -0.08
ETN	-0.64	-0.64	-0.77 – -0.52	-0.20	-0.20	-0.35 – -0.05
APR	-0.36	-0.36	-0.43 – -0.29	-0.09	-0.09	-0.16 – -0.02
Model E1, Biologic Experienced[2]						
Placebo	-0.13	-0.13	-0.29 – 0.02	---	---	---
SEC300	-0.39	-0.39	-0.62 – -0.15	-0.43	-0.43	-0.88 – 0.01
UST	-0.32	-0.32	-0.55 – -0.09	0.00	0.00	-0.27 – 0.27
Model E2, Biologic Naïve[3]						
Placebo	-0.26	-0.26	-0.30 – -0.22	---	---	---
SEC150	-0.43	-0.44	-0.56 – -0.29	-0.09	-0.09	-0.23 – 0.06
SEC300	-0.51	-0.51	-0.66 – -0.38	-0.08	-0.08	-0.21 – 0.06
UST	-0.48	-0.48	-0.58 – -0.38	-0.09	-0.09	-0.19 – 0.01
CZP	-0.47	-0.47	-0.56 – -0.37	-0.12	-0.12	-0.20 – -0.02
GOL	-0.48	-0.49	-0.59 – -0.35	-0.11	-0.11	-0.19 – -0.01
ADA	-0.50	-0.50	-0.58 – -0.41	-0.13	-0.13	-0.21 – -0.06
INF	-0.61	-0.60	-0.72 – -0.50	-0.15	-0.14	-0.24 – -0.07
ETN	-0.59	-0.59	-0.72 – -0.49	-0.15	-0.14	-0.26 – -0.06
APR	-0.36	-0.36	-0.43 – -0.29	-0.09	-0.09	-0.16 – -0.02

Notes:

1. Data acquired from Table 147 in Corbett et al.
2. Data acquired from Table 49 in Corbett et al.
3. Data acquired from Table 148 in Corbett et al.

PASI50/PASI75/PASI90 Response

The reporting from PsARC and HAQ was repeated for PASI, as illustrated in Figure B-4 and Figure B-5. While there was interest in capturing PASI50, PASI75, and PASI90 results – these thresholds were not included in all trials. As before, the many redacted values and unclear provenance of some required values served as barriers to including the relevant NMAs within the reproducible workflow. The available data for the NMA of PASI response is reported in Table B-6 and Table B-7. As with the other two outcomes, implementing the PASI NMAs using the available data was not considered as it would have constrained the evaluation of comparators in the decision model. Instead, the original NMA results for both sub-populations, which are presented in Table B-8, were treated as the data source for this parameter set.

