# Advancing the methods and accessibility of cost-effectiveness and value of information analyses in health care

by

Mohsen Sadatsafavi

## A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

#### DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Pharmaceutical Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2012

© Mohsen Sadatsafavi, 2012

#### Abstract

This thesis comprises three methodological advancements that address important issues related to cost-effectiveness analysis (CEA) and expected value of information (EVI) analysis in health technology assessment. Aims: 1) To develop a practical sampling scheme for the incorporation of external evidence in CEAs conducted alongside randomized controlled trials (RCT); 2) To develop non-parametric methods for the calculation of the expected value of sample information (EVSI) for RCT-based CEAs; 3) To develop a computationally efficient algorithm for the calculation of single-parameter expected value of partial perfect information (EVPPI) for RCT-based and model-based CEAs. The theories and methods laid out in this work are accompanied by real-world CEA and EVI analyses of the Canadian Optimal Therapy of Chronic Obstructive Pulmonary Diseases (OPTIMAL) trial, a RCT on combination pharmaceutical therapies in chronic obstructive pulmonary diseases (COPD). Results: 1) The 'vetted bootstrap' is a semi-parametric algorithm based on rejection sampling and bootstrapping that allows the incorporation of external evidence into RCT-based CEAs. Implementing this method to incorporate external information on the effect size of treatment in the OPTIMAL trial required only minor modifications to the original CEA algorithm. 2) A Bayesian interpretation of the bootstrap allows non-parametric calculation of EVSI through two-level resampling. In the case study, incorporation of missing value imputation and adjustment for covariate imbalance in EVI calculations generated EVSI and the expected value of perfect information (EVPI) values that were significantly different than those calculated conventionally, demonstrating the flexibility of this method and the potential impact of modeling such aspects of the analysis on EVI

calculations. 3) The new method enabled the calculation of EVPPI for the effect size of treatment for the exemplary RCT data, and showed a significant (up to 25 times in terms of root-mean-squared error) improvement in efficiency compared to the conventional EVPPI calculation methods in a series of simulations. Summary: This thesis provides several original advancements in the methodology of the CEA and EVI analysis of RCTs and enables several analytical approaches that have hitherto been available only through parametric modeling of RCT data.

#### Preface

This thesis is the culmination of research initiated and developed by Mohsen Sadatsafavi. The author defined the research questions explored in this thesis, conducted the statistical analyses required and wrote the four thesis-based manuscripts. The co-authors, some of whom were members of the primary author's thesis committee, provided methodological and clinical advice throughout the execution of the thesis work.

The Optimal Therapy of COPD trial (OPTIMAL) was a Canadian multi-centre randomized, controlled trial. Mohsen Sadatsafavi was part of the team that conducted the costeffectiveness analysis for this study (Najafzadeh M, Marra C, Sadatsafavi M, Aaron S, Sullivan S, Vandemheen KL, et al. Cost effectiveness of therapy with combinations of long acting bronchodilators and inhaled steroids for treatment of COPD. Thorax. 2008;63(11):962-967). The data for this study was used as a running example through this thesis

Full permission to use and re-analyze the data from this study was granted by the principal investigator of the study, Dr. Shawn Aaron, University of Ottawa.

Manuscripts related to this thesis:

1) Mohsen Sadatsafavi, Carlo Marra, Lawrence McCandless, Stirling Bryan. The vetted bootstrap: a practical sampling algorithm to incorporate external evidence in trial-based CEAs (under preparation) [related to Chapter 2] 2) Mohsen Sadatsafavi, Stirling Bryan, Larry Lynd, Carlo Marra. Two-level resampling as a novel method for calculation of expected value of sample information in economic trials (under preparation)

3) Mohsen Sadatsafavi, Mehdi Najafzadeh, Nick Bansback, Lawrence McCandless, Carlo Marra. Efficient algorithm for calculation of single-parameter expected value of partial perfect information (under preparation) [related to Chapter 4]

## **Table of Contents**

Abstract		ii
Preface		. iv
List of Ta	bles	viii
List of Fig	gures	. ix
Acknowl	edgements	x
Dedicatio	on	. xi
Chapter	1. Introduction	1
1.1	Research statement	1
1.2	Introduction	3
1.2.3	1 Types of economic evaluations	5
1.2.2	2 Outputs of an economic evaluation	8
1.2.3	3 Synthesizing information (evidence synthesis)	10
1.2.4	4 Decision making under uncertainty	16
1.2.	5 The bootstrap method for the analysis of uncertainty in RCT-based CEAs	18
1.2.0	6 Expected Value of Information (EVI)	20
1.3	Gaps in current knowledge that inspired this dissertation	28
1.4	Context	29
1.5	The OPTIMAL clinical trial as a case study	29
1.6	Contents of this thesis	32
Chapter	2. The Vetted Bootstrap, a practical algorithm for the incorporation of external	
evidence	e in RCT-based CEAs	34
2.1	Background	34
2.2	The vetted bootstrap: a practical approach to evidence synthesis in RCT-based CEAs	35
2.3	Theory and methods	36
2.4	A Bayesian interpretation of the bootstrap	37
2.5	Cost-effectiveness analysis without the incorporation of external evidence	39
2.6	Incorporating external evidence	40
2.7	Algorithm: the vetted bootstrap	41
2.8	Weighted bootstrap as an alternative to the vetted bootstrap	42
2.9	Case studies	43
2.9.3	1 Analogy between the vetted bootstrap and parametric analysis of bivariate	
norr	nal data	43
2.9.2	2 A real-world RCT-based CEA	45
2.10	Discussion	55
Chapter	3. Pulling EVI by its bootstrap: a framework for non-parametric expected value of	
informat	ion analysis of RCT-based CEAs	60
3.1	Background	60
3.2	Context	61
3.3	EVPI and EVSI definitions in a non-parametric framework	62
3.4	Bootstrap as a method of sampling from the "distribution of a distribution"	63
3.5	Case studies	64
3.5.3	1 A stylized example	64

352	Example from the OPTIMAL trial	68		
3.6 Dis		80		
Chapter 4.	A heuristic algorithm for calculation of single-parameter expected value	of partial		
nerfect Info	rmation	85		
4 1 Ba	skaround			
4.1 Du	stands			
421	Context and notations			
422	Concent	89		
43 De	ciding on the number of segmentation point, and a visual tool for model cl	hecking 97		
4.4 Sin	nulation studies	100		
441	Comparing the performance of the new algorithm and the conventional	two-level		
simulat	ion method for model-based CEAs.	100		
4.5 Ca	se study of EVPPI calculations for the OPTIMAL trial	105		
4.6 Im	plementation			
4.7 Dis	cussion			
Chapter 5.	Integrated discussion and conclusions			
5.1 Ma	in contributions			
5.2 Lin	nitations and future research			
5.3 Pu	ting this research in context			
5.4 Kn	owledge transfer and exchange			
5.5 Co	ncluding remarks	124		
Bibliography				
Appendices				
Appendix	A.1: R code for stylized example of chapter 2.	136		
Appendix A.2: R code for stylized example of chapter 3				
Appendix	A.3: Exact calculation of the Expected Value of Partial Perfect Information	for		
model 1	·	140		
Appendix	A.4: Excel add-in for single parameter EVPPI calculation	144		
Appendix	A.5: R code for the efficient algorithm for EVPPI calculation	148		

## List of Tables

Table 1-1: Characteristics of patients and clinical and economic results in the OPTIMAL trial	32
Table 2-1: Outcomes of the OPTIMAL CEA without and with the incorporation of external   evidence	50
Table 3-1: Non-parametric EVSI calculation for the stylized example	66
Table 3-2: The effect of adjusting for covariate imbalance on EVPI calculations	76
Table 3-3: EVPI for various design and scenarios for the OPTIMAL trial	78
Table 3-4: Impact of different aspects of the detailed analysis on the EVPI	80
Table 4-1: Results of the simulation analysis comparing the performance of the novel and two level Monte Carlo method for EVPPI calculation10	- 04