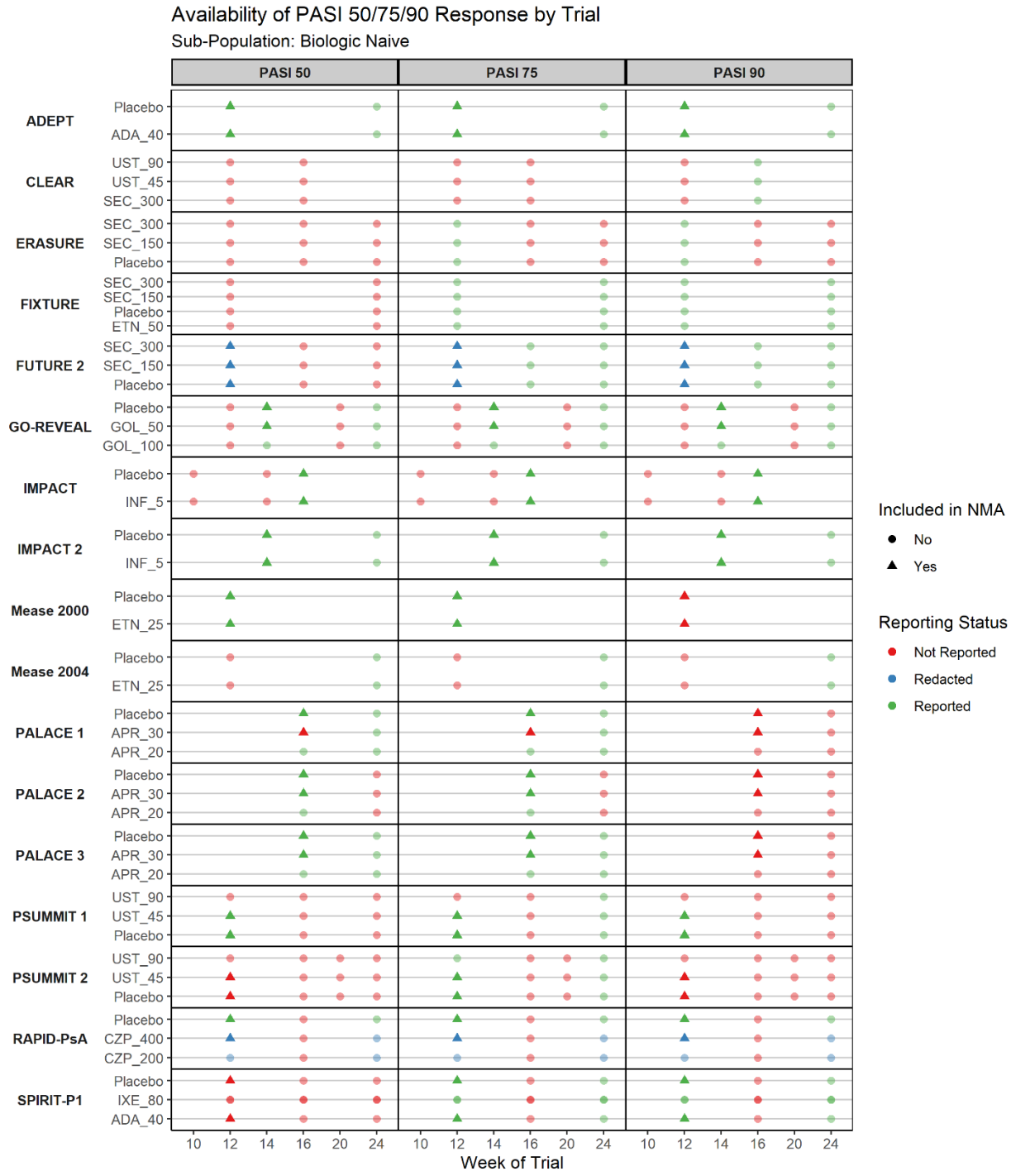


Figure B-4. PASI Response Availability for the Biologic Naive Sub-Population

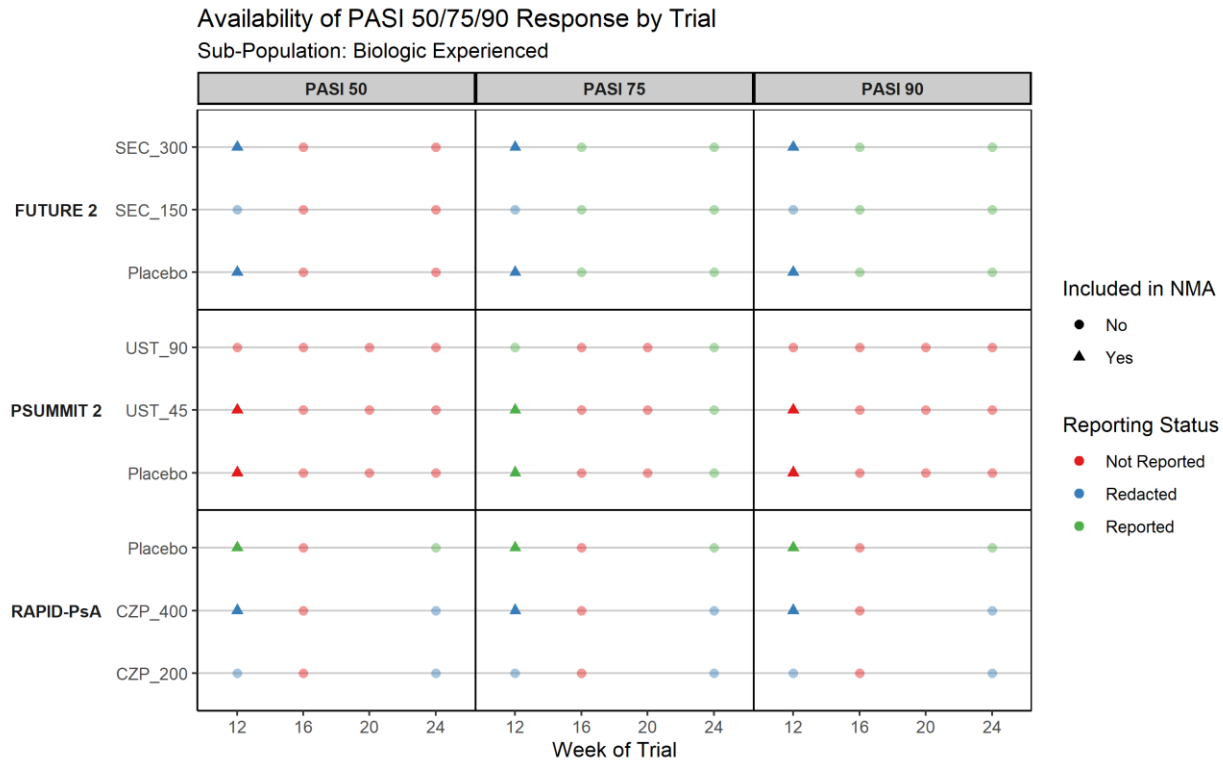


Figure B-5. PASI Response Availability for the Biologic Experienced Sub-Population

Table B-6. Inputs to Network Meta-Analysis of PASI Response (Biologic Naive)

Trial	Arm (Dose)	Week	N[1]	PASI50[2]	PASI75[2]	PASI90[2]
ADEPT	ADA (40)	12	69	50	34	21
	Placebo	12	69	10	3	0
FUTURE 2	SEC (150)	12	36	Redacted[3]	Redacted[3]	Redacted[3]
	SEC (300)	12	30	Redacted[3]	Redacted[3]	Redacted[3]
	Placebo	12	31	Redacted[3]	Redacted[3]	Redacted[3]
GO-REVEAL	GOL (50)	14	109	63	44	22
	GOL (100)	14	108	83	63	26
	Placebo	14	79	7	2	0
IMPACT	INF (5)	16	22	22	15	8
	Placebo	16	16	0	0	0
IMPACT 2	INF (5)	14	83	68	53	34
	Placebo	14	87	8	2	0
Mease 2000	ETN (25)	12	19	8	5	–[4]
	Placebo	12	19	4	0	–[4]
PALACE 1	APR (20)	16	82	36	18	–[4]
	APR (30)	16	74	–[4]	–[4]	–[4]
	Placebo	16	68	11	3	–[4]
PALACE 2	APR (20)	16	80	27	15	–[4]
	APR (30)	16	77	32	17	–[4]
	Placebo	16	74	10	2	–[4]
PALACE 3	APR (20)	16	91	31	19	–[4]
	APR (30)	16	90	38	20	–[4]
	Placebo	16	89	22	7	–[4]
PSUMMIT 1	UST (45)	24	145	89	56	28
	UST (90)	24	–	–[4]	–[4]	–[4]
	Placebo	24	146	31	13	6
PSUMMIT 2	UST (45)	24	36	–[4]	17	–[4]
	UST (90)	24	36	–[4]	17	–[4]
	Placebo	24	30	–[4]	1	–[4]
RAPID-PsA	CZP (200)	12	–	Redacted[3]	Redacted[3]	Redacted[3]
	CZP (400)	12	–	Redacted[3]	Redacted[3]	Redacted[3]
	Placebo	12	66	18	11	3
SPIRIT-P1	ADA (40)	12	68	–[4]	23	15
	Placebo	12	67	–[4]	5	1

1. Number randomized.
2. Number of responders at the specified threshold level.
3. Redacted values remain unpublished. Trial registry data updated in June 2020 with week 24 values.
4. Response at threshold level not reported.

Table B-7. Inputs to Network Meta-Analysis of PASI Response (Biologic Experienced)

Trial	Arm (Dose)	Week	N[1]	PASI50[2]	PASI75[2]	PASI90[2]
FUTURE 2	SEC (150)	12	22	Redacted[3]	Redacted[3]	Redacted[3]
	SEC (300)	12	11	Redacted[3]	Redacted[3]	Redacted[3]
	Placebo	12	12	Redacted[3]	Redacted[3]	Redacted[3]
PSUMMIT 2	UST (45)	24	44	–[4]	14	–[4]
	UST (90)	24	44	–[4]	14	–[4]
	Placebo	24	50	–[4]	1	–[4]
RAPID-PsA	CZP (200)	12	–	Redacted[3]	Redacted[3]	Redacted[3]
	CZP (400)	12	–	Redacted[3]	Redacted[3]	Redacted[3]
	Placebo	12	20	5	1	1

1. Number randomized.
2. Number of responders at the specified threshold level.
3. Redacted values remain unpublished. Trial registry data updated in June 2020 with week 24 values.
4. Response at threshold level not reported.