## List of Figures

Figure 1-1: Illustration of informing decisions in health care based on the evidence2
Figure 2-1: An illustrative example of the parametric Bayesian inference vs. the vetted bootstrap approach
Figure 2-2: Incremental cost-effectiveness ratio (ICER) without and with incorporation of external evidence
Figure 2-3: Cost-effectiveness plane (CE-plane) for TFS vs. TP without and with incorporation of external evidence
Figure 2-4: Cost-effectiveness acceptability curve (CEAC) without and with incorporation of external evidence
Figure 3-1: Parametric vs. non-parametric EVSI calculations for the stylized example
Figure 3-2: Schematic illustration of the two-level bootstrap approach for EVSI calculation72
Figure 3-3: EVSI per individual for a future study with similar design as the OPTIMAL study with a range of sample sizes
Figure 4-1: Schematic illustration of the segmentation approach to EVPPI calculation90
Figure 4-2: The running cumulative sum of incremental net benefits for three parameters of a decision-analytic model
Figure 4-3: Schematic illustration and parameter specification for model 1 (left) and model 2 (right)101
Figure 4-4: EVPPI as a function of willingness-to-pay for the effect size of TFS vs. TP, using Bayesian and approximate Bayesian bootstrap107
Figure 4-5: Scatter plot of the PSA (left panel) and S` (right panel) for the effect size of TFS vs. TS for two WTP values using Bayesian and approximate Bayesian bootstrap

#### Acknowledgements

I am grateful of the generous support from Dr. Carlo Marra, my supervisor during my PhD. Dr. Marra allowed me to choose a topic that would interest me, and provided a supportive and friendly environment for me and other students. Working on this thesis at his lab was challenging but fun, and help was always around at critical times. I hope I can benefit from his mentorship for the future. My gratitude also to my committee member Dr. Stirling Bryan, who generously dedicated his time and resources. I truly appreciate the fruitful and enjoyable discussions with him. I also thank my co-supervisor, Dr. Lawrence McCandless, for sharing his deep statistical expertise along the way, and for the lively debates! I gratefully acknowledge my other committee members Drs. Larry Lynd and Craig Mitton for the time and space given to me in their busy schedules.

Many other people have helped me during this journey. My fellow graduate and colleague, Dr. Mehdi Najafzadeh, was a reliable friend as well as a source of determination and fascination with science. The contribution of my friends Drs. Antonio Avina-Zubieta, Jenifer Davis, and Mary De Vera to my survival during this period is non-negligible. I thank Jamie Thomas for kindly reading large parts of my thesis.

I was generously supported by a fellowship award from the Canadian Institute of Health Research (CIHR) as well as the Bisby fellowship award from the same institution.

Х

## Dedication

This thesis is dedicated to my parents, Aghamir Sadatsafavi and Zahra Banikarim, for their endless love, support, and encouragement.

#### **Chapter 1. Introduction**

#### **1.1 Research statement**

The goal of this thesis is to advance Bayesian methods for cost-effectiveness analysis (CEA) and expected value of information (EVI) analysis of healthcare technologies, with emphasis on the studies conducted alongside randomized controlled trials (RCTs) that use bootstrapping to carry out the CEA. RCTs provide a desirable framework for CEAs because they provide internally valid and often timely information on the comparative performance of health technologies<sup>1</sup>. The bootstrap method is a popular method among applied health economists for quantifying uncertainty in RCT-based CEAs and is one of the methods recommended by current guidelines<sup>1</sup>. An alternative approach to conduct a RCT-based CEA is to perform a fully parametric analysis. The latter can be performed in a Bayesian context, enabling analysts to incorporate prior evidence in the analysis and allowing the calculation of EVI measures. Currently, however, for an analyst using the bootstrap method and intending to incorporate external evidence or perform EVI analysis, a paradigm shift to the parametric framework is required. This thesis attempts to provide this large camp of analysts with alternative methods of Bayesian evidence synthesis and EVI calculation. The overall theme of the thesis can be described in *Figure 1-1*. This is an illustration of the steps that need to be taken whenever making decisions on the adoption of health technologies or conducting studies on the performance of such technologies. Each chapter of my thesis contributes to the methodology in one of the three steps in this figure.





CEA: cost-effectiveness analysis, EVPI: expected value of perfect information, EVPPI: expected value of partial perfect information, EVSI: expected value of sample information

Chapter 2 introduces the *vetted bootstrap*, a practical sampling algorithm for incorporation of external evidence in RCT-based CEAs. Chapter 3 suggests a non-parametric approach for calculation of the expected value of perfect information (EVPI) and the expected value of sample information (EVSI) for RCT-based CEAs based on the bootstrap. Chapter 4 provides a heuristic method for the calculation of single-parameter expected value of partial perfect information (EVPPI) that is applicable to both RCT-based and model-based CEAs. All three chapters are accompanied by case studies based on the data from the OPTIMAL trial <sup>2,3</sup>.

The remainder of this chapter is organized as follows: I discuss the inevitability of decision making under constraint resources in health care. This will lead to the notion of economic

evaluations which I describe and classify in some detail. The concept of synthesizing evidence to inform decisions is presented, and I review the relevant literature on Bayesian evidence synthesis for RCT-based CEAs. This is followed by discussing the quantification of uncertainty in the evidence with an emphasis on the role of the bootstrap method in quantifying uncertainty in RCT-based CEAs. A connection is then made to the notion of the irrelevance of inference in decision making and the debate is framed in this context to discuss the expected value of information (EVI) methodology. The EVI literature in medical decision making is reviewed. I conclude the chapter by highlighting some of the gaps in the knowledge base that motivated this research, elaborating on the general context governing the subsequent chapters, and describing the clinical trial that acts as a running example throughout this thesis.

#### **1.2 Introduction**

In the healthcare field (as in all other realms of life), a decision maker faces choices. For example, a policy maker for a public healthcare system may need to decide on public insurance coverage of either drug A or B for the treatment of a disease, or a physician at bedside may need to decide whether to make a diagnosis based on available information or ask for more diagnostic tests. A rational decision maker should have an implicit or explicit objective based on which the merits of competing decisions are compared. Decision making at the public health level is often governed by policy makers interested in maximizing some measure of population health given a finite financial budget <sup>4</sup>. As such, the decision on the adoption of competing health technologies (henceforth called the *adoption decision*) in health economics often

concerns two entities: the cost and effectiveness of health technologies<sup>\*</sup>. Economic evaluation is a set of methods and concepts comprising the philosophy, theory, methodology, and professional practice necessary to address such decisions in an objective and rational manner <sup>4</sup>.

Facing choices in healthcare is inevitable, even when only one technology is available for a health condition. Consider the situation in which the only treatment for a disease is pharmacotherapy with a specific drug. Here the option of using the available treatment can always be compared to the option of not treating patients at all. Countless other questions surface as well: should the technology be used only in a subgroup of individuals and at what cutoff; what level of intensity; delayed versus immediate treatment; and so on. Such decisions, at times, are based on educated guesses, expert opinion, or consensus. Nevertheless, an objective and quantitative approach to health technology assessment has several clear advantages. It helps comprehensively quantify all the relevant options, clearly determine the viewpoint assumed in the analysis, and optimally use the available evidence in making such decisions <sup>4</sup>.

A fundamental reality underlying economic evaluation of health technologies is that because the amount of budget the decision maker can use is limited, choosing a particular technology means fewer financial resources would be available to the others. Since the budget in a given jurisdiction is often set at a fixed value for the entire health care system by legislative authorities, the competition for resources transcends the boundaries of any specific health

<sup>&</sup>lt;sup>A</sup> A health technology, throughout this document, is widely defined as any form of healthcare product that is used for the screening, diagnosis, monitoring, or treatment of health conditions in humans. Examples include a guideline for general practitioners to treat hypertension, a screening program for cervical cancer, a blood test for detection of a toxin in the body, a drug for common cold, and a surgical procedure for treatment of pancreatitis.

condition<sup>\*</sup>. In other words, using a more expensive health technology for a particular disease means there will be fewer resources available to the technologies for all other health conditions. The question an economic evaluation is set to address is whether the additional benefits generated by adopting a technology are greater than such opportunity losses from the reduction in other technologies—that is, whether such re-allocation is efficient <sup>4</sup>.

#### 1.2.1 Types of economic evaluations

Economic evaluation studies can be categorized in different ways <sup>5</sup>. One approach is to classify according to the metrics used in quantifying the trade-offs between costs and health effects. In cost-effectiveness analysis (CEA), the incremental costs and health outcomes for competing technologies are calculated <sup>6(p96)</sup>, and health outcomes are quantified as either condition-specific (e.g., disease exacerbation avoided, or cancer-free years) or in more generic units (e.g., life years gained). Cost-utility analysis (CUA) is a specific form of the CEA in which the effectiveness outcome is expressed in a unit that is not specific to any disease and allows preferred health states to receive quantitatively higher weights compared to less preferable states, thus the difference in both quantity and quality of life between choices is materialized on a single scale. Some effectiveness outcomes satisfying these characteristics include quality-adjusted life years (QALY), disability-adjusted life years (DALYs) and healthy year equivalents (HYEs)<sup>7</sup>. Such specification of outcomes for the CUA enables the comparison of health

It also transcends into other fields if the legislative body is using a finite budget to make decisions across economic sectors. For example, for a national legislative body a decision on the adoption of a medication competes with a decision on building a dam, as both ultimately compete for a limited government budget. This generalization, nevertheless, has little relevance for the methods presented in this thesis.

technologies across unrelated health conditions. Finally<sup>\*</sup>, cost-benefit analysis (CBA) is a form of economic evaluation that requires analysis outcomes to be expressed in monetary units. As the CBA values benefits and costs in monetary units, it is not limited to only making decisions within health care, and can span within and across sectors <sup>6(p129)</sup>.

The methods developed throughout this thesis are applicable to all three forms of economic evaluation mentioned above. However, as mentioned earlier, the CUA is a sub-class of the CEA. Also, decision rules that are used to designate a health technology as cost-effective often (implicitly or explicitly) assign a monetary value for the unit of effectiveness. This practice allows the transformation of effectiveness outcomes to monetary units. As such, the CEA and the CBA, at the time of decision making, are often (at least statistically) indistinguishable<sup>8†</sup>. Given this, and for consistency, I will use the term CEA for comparative economic evaluations throughout this thesis.

An alternative way of classifying economic evaluations is in terms of the framework for the analysis. From this perspective, CEAs can broadly be classified as those that aggregate evidence from various resources in a decision-analytic format versus those that mainly estimate cost and effectiveness outcomes from patient-level data. For ease, the former group will be referred to as model-based CEAs given that a decision-analytic simulation model is often used as the framework; the latter group will be referred to as RCT-based CEAs, as RCTs are by far the most

<sup>&</sup>lt;sup>\*</sup> Another form of economic evaluation, the cost-minimization analysis, is not reported here as the focus is on the evaluations that compare both cost and health outcomes without the assumption of equivalence.

<sup>&</sup>lt;sup>•</sup> This holds when a unique willingness-to-pay (WTP) value is used to value the QALY. In fact, the logical equivalence of the CEA and CBA using a fixed WTP in contrast with the two very different assumptions underlying each approach (with regard to the normative perceptions of the role of health versus other goods in the society) has led some investigators to reject the theoretical validity of using a fixed WTP altogether <sup>9</sup>.

common type of experiment informing such analyses, although observational studies can be used for the same purpose <sup>10</sup>. In model-based CEAs, a decision-analytic model is created to simulate various aspects of the health condition under study and the impact of competing technologies on this condition is simulated to generate estimated values of cost and effectiveness outcomes associated with each alternative. On the other hand, in typical RCTbased CEAs, statistical inference is made on the individual-level data of the RCT to estimate the cost and effectiveness outcomes associated with each choice. A RCT useful for such analysis should incorporate realistic situations rather than laboratory-type conditions in its design (effectiveness versus efficacy trials). In addition, the resource use and effectiveness outcomes are often collected for each individual to enable inference on cost-effectiveness. RCTs with such characteristics are often called pragmatic trials <sup>11,12</sup>.

RCT-based CEAs comprise a large fraction of contemporary CEAs. A growing number of trials incorporate economic end-points at the design stage and there are established protocols and guidelines for conducting economic evaluation alongside a RCT <sup>1,13</sup>. Some advantages of RCT-based CEAs are practical. For example, the approval of new health technologies in many jurisdictions requires evidence generated through RCTs. Therefore RCTs are a timely (and sometimes the only) source of evidence on the performance of emerging health technologies. RCTs also have theoretical advantages for deriving a CEA. They are desirable frameworks to inform decisions because of their high degree of internal validity <sup>14</sup>. The randomization process (statistically) protects against known and unknown confounders, which, along measures to prevent performance, detection, and attrition biases, can minimize systematic errors <sup>15</sup>.

such trials can be generalized to different settings <sup>16</sup>. Decision-analytic modeling often times involves parametric evidence synthesis. One important and often overlooked aspect of evidence synthesis is modeling the correlation structure among the parameters representing the evidence <sup>17,18</sup>. In RCT-based CEAs, cost and effectiveness outcomes are estimated at the individual level, and the correlation between them and with other parameters is captured directly through patients' experience during the trial (for example, one would expect patients who experience fewer respiratory exacerbations in a COPD trial to report higher health state utility values and lower costs; however, the analyst needs not make an explicit assumption on the impact of exacerbation on costs and quality of life as such an impact is already captured through patients' experiences within the trial). The bootstrap method of RCT-based CEA indeed relies on such 'captured' correlation to enable evidence synthesis without parametric modeling.

The main focus of this thesis is RCT-based CEAs, but the methods developed in Chapter 2 and 4 can be applied to model-based CEAs as well. I will discuss the applications of the methods for model-based CEAs wherever they arise.

#### **1.2.2** Outputs of an economic evaluation

In order to maximize the reallocation of resources, health technology assessment should be concerned with the incremental cost and effectiveness of competing technologies <sup>19</sup>. A typical figure of merit to inform such reallocation decisions is the incremental cost-effectiveness ratio (ICER):

$$ICER \equiv (C_A - C_S)/(E_A - E_S),$$

where the letters C and E represent, respectively, the costs and effectiveness, and the indices A and S indicate the alternative and standard (or no) treatment, respectively.

If the data on the comparative cost-effectiveness of all health technologies in a health jurisdiction is available, then the problem of maximizing the health outcome given a fixed budget can be solved through an iterative approach based on creating a league table and ranking strategies by their ICERs <sup>20</sup>. In practice, however, it is very common to compare the ICER to the decision maker's willingness-to-pay (WTP) value <sup>21</sup>. The WTP is the maximum monetary value the decision maker is willing to pay for a unit of effectiveness. For the QALY as the effectiveness outcome, a widely accepted value is \$50,000 when the decision maker is the society itself; that is, it is widely held that society is willing to pay up to \$50,000 for gaining one QALY in an individual. Nevertheless, recent studies based on the empirical assessment of adopted technologies suggest that societies are actually paying much different values for QALYs <sup>22</sup>.

The decision based on the ICER and a non-negative WTP can be formulated in this way:

$((E_A - E_S) > 0, ICER < WTP)$	adopt the alternative technology
$\left\{ (E_A - E_S) < 0, ICER > WTP \right\}$	adopt the alternative technology
( otherwise	adopt the standard technology $ig)$

Note that comparison of the ICER with a decision threshold based on a WTP value per unit of effectiveness blurs the boundaries between the CEA and the CBA, because the decision rule described above can be transformed into

$$\begin{cases} (E_A - E_S) \times WTP - (C_A - C_S) > 0 & adopt the alternative technology \\ otherwise & adopt the standard technology \end{cases}$$

The term  $(E_A - E_S) \times WTP - (C_A - C_S)$  is called the incremental net monetary benefit (INMB) <sup>23</sup> as  $(E_A - E_S) \times WTP$  is the incremental effectiveness outcome transformed to a monetary value. Similarly, one can use the term  $(E_A - E_S) - (C_A - C_S)/WTP$ , called the incremental net health benefit (INHB), in the same decision rule <sup>24</sup>, as the term  $(C_A - C_S)/WTP$  transforms the incremental cost to its equivalent effectiveness value. In other words, a decision rule comparing the ICER against a WTP is equal to a decision rule that compares the incremental net (monetary or health) benefit with zero. In practice, inference based on the net benefit is usually subject to sensitivity analysis based on a plausible range for the WTP.

#### **1.2.3** Synthesizing information (evidence synthesis)

The evidence to support the adoption decision can be obtained from various sources. For example, evidence can be obtained from empirical (interventional or observational) studies, expert opinion, or secondary analyses (such as meta-analyses). It is obvious that as long as evidence is not generated through a mechanism that introduces bias in favor of some choices over the others, then the incorporation of more evidence increases the chance of making the correct decision; therefore the most efficient decisions are those that are based on comprehensive evidence synthesis. In model-based CEAs, where the evidence is represented through model parameters, evidence synthesis is an integral part of the effort in conducting the CEA and comprehensiveness is a matter of the rigor of the analyst. In RCT-based CEAs, on the other hand, as inference is often directly made using the individual-level outcomes, there has been less emphasis on the practice of evidence synthesis. This has made RCT-based CEAs the target of some valid criticism. Sculpher et al. <sup>25</sup>, for example, consider the lack of comprehensive evidence synthesis alongside RCT-based CEAs to be their most damning criticism.