Table B-8. Network Meta-Analysis Results for PASI Response

Tx	Mean	Median	95% CrI
Model F1, Biologic Naive[1]			
Placebo	1.02	1.02	0.90 – 1.15
SEC300	-1.94	-1.94	-2.63 – -1.28
SEC150	-1.88	-1.87	-2.54 – -1.24
UST	-1.14	-1.13	-1.41 – -0.87
CZP	-0.88	-0.88	-1.24 – -0.52
GOL	-1.65	-1.64	-2.10 – -1.21
ADA	-1.48	-1.48	-1.83 – -1.14
INF	-2.41	-2.41	-2.84 – -2.01
ETN	-0.80	-0.80	-1.64 – 0.03
APR	-0.75	-0.75	-0.99 – -0.51
PASI Threshold (z_j)			
zPASI50	0.00	0.00	0.00 – 0.00
zPASI75	0.59	0.59	0.52 – 0.65
zPASI90	1.15	1.15	1.06 – 1.25
Model F1, Biologic Experienced[2]			
Placebo	1.35	1.35	0.59 – 2.19
SEC300	-2.51	-2.51	-4.01 – -1.23
UST	-1.66	-1.66	-2.73 – -0.83
PASI Threshold (z_j)			
zPASI50	0.00	0.00	0.00 – 0.00
zPASI75	0.87	0.87	0.28 – 1.84
zPASI90	1.48	1.48	0.70 – 2.56
Model G2, Biologic Naïve[3]			
Placebo	1.02	1.01	0.89 – 1.15
SEC300	-1.86	-1.86	-2.33 – -1.36
SEC150	-1.79	-1.80	-2.23 – -1.32
UST	-1.35	-1.34	-1.60 – -1.12
CZP	-1.43	-1.42	-1.89 – -1.04
GOL	-1.13	-1.14	-1.50 – -0.67
ADA	-1.42	-1.42	-1.67 – -1.17
INF	-1.79	-1.80	-2.17 – -1.31
ETN	-0.85	-0.85	-1.48 – -0.20
APR	-0.82	-0.82	-1.00 – -0.64
PASI Threshold (z_j)			
zPASI50	0.00	0.00	0.00 – 0.00
zPASI75	0.58	0.58	0.52 – 0.65
zPASI90	1.14	1.14	1.04 – 1.24

Notes:

1. Data acquired from Table 155 in Corbett et al.
2. Data acquired from Table 55 in Corbett et al.
3. Data acquired from Table 157 in Corbett et al.

Appendix C

Constructing Treatment Sequences for Uncertainty Demonstration

The expanded set of comparator sequences and the source treatments are presented in Table C-1. As the objective of the corresponding simulation was to demonstrate specific programming features, the constructed sequences were not intended to reflect clinical reality. Consistent with Corbett et al., candidate treatments had to be approved for use in adult Psoriatic Arthritis (PsA) at the time of the original appraisal. This led to the exclusion of APR and restriction of SEC dosing by sub-group. Treatment eligibility was further constrained by the available evidence synthesis data for three outcomes (PsARC, HAQ, PASI) based on biologic experience. In first- and second- line therapy, patients were assumed to be biologic naïve. Those who transitioned to a third-line biologic were assumed to be experienced. To simplify the comparators, it was assumed that a treatment could only be trialled once, and subsequent lines of treatment had to belong to a different class from the preceding biologic (anti-TNF vs anti-IL).

Table C-1. Comparator Treatment Sequences for Uncertainty Demonstration

Eligible Treatments			
Class	**Biologics**		
Positions 1 and 2 (Biologic Naïve)			
Anti-TNF	ETN, INF, ADA, GOL, CZP		
Anti-IL	UST, SEC150, SEC300		
Position 3 (Biologic Experienced)			
Anti-TNF	--[1]		
Anti-IL	UST, SEC150, SEC300		
Included Comparators[2]			
SEC150_ETN_UST[3]	SEC300_ETN_UST[3]	SEC150_INF_UST	SEC300_INF_UST
SEC150 ADA UST	SEC300 ADA UST	SEC150 GOL UST	SEC300 GOL UST
SEC150 CZP UST	SEC300 CZP UST	UST ETN SEC300	SEC150 ETN SEC300
UST INF SEC300	SEC150 INF SEC300	UST ADA SEC300	SEC150 ADA SEC300
UST GOL SEC300	SEC150 GOL SEC300	UST CZP SEC300	SEC150 CZP SEC300

Notes:

1. Biologic Experienced data not available for treatments in the Anti-TNF class.
2. Not shown: BSC was included as a comparator.
3. Original treatment sequence. The third original sequence, CZP-ETN-UST, was excluded because CZP/ETN are both Anti-TNF.

Appendix D

Availability of Code and Data

The implementation of strategies to enhance the reproducibility of Health Technology Assessments for this thesis is reflected through the development of the `HTATools4PsA` R package. This R package represents a collection of functions and data sets which were used for the economic evaluation of biologic treatment for patients with Psoriatic Arthritis. The version of the package used for this thesis has been preserved in a Zenodo repository, as per the below citation.

> Wagner, DJ. (2021). HTATools4PsA: Tools for the Economic Evaluation of Biologic Treatment for Psoriatic Arthritis (Version 0.8.3). https://doi.org/10.5281/zenodo.5784367

Installation of the package will follow a simple process:

- Make sure the following R packages are already installed: `devtools`, `fs`, `tidyverse`, and `RefManageR`.
- Download the package ZIP file from Zenodo using the above supplied Digital Object Identifier.
- Install the package using the `install_local()` function from the devtools R package.
    - This function will require the path for the downloaded ZIP file.
    - Once the installation is successful, the original ZIP file can be deleted.