There is a fundamental difference between RCT-based and model-based CEAs. For the former, the RCT data itself is a rich source of evidence for the decision, whereas for the latter, a decisions model is mostly a framework for propagating evidence to generate outcomes of interest<sup>\*</sup>. Thus, a decision model can be seen as a function from model parameters to the outcomes of interest. It follows then, for RCT-based CEAs, the practice of evidence synthesis from sources external to the RCT is actually combining historical (prior) and trial (current) information. A quantitative method for carrying out such a task is Bayesian evidence synthesis <sup>26</sup>. The Bayesian approach formalizes the procedure of incorporating pre-study beliefs, which are subsequently influenced by the results of an experiment, such as a clinical trial, to yield revised beliefs <sup>27</sup>.

#### 1.2.3.1 Review of literature on Bayesian evidence synthesis for RCT-based CEAs

In the mainstream biostatistics literature, Bayesian analysis of clinical trials is an active and flourishing area of research <sup>27–31</sup>. Such methods, however, have mainly focused on the statistical inference on a single trial outcome, the effect size, incorporating prior knowledge from previous trials, other experimental studies, expert opinion, and 'off the shelf' priors <sup>27</sup>. For RCT-based CEAs the interest is in the joint inference on costs and effectiveness (or inference on

<sup>&</sup>lt;sup>\*</sup> One can argue that the model structure is massively influenced by the information that the analyst has about the underlying health condition. Nevertheless, as long as issues such as model uncertainty are not taken into account, this type of information is not relevant in this thesis.

net benefit) <sup>32</sup>. If external evidence on cost and/or effectiveness is available, then the analyst can use the aforementioned Bayesian methods to combine such information with trial results.

Such a paradigm, combining prior and trial information defined on costs and/or effectiveness outcomes, has been the dominant practice in evidence synthesis for RCT-based CEAs. An early step towards adopting a Bayesian approach to RCT-based CEAs was taken by Briggs <sup>33</sup>. He suggested that a Bayesian interpretation of cost-effectiveness results offers several advantages over the frequentist paradigm including the capacity to incorporate prior evidence and the ability to make probability statements for the outcomes while retaining the robustness of the frequentist approach. Another early work in Bayesian CEA was by Heitjan et al.<sup>34</sup>. Here what captured the authors' attention was the difficulty in presenting the results of inference around the ICER as a ratio statistic. The adoption of the Bayesian paradigm was for the purpose of making probability statements around the true ICER value on which basis better inferential measures for the ICER were suggested. Given such an objective for using a Bayesian paradigm, no effort was made to incorporate any external evidence and vague priors were used in the case study. Choosing a Bayesian framework in order to provide more interpretable or robust estimates of the outcome of interest, rather than incorporating external evidence, is not limited to this example. As another example, empirical Bayes methods have been proposed in the context of multi-national RCTs in order to provide better estimates of costs and costeffectiveness outcome for individual countries by borrowing evidence from results in other countries in a hierarchical Bayesian paradigm <sup>35,36</sup>.

Al et al. <sup>37</sup> performed one of the first analyses that used informative priors based on external evidence for a RCT-based CEA. The authors used data from two trials that compared costs and effectiveness of stent implantation versus balloon angioplasty in patients with cardiovascular diseases. The outcomes of the first RCT were used as the prior information for the second study. Evidence synthesis was fully parametric with the use of conjugate priors, which resulted in closed-form equations for the posterior probability distributions. The analysis was separately performed on costs and effectiveness, and also on the joint distribution of cost and effectiveness. The authors concluded that different prior distributions may lead to different decisions.

A more systematic treatment of Bayesian evidence synthesis for RCT-based CEAs was provided by O'Hagan et al. <sup>38–40</sup>. The authors introduced a series of methods developed for the incorporation of prior information on costs and/or effectiveness in RCT-based CEAs when cost and effectiveness could be defined on a variety of scales. Because the posterior distributions were complex and direct inference was not possible, samples from the posterior distributions were derived using Markov Chain Monte Carlo (MCMC) techniques. Nixon et al. <sup>41</sup> further developed the methodology and provided a coherent set of Bayesian methods to adjust for imbalance in the distribution of baseline covariates, to provide subgroup-specific costeffectiveness outcomes, and to allow for differences between centres in a multicentre study using a hierarchical model. These methods consider costs and effects jointly, and allow for the typically skewed distribution of cost data.

As an alternative to MCMC methods, Heitjan et al. <sup>42,43</sup> used importance sampling to draw samples from the posterior joint distribution of cost and effectiveness outcomes. The case study for this report involved a RCT comparing two different interventions for the management of cardiovascular patients. Data from a pilot study was used to construct prior information on the effectiveness (survival) and informed reasoning to construct a subjective prior for the distribution of the costs.

All of the above-mentioned studies were mainly methodological and examples were used for illustrative or pedagogical purposes. RCT-based CEAs in which the results were the main message of the study and for which a Bayesian paradigm was adopted are uncommon <sup>44–46</sup>. Interestingly, in all such RCTs, vague priors have been used for the analysis. Therefore, the choice of a Bayesian paradigm in such RCT-based CEAs seems to have been for reasons other than evidence synthesis. This could be due to a preference for making probability statements about the CEA outcomes, which are more intuitive than frequentist measures, or perhaps because of the recent availability of software that makes practical MCMC methods available to applied statisticians <sup>47</sup>.

Nevertheless, it is hard to argue that no external information relevant to the adoption decision was available at the time the aforementioned RCT-based CEAs were conducted. One factor that might have restricted the analysts to non-informative priors is that, in most cases, the inference was made directly on the joint distribution of the cost and effectiveness outcomes without identifying any extra parameters during the analysis. In such a context, the prior information too could have only been defined on the cost and effectiveness outcomes. But prior

information on cost and effectiveness outcomes is rarely available and if it is, it is notoriously difficult and often inappropriate to transfer from one setting to another <sup>48</sup>. An example for this phenomenon was reported by Briggs <sup>49</sup>. Here fixed-effects analysis on the normal data was chosen for Bayesian evidence synthesis on the net benefit scale for a British RCT on blood pressure control in patients with diabetes. Prior information on costs and effectiveness from a Swedish study was available, but the author concluded that such information cannot really be transferred across the two jurisdictions and an analysis with non-informative priors was adopted.

#### 1.2.3.2 Evidence synthesis: trialist versus economist

What makes the effectiveness and costs so difficult to transfer from one setting to another is that they are, to a large extent, affected by the specific settings in the jurisdiction in which they are measured (e.g., price units for medical resources, practice patterns, organizational peculiarities, and so on). On the contrary, evidence on the aspects of the intervention that relate to the pathophysiology of the underlying health condition and the biologic impact of treatment, such as the effect size of treatment or rate of adverse events, are less affected by specific settings and are therefore more transferable. This puts the economist in a difficult situation for a RCT-based CEA: inference is made directly on the cost and effectiveness variables, but evidence is available on some other aspects of treatment that is not necessarily identified during the CEA. Here arises an important philosophical difference between the trialist and economist in the analysis of RCT data with important implications for the evidence synthesis methods used by the two camps.

- The trialist is a 'discoverer', and is interested in making inference based on RCT data primarily on the effect size and secondarily on other aspects of treatment such as safety or compliance. These measures are conceptually distinct enough to be analyzed and reported separately and trialists have a full arsenal of standard statistical methods at their disposal for such analyses. Concordantly, incorporation of prior knowledge with trial data has largely remained within such strata of results.
- The economist should inform the adoption decision. The adoption decision is an allinclusive one and cost-effectiveness is a complex function of all aspects of a health technology. As such, evidence external to the trial on any aspect of technology has bearings on the results of the CEA, and the economist does not have the luxury of dissecting RCT results into different components.

This important difference between the two camps results in specific challenges for the economist, which might not have been addressed by advances in Bayesian analysis of RCTs. Chapter 2 specifically deals with this problem.

#### **1.2.4 Decision making under uncertainty**

It is very unlikely that all the information required for an adoption decision is at hand. Evidence coming from the literature is accompanied by uncertainty due to the finite sample size of the original reports and between-study variation. If evidence is elicited from expert opinion, uncertainty lies in experts' doubt, varying levels of analysts' trust in experts' opinions, and the heterogeneity of opinion amongst experts. In RCT-based CEAs, an essential source of uncertainty is the finite sample size of the RCT. In addition, some quantities that are used in the analysis (such as unit costs) might be uncertain and sometimes there is doubt in the similarity of the RCT population and the population for which the decision will be applied. Such uncertainties will indeed result in uncertainties in the estimated cost and effectiveness outcomes.

Decision-theoretic arguments postulate that if the goal is to maximize the benefit of decisions across the population, and in the absence of irreversibilities and sunk costs (costs that will not be recovered in case the decision is annulled), a risk-neutral decision maker should choose the option with the highest expected benefit <sup>50</sup>. The calculation of the expected benefit requires proper propagation of uncertainty from the underlying evidence to the cost-effectiveness outcomes <sup>17</sup>. In the presence of uncertainty, one relevant metric for decision making is:

$$ICER \equiv \frac{\mathbf{E}(C_A) - \mathbf{E}(C_S)}{\mathbf{E}(E_A) - \mathbf{E}(E_S)}$$

(or its net benefit equivalent), with **E** indicating expectation which should incorporate all sources of uncertainty as described above<sup>\*</sup>. A robust, unbiased, and easy to implement method of calculating the expected values of costs and effectiveness is to perform a Monte Carlo simulation. This can be done by randomly drawing from the uncertain parameters in model-based CEAs <sup>51</sup> and through bootstrapping in RCT-based CEAs <sup>52,53(pp54–58)</sup>, and averaging the calculated outcomes. This method in the health economics literature is commonly referred to

<sup>&</sup>lt;sup>\*</sup> In many contemporary model-based CEAs, only the point estimate of parameters is used for the calculation of the ICER in the so-called base case analysis. This is indeed different from the ICER described above (because the function of an expected value is different from the expected value of a function). Therefore, the ICER calculated by using the point estimate of the parameters can only be taken as an approximate value for the true value of the ICER that should inform the adoption decision.

as probabilistic sensitivity analysis (PSA) <sup>17,51</sup>. For very simple decision-analytic models, sometimes the joint distribution of the cost and effectiveness outcomes can be analytically calculated. Likewise, for RCT-based CEAs, there are parametric methods that can be used to calculate the expected value of cost-effectiveness outcomes <sup>54,55</sup>. Overall, however, the PSA remains a very popular method of quantifying uncertainty in a CEA.

#### 1.2.5 The bootstrap method for the analysis of uncertainty in RCT-based CEAs

The bootstrap method is one the most popular methods for the analysis of uncertainty in RCTbased CEAs <sup>52,54,56,57</sup>. In this approach, the distribution from which data for a trial arm is sampled (often referred to as the *population distribution*) is approximated by the empirical distribution of the sample. An empirical distribution of a real-valued vector x of size n is the probability distribution constructed by putting a probability mass of 1/n on each element of x<sup>58</sup>. The sampling distribution of the population distribution, and hence any statistics based on that, can then be estimated by sampling from the empirical distribution (i.e., sampling with replacement with the same number of observations as are in the original data set x). For a RCT, the data structure is more complicated than that of a real-valued vector. But the bootstrap remains a valid inferential technique as long as bootstrapping mechanism mirrors the mechanism that has generated the data <sup>59</sup>, which, for parallel arm RCTs, means obtaining bootstrap sets separately within each arm of the RCT <sup>52,53,60</sup>. The data generated through bootstrapping can be used to calculate the ICER, estimate a confidence region around the ICER, draw the acceptability curves, and so on. The popularity of the bootstrap method for RCT-based CEAs can be attributed to several factors. First, inference in CEAs is on the joint distribution of cost and effectiveness. This makes the familiar univariate statistical tools for parametric inference unavailable. Another problem is the distribution of costs and effectiveness values <sup>61,62</sup>. Costs are often right-skewed and zero-inflated. An outcome like health state utility value can take any values from less than zero to one. Joint modeling of such outcomes, adjusting for covariate imbalance, and imputation of missing and censored values in a unified framework require statistical expertise and will likely result in a context-specific statistical model with a multitude of parametric assumptions. The bootstrap on the other hand is a robust non-parametric approach which is simple to implement, involves fewer distributional assumptions, and allows the propagation of joint uncertainty on all such aspects through an iterative resampling scheme. However, one should note that the validity of the bootstrap approach is based on two asymptotics: the sample size of the RCT approaches the size of the target population, and the statistics of interest is estimated from infinite cycles of bootstrapping <sup>53(p55)</sup>.

Note that the above description of the bootstrap is in line with its popular frequentist interpretation (as it estimates the sample distribution of the parameter of interest). Rubin et al. added a Bayesian dimension to the bootstrap by introducing the Bayesian bootstrap <sup>63</sup>. The Bayesian bootstrap is very similar both in operation and in numerical results to the conventional bootstrap <sup>63,64</sup>. However, theoretically, it estimates the posterior distribution of the parameter of interest. Such a Bayesian interpretation of the bootstrap motivates some of the widely accepted statistical methods such as the non-parametric imputation of missing data

<sup>65</sup>. A more detailed description of the Bayesian bootstrap will be provided in Chapter 2 as it plays a pivotal role in the method developed in that chapter.

#### **1.2.6 Expected Value of Information (EVI)**

Despite the need for basing the adoption decision on the expected value of cost and effectiveness outcomes, the consumers of cost-effectiveness studies often remain interested in how strong the evidence is in support of some decisions over others. The data generated through the PSA can be used to communicate the level of uncertainty in various ways. Examples include the cost-effectiveness plane, cost-effectiveness acceptability curve, or confidence regions and intervals around the ICER and net benefit <sup>60,66</sup>. Unfortunately, aside from illustrative purposes, none of these methods have any rationale in decision making and the adoption decision should always remain based on the expected value of the costs and effectiveness of comparative options <sup>50</sup>. This raises the question as to the relevance of such methods for the presentation of uncertainty in medical decision making.

From a decision-theoretic viewpoint, uncertainty in a decision only matters when the decision maker is interested in collecting more evidence <sup>50</sup>. Decisions that are made in the face of substantial uncertainty benefit more from additional evidence than those for which there is little doubt in the merit of the chosen option. Expected value of information (EVI) analysis includes a set of concepts and methods stemmed from decision theory that deal with the impact of uncertainty on the outcome of a decision <sup>50,67–71</sup>.

From this perspective, information is valuable because it increases the chance that the right decision is made. Thus, there is an opportunity loss associated with uncertainty, which is

generally determined by the combination of the probability that a decision based on the current information is wrong on one hand and the marginal loss of benefit due to the wrong decision on the other hand <sup>67</sup>. This means investment in obtaining more information (evidence) can be associated with a return on investment because it increases the chance of choosing the optimal option. EVI methods allow for the quantification of such return on investment and present its magnitude in the same unit as the benefit (often defined on the NMB scale) of the decision <sup>67</sup>. There are several measures of EVI, some of the most widely used are:

1) <u>Expected Value of Perfect Information (EVPI)</u>: the EVPI quantifies the overall value of resolving all uncertainty in the decision task <sup>72</sup>. It answers the question: *what is the expected gain in benefit by completely resolving uncertainty around all evidence used in making the decision?* For a risk-neutral decision maker, it is defined as the value of decision situation with perfect information *minus* the value of current decision situation:

$$EVPI \equiv \boldsymbol{E}_{\boldsymbol{\Omega}} \{ \max_{i} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \} - \max_{i} \{ \boldsymbol{E}_{\boldsymbol{\Omega}} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \},\$$

where  $\Omega$  is the set of all stochastic (uncertain) quantities informing the decision, **Y** is the set of non-stochastic quantities, and  $NB_i$  is the function calculating the net benefit for the  $i^{th}$  option<sup>\*</sup>. Alternatively, the EVPI can be seen as the opportunity loss in the decision with current information, as such opportunity loss will be completely avoided with the availability of perfect information <sup>71</sup>.

<sup>&</sup>lt;sup>\*</sup> This notation departs from the style most often used in the literature because the equations for EVI measure are mostly discussed in the context of decision-analytic CEAs. The notation adopted here is intended to be more flexible. For example, *Q*, the stochastic entity, can be the set of uncertain parameters in decision-analytic CEAs, or the population distribution of the RCT in RCT-based CEAs.

2) Expected Value of Partial Perfect Information (EVPPI, or partial EVPI): This is the expected gain in benefit by completely resolving uncertainty around specific aspects of the evidence used in the decision task <sup>73</sup>. It answers the question: *what is the expected gain in benefit by completely resolving uncertainty around a selected set of evidence in a decision-making task?* It is also equal to the opportunity loss due to uncertainty in the value of the subset of evidence. If the subset of evidence of interest is denoted by  $\Omega^{I}$  and the rest (complementary set) of the evidence by  $\Omega^{C}$  then

$$EVPPI_{\Omega^{I}} \equiv \boldsymbol{E}_{\Omega^{I}} \max_{i} \{ \boldsymbol{E}_{\Omega^{C} \mid \Omega^{I}} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \} - \max_{i} \{ \boldsymbol{E}_{\Omega} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \}.$$

3) Expected Value of Sample Information (EVSI): This is the expected value of gain in benefit obtained from conducting an experiment with a given design and sample size <sup>68</sup>. It answers the question: what is the expected gain in benefit in conducting an experiment with a particular design and sample size that gives us information for a particular decision making task? If the future study provides us with a probability distribution *D* that carries information about  $\Omega^{I}$ , then EVSI can be calculated as

$$EVSI_{\boldsymbol{\Omega}^{l}} \equiv \boldsymbol{E}_{\boldsymbol{D}} max_{i} \{ \boldsymbol{E}_{\boldsymbol{\Omega}^{c} \mid (\boldsymbol{\Omega}^{l} \mid \boldsymbol{D})} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \} - max_{i} \{ \boldsymbol{E}_{\boldsymbol{\Omega}} [NB_{i}(\boldsymbol{\Omega}, \boldsymbol{Y})] \}.$$

Again, an alternative way of deriving the EVSI is to see it as the difference between the EVPI before and after conducting the study <sup>71</sup>.

The above equations for EVPI, EVPPI, and EVSI typically calculate per-individual values (because  $NB_i(.)$  typically returns per-individual net benefit). To clarify, a per-individual EVSI of \$400 (on the NMB scale) associated with a future RCT means that the adoption decision, after the trial

results become available and the evidence is incorporated into the decision, will be associated with an expected extra \$400 gain in the NMB compared to the current decision for every instance the cost-effective technology is used. Because the cost-effective technology will be used for many patients, the overall *population EVSI* provides the expected return on investment from the study. This extension from individual to population values apply to the EVPI and EVPPI as well. In general, population EVI is the product of per-individual EVI and the total number of times the adoption decision is expected to be made. If patients arrive at rate *r* in the future, and if the results of the study will become available after time *t* and the decision will be relevant for a time T (*T*>*t*, and assuming both *t* and *T* are multiplicative of the time unit), then the population EVI, discounting the future events at rate *c*, can be calculated as <sup>74</sup>

$$EVI_{population} = \sum_{i=t}^{T} (r. EVI) / (1+c)^{i}$$

Note that the later the information becomes available, the smaller the population EVI will be. For a RCT, this is the time spent for planning, recruitment, analysis, and dissemination of the results, during which the population does not receive the benefit from the information provided by the trial. Therefore, unlike per-individual EVSI, the population EVSI will not necessarily be a monotonically increasing function of the sample size.

#### 1.2.6.1 Review of literature on EVI

#### 1.2.6.1.1 EVI methods in medical decision making

EVI methods have firm foundations in statistical decision theory and have been successfully used in other fields such as operations research and environmental risk analysis <sup>75–78</sup>. An EVI

approach to evaluate uncertainty in medical decision making was at first focused on modelbased CEAs. The EVPI was the first EVI measure introduced to the health economics community as an index for measuring the sensitivity of decision-analytic models to the variation in their inputs. The EVPI was argued to be a theoretically sound and a practically valid method for sensitivity analysis <sup>67</sup>. Claxton <sup>50</sup> more directly linked EVI measures to the prioritization of future research and argued that EVI can be used to inform the decision on acquiring more information once the adoption decision is made. Methods for EVI analysis alongside RCT-based CEAs soon followed <sup>70,71,79–86</sup>.

Among the EVI measures, the EVSI has a particular appeal in the context of RCT-based CEAs because the EVSI, along with the estimate of the overhead and per-subject costs of the trial, can be used to inform the design and specifically the optimal sample size of subsequent RCTs<sup>71</sup>. This provides a theoretically rigorous alternative for contemporary methods for sample size calculations in RCTs. Traditionally, the statistical analysis of RCTs has been based on the discovery of the differential effects in the efficacy of competing interventions based on frequentist hypothesis testing. When a RCT is aimed at reporting the statistical significance on the primary outcome, the sample size calculation is concordantly based on power analysis, giving the trial a pre-specified level of type I and type II errors<sup>87</sup>. However, given that the results of (pragmatic) RCTs are eventually used to inform the adoption decision, it has been argued that designing a RCT should directly incorporate such considerations<sup>88</sup>.

This approach brings the trial design and the adoption decision into a single framework of constraint optimization: how to best use a fixed budget to maximize population health through

adopting optimal technologies and prioritizing research on the comparative performance of technologies. Choosing the relevant course of action in this context will be informed by the CEA and EVI analyses. In the absence of irreversibilities, one possible scenario can be formulated as follows: the adoption decision is made based on the CEA. Next, the EVPI is used to evaluate if there is any need for collecting more evidence. If the EVPI is positive, one can then use the EVPPI to assess if research in any specific dimensions of evidence is particularly promising. For example, if the uncertainty in the quality of life of patients under treatment is the only important source of uncertainty, then a non-randomized study measuring this parameter in a sample of patients and controls might be an economical way of gathering the required information. If there are several sources of uncertainty, conducting a future RCT that will provide evidence on all such parameters might be justified. Once a particular study is found to be potentially worthwhile, the next step is to find the best design and the sample size of such study. The most optimal design of a RCT is the one that maximizes the difference between the population EVSI and the cost of the trial. This difference is called the expected net gain (ENG)  $^{71}$ , or the expected net benefit of sampling (ENBS)<sup>89</sup>. A future RCT is worthwhile if the ENG is positive, and the optimal sample size is the one that maximizes the ENG.

It should be noted that in implementing new technologies there are often sunk costs arising from reversing information flow, change in practice to health care providers, and irreversibilities such as necessary capital equipment, training, and so on that might not be reallocated to alternative uses <sup>83</sup>. These costs might be high enough that once the adoption decision is made, it becomes irreversible <sup>90</sup>. Consideration of such irreversibilities will create situations in which delaying the adoption of the optimal technology might be beneficial, and
the relevant course of action will become one of 'adopt and no trial', 'adopt and trial', or 'delay and trial'. Again, the CEA and EVI, in conjunction with the estimate of irreversible and sunk costs, will provide a rigorous answer <sup>83</sup>.

#### 1.2.6.1.2 EVI calculations for RCT-based CEAs

Compared to EVI analysis for decision-analytic models, which has a deep root in the operations research and risk analysis literature <sup>78</sup>, EVI analysis for RCT-based CEAs is a younger discipline that is more restricted to the health care field. Methods have been developed for the calculation of EVPI, EVPPI, and EVSI for RCT-based CEAs <sup>70,71,79,79-86,91</sup>. Such methods are almost entirely based on the normal approximation of the distribution of the expected value of the (incremental) net benefit for trial arms. This turns the output of the RCT into a multivariate normal distribution of net benefits for which the algorithms for EVI analysis for model-based CEAs can readily be applied (such multivariate distribution can indeed be considered a simple decision-analytic model). For example, EVI calculation using Monte Carlo simulation by sampling from the distribution of the NMB has been used to calculate EVPI, EVSI, and EVPPI by Koerkamp et al. <sup>86</sup>. The authors analyzed patient-level data on a RCT comparing two treatments for intermittent claudication. The expected value of the NMB associated with each treatment was assumed to follow a normal distribution. In addition to the EVPI, the EVSI and EVPPI for the whole as well as sub-sets of parameters were calculated.

In some general situations, EVI equations based on such normal approximation can be expressed in closed form, providing a more elegant and efficient solution than Monte Carlo methods. For example, for a two-arm RCT, when the expected value of the incremental net

benefit for the current and future RCTs are assumed to be normally distributed, the EVPI and EVSI can be calculated using equations that are related to the unit normal loss integral <sup>71</sup>. Eckermann et al. developed this paradigm further for the situation where there is a reversibility cost, and provided equations for the expected value of benefit associated with the option of 'delay and trial' <sup>83</sup>. This framework has been extended to multi-stage RCTs <sup>80</sup>, when the implementation of the optimal technology is not perfect <sup>79</sup>, when the decision maker is a third party (industry) that aims at maximizing their expected profit <sup>82</sup>, and when different jurisdictions have the option of conducting their own RCTs and/or borrowing evidence from the RCTs conducted by other jurisdictions <sup>84</sup>.

For model-based CEAs, the EVPI can directly be calculated from the PSA data <sup>17</sup>. Some authors have already mentioned that the data generated through bootstrapping for RCT-based CEAs can be used to calculate the EVPI in the same way <sup>86,92</sup>. However, the same authors mention that it is not generally known how to perform EVSI calculations using the bootstrap, a topic that will be dealt with in Chapter 3 of this thesis. Calculating the EVPPI for RCT-based CEA has not been thoroughly explored in the literature. In the study by Koerkamp et al. <sup>86</sup>, the EVPPI was calculated for components of cost and effectiveness parameters by assigning a multivariate normal distribution to the cost and QALY components. The methodology used for such calculations does not seem to be readily extendable to calculate EVPPI for other parameters such as the effect size of the treatment, because the framework for the analysis directly generates a probability distribution for the cost-and effectiveness outcomes without identifying such intermediary parameters. This is the focus of the research presented in Chapter 4.

#### **1.3** Gaps in current knowledge that inspired this dissertation

Because the bootstrap is such a popular method in RCT-based CEAs, it would be attractive to further expand this framework. Currently, investigators who intend to incorporate external evidence into RCT-based CEAs or perform EVI analysis require a paradigm shift to parametric modeling. In this context, some of the main gaps in the current knowledge can be summarized as follows:

1) In its current form, the bootstrap approach in the analysis of RCT-based CEA does not allow the incorporation of external information in the analysis. The need for the incorporation of external information, whenever available, in CEAs should not be affected by the choice of the analytic framework. As such, it is desirable to develop methods for evidence synthesis for RCTbased CEAs that use the bootstrap.

2) The current methods of EVSI calculations for RCT-based CEAs are largely based on the normal approximation of the expected value of the incremental cost and effectiveness (or net benefit) across treatment arms. The central limit theorem seems to make this a theoretically justifiable approach. However, making this approximation requires some simplifications in modeling the uncertainty due to censored and missing values and covariate imbalance. It is desirable to develop methods that can more accurately model such realistic aspects of the analysis of RCT-based CEAs. In addition, if the bootstrap is used for a RCT-based CEA, it will be natural to use an EVI method that is based on the same paradigm and the implicit assumptions inherent in the bootstrap method.

3) EVPPI calculation for RCT-based CEAs is largely undeveloped. After the results of a RCT become available, it would be of interest to evaluate if there is a need for generating further evidence for particular aspects of the decision so that any future study can be geared towards measuring such parameters. Again, it is desirable to use such EVPPI calculations based on the bootstrap method of RCT-based CEAs given its popularity and familiarity among applied health economists.

#### **1.4 Context**

The methods described in this thesis and the numerical results are based on a set of assumptions that will be mentioned here in order to avoid repetition. Throughout this work, it is assumed that the decision maker is a risk-neutral agent that tries to maximize the benefit to the target population, which is the population from which the RCT data is sampled. I focus on decision problems with discrete alternatives, i.e. that out of the total *D* alternative decisions, the decision maker will choose the one with the highest NB. The decision-maker has access to a fixed financial budget and has a fixed WTP value for a unit of effectiveness. The results of the CEA and EVI are all reported as per individual and no attempt has been made to transform the reported results into their population values as such calculations are for the most part independent of the method of per-individual EVI calculation.

#### 1.5 The OPTIMAL clinical trial as a case study

The OPTIMAL clinical trial is used as a running example for this thesis <sup>2,3,93</sup>. Although several parts of this thesis are inspired by issues surrounding the analysis of the OPTIMAL RCT, none of the methods developed is specific to any aspect of the OPTIMAL study, and the case studies are

selected in order to provide an opportunity to discuss the practical aspects of implementing the methods.

The OPTIMAL clinical trial was a multi-centre Canadian study evaluating the benefit of combination therapy in patients with chronic, obstructive pulmonary disease (COPD). It was conducted among 27 academic and community medical centres in Canada from 2003 to 2006. The study included 449 patients randomized into three treatment groups: tiotropium plus placebo (TP, the current standard of care, N=156), tiotropium plus salmeterol (TS, N=148), or tiotropium plus fluticasone-salmeterol (TFS, N=145). The hypothesis was that combination therapy in COPD improves outcomes, and the primary outcome measure was the proportion of patients who experienced a respiratory exacerbation by the end of follow-up (52 weeks).

Basic demographic characteristics and clinical and economic results of the OPTIMAL trial are presented in *Table 1-1.* At the end of follow-up, the proportion of patients who experienced at least one respiratory exacerbation was not significantly different across the groups (TP=62.8%, TS=64.8%, TFS=60.0%, P=0.69). The difference in this proportion between TS and TP was -2.0 percentage points (95% CI -12.8 to 8.8), and for the TFS vs. TP, it was 2.8 percentage points (95% CI -8.2 to 13.8).

Nevertheless, patients in combination therapy showed improvements in a number of secondary outcomes: TFS improved lung function (p-value=0.049) and disease-specific quality of life (p-value=0.01) and reduced the number of hospitalizations for COPD exacerbation (incidence rate ratio, 0.53 [95% CI 0.33 to 0.86]) and all-cause hospitalizations (incidence rate ratio, 0.67 [95%

CI 0.45 to 0.99]) compared with the TP. In contrast, the TS did not statistically improve lung function, hospitalization rates, or quality of life compared with the TP.

The OPTIMAL trial protocol included a concurrent prospective economic analysis. Data on resource use and quality of life (measured using a disease-specific quality of life questionnaire and converted into utility values <sup>94</sup>) were collected during the trial. The cost-effectiveness analysis revealed that the average patient in the TP group generated \$2,678 in direct medical costs (costs are all in 2008 Canadian dollars), while the TS and TFS generated \$2,801 and \$4,042, respectively. The average QALY for the TP group was 0.7092. When QALYs were adjusted for baseline utility, the TS group showed a decrease of –0.0052 and the TFS group showed an increase of 0.0056 in QALY compared to that of the TP group. As such, the TS strategy was dominated by TP and TFS. Compared with the TP, the TFS strategy resulted in ICER of \$243,180 per QALY gained. At the willingness-to-pay (WTP) value of \$50,000/QALY, the TP strategy remained the best choice, and in probabilistic sensitivity analysis, it had 78% chance of being the cost-effective option.

For a full review, the reader can refer to the publication on the design  $^{93}$ , results of the main analysis <sup>2</sup>, and the economic evaluation <sup>3</sup> of the OPTIMAL study.

	TP	TS	TFS		
Ν	156	148	145		
Mean age (SD), y	68.1 (8.9)	67.6 (8.2)	67.5 (8.9)		
Women, %	46.2	42.6	42.1		
Prebronchodilator lung function	1.01 (0.38)	1.00 (0.44)	1.05 (0.38)		
Baseline utility	0.6919	0.7055	0.6931		
Patients with 1 acute exacerbation, n (%)	98 (62.8)	96 (64.8)	87 (60.0)		
Absolute risk reduction (95% CI), percentage points	reference	-2.0 (-12.8 to 8.8)	2.8 (-8.2 to 13.8)		
All exacerbations, n	222	226	188		
Mean cost	2678 (1950 to 3536)	2801 (2306 to 3362)	4042 (3228 to 4994)		
Mean QALY	0.7092 (0.6953 to 0.7228)	0.7124 (0.6931 to 0.7310)	0.7217 (0.7034 to 0.7389)		
Mean adjusted QALY	reference	-0.0052 (-0.0088 to 0.0032)	0.0056 (-0.0142 to 0.0251)		

#### Table 1-1: Characteristics of patients and clinical and economic results in the OPTIMAL trial

*QALY: quality-adjusted life year, TP: tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol, SD: standard deviation, CI: confidence interval* 

#### **1.6 Contents of this thesis**

Overall, the contribution of this thesis research to the knowledge base in the health economics literature includes a coherent set of methods that can be seen as an extension of bootstrap method for the CEA of RCTs. Chapter 2 discusses the Bayesian interpretation of the bootstrap and, based on such interpretation, introduces the *vetted bootstrap*, a practical sampling algorithm for incorporation of external evidence in RCT-based CEAs. Chapter 3 takes the framework proposed in Chapter 2 one step further, and suggests a non-parametric approach for calculation of the expected value of perfect information (EVPI) and the expected value of sample information (EVSI) for RCT-based CEAs based on the bootstrap. Chapter 4 provides a heuristic method for the calculation of single-parameter expected value of partial perfect information (EVPPI), which is applicable to both model-based and RCT-based CEAs. Chapter 5 provides an integrated discussion elaborating on the common features of the methods developed throughout this thesis and discusses some limitations and suggestions for future research. Some details that I deemed not to be an integral part of the argument are presented in a series of appendices. The appendices also contain computer codes for the implementation of the algorithm described in Chapter 4. Also provided in the appendices are the computer codes for the stylized examples in this thesis so that the reader can reproduce the results or uses different input values for further examination. The code used for the analysis of the OPTIMAL trial is not provided here as it is not a generic, reusable computer code, but the reader can access and examine the code at http://www.core.ubc.ca/~msafavi/thesis

# Chapter 2. The Vetted Bootstrap, a practical algorithm for the incorporation of external evidence in RCT-based CEAs.

#### 2.1 Background

In the contemporary Bayesian analysis of RCTs the focus is on statistical inference on a single trial outcome, most often the effect size, and the incorporation of prior knowledge too remains focused on such outcomes <sup>27–29,31</sup>. For trial-based CEAs, if external evidence on cost or effectiveness (or the net benefit) is available, then the analyst can use such Bayesian methods to combine this information with trial results. This has been the dominant paradigm in the Bayesian analysis of RCT-based CEAs <sup>34,37,39,43,49</sup>. However, prior information on cost and effectiveness is rarely available and if it is, it is often inappropriate to transfer to other settings <sup>95</sup>. In the context of a RCT-based CEA, external evidence is often available on parameters that relate to the underlying biology of the health condition and the impact of treatment. Such parameters indeed affect cost and effectiveness, but are not necessarily quantified in the course of a RCT-based CEA.

As a motivating example, consider a hypothetical trial in which the economist is interested in treatment costs (direct and total costs) and effectiveness (quality adjusted life years and mortality), all of which are collected at the individual level. Without any external evidence, the economist can make direct inference on the joint distributions of cost and effectiveness outcomes across the trial arms. Yet, imagine there is also external evidence on the treatment effect size from another RCT, as well as adverse event rates for the control arm from an observational study. How can external evidence on such parameters be incorporated in the

analysis and then be propagated to the cost and effectiveness outcomes? One way to do so is to create a parametric model to connect cost-effectiveness outcomes with parameters for which external evidence is available. The model can be updated using a variety of techniques such as Markov Chain Monte Carlo (MCMC) methods or through the maximization of likelihood as used in the Confidence Profile method <sup>96</sup>. However, the model must connect several parameters through link functions, regression equations, and error terms. This involves a multitude of parametric assumptions, and there is always the danger of model misspecification <sup>61,62</sup>. In addition, implementing such a model and comprehensive model diagnostics are not an easy undertaking. The 'vetted bootstrap', presented in this chapter, requires fewer parametric assumptions and is generally much easier to implement.

### 2.2 The vetted bootstrap: a practical approach to evidence synthesis in RCTbased CEAs

The vetted bootstrap is an extension of the popular bootstrap method of RCT-based CEAs <sup>52,54,97</sup>. It is a semi-parametric method in that it requires a parametric specification of the external evidence while avoiding parametric assumptions on the cost-effectiveness outcomes and their relationship with the external evidence. The method presented is a form of rejection sampling <sup>98</sup> applied to the bootstrap sample – an approach that is very simple to implement.

The remainder of this chapter is structured as follows: after outlining the context, a Bayesian interpretation of the bootstrap is presented. Next, the theory of the incorporation of external evidence into such sampling scheme is explained. A highly stylized example shows the step-by-step implementation of the vetted bootstrap method in a simple scenario for which an

(approximate) analytical solution exists. A case study featuring the OPTIMAL trial shows the practical aspects of implementing such a method. A discussion section on the various aspects of the new method and its strengths and weaknesses compared to parametric approaches concludes the chapter.

#### 2.3 Theory and methods

Let  $\mathbf{X} = \{X_{ij}: i = 1, 2, ..., N, j = 1, 2, ..., n_i\}$  represent the individual-level data of an N-arm clinical trial, with sample size of  $n_i$  in the  $i^{th}$  arm, where  $X_{ii}$  is the data of the  $j^{th}$  person in the  $i^{th}$ arm.  $X_{ij}$  is the set of all the measured quantities that are required to inform the costeffectiveness analysis. For example, it can consist of the individual-level resource use and quality of life weights, baseline covariates that are used for adjusting outcomes, as well as the parameters for which external evidence is available. I do not index the elements inside  $X_{ij}$  as the unit of sampling in subsequent sections is  $X_{ij}$  in its entirety. Let  $\mathbf{F} = \{F_1, F_2, \dots, F_N\}$  be the set of unknown population distributions from which the data for each individual within each arm of the trial is generated ( $\forall i, j X_{ij} \sim_{i.i.d} F_i$ , which I also indicate by the short-hand notation  $X \sim F$ ). Let  $\varphi$  be the vector of some quantities of interest relevant for cost-effectiveness analysis (e.g., expected cost and effectiveness values for each intervention, or the net monetary benefit)<sup>\*</sup>. Let the functional form  $\varphi(\mathbf{F})$  represent the unobserved *population value* of  $\varphi$  that can be identified if **F** is fully specified. Likewise let  $\theta$  be the vector of some parameters for which external evidence is available (henceforth called auxiliary parameters). Examples include

<sup>&</sup>lt;sup>\*</sup> While the measures of economic evaluation are of interest here,  $\varphi$  can be generalized to many other types of inference as well.

the odds ratio (OR) of treatment success between the two arms of the trial, the rate of adverse events estimated from an observational study, or compliance rate estimated from another dataset. Again, I use the functional form  $\theta(\mathbf{F})$  to represent the unobserved population value of  $\theta$ . The auxiliary parameters should have a fundamental property:  $\theta(\mathbf{F})$  should be identified (unique) for each  $\mathbf{F}$ , but many different  $\mathbf{F}$ s can have the same  $\theta(\mathbf{F})$ . This definition seems to hold for the most common forms of external evidence, such as the examples used above. Let  $\pi_{\theta}(.)$  be the probability density function of  $\theta$  representing our external information. The distribution assigned to the external evidence should satisfy the following conditions<sup>\*</sup>: 1)  $\pi_{\theta}(.)$ should be constructed without any influence from trial data  $\mathbf{X}$  (no hindsight bias). 2)  $\pi_{\theta}(.)$ should have a finite maximum value. This condition will be violated for some distributions such as  $Beta(\alpha, \beta)$  when either  $\alpha$  or  $\beta$  is less than one. 3)  $\pi_{\theta}(.)$  should be a continuous distribution. This is related to the previous condition as discrete or mixed-type distributions have probability density functions that take infinite values.

My goal in this chapter is to generate random samples from the posterior distribution of  $\varphi$  having observed the RCT data and the distribution of the external evidence. I use the short-hand notation  $P(\varphi | \mathbf{X}, \pi_{\theta})$  for this quantity.

#### 2.4 A Bayesian interpretation of the bootstrap

A key step for the vetted bootstrap method is to treat probability distributions as random entities that, much like random variables, are subject to probability calculations and statistical

<sup>&</sup>lt;sup>\*</sup> Some of these conditions might be relaxed in specific situations. For example, the second condition should not necessarily hold for the weighted bootstrap method; the third condition can also be relaxed if the population value only takes discrete values.

inference, including the Bayes' rule. If we treat the population distribution  $\mathbf{F}$  as a random entity, we can use Bayes' rule to update our knowledge of  $\mathbf{F}$  based on the observed  $\mathbf{X}$ :

$$P(\mathbf{F}|\mathbf{X}) \propto \pi_{\mathbf{F}}(\mathbf{F}). L(\mathbf{X}|\mathbf{F}), \tag{1}$$

where  $\pi_F(F)$  is our prior distribution on **F**, P(F|X) is our posterior distribution of **F** having observed the trial data **X**, and L(X|F) is the likelihood of data.

If prior and posterior distributions on  $\mathbf{F}$  are from a parametric family indexed by a set of distribution parameters, then the randomness of F would translate into the randomness of those parameters. However, one can perform such Bayesian inference non-parametrically: Rubin  $^{63}$  showed that if we assume a prior non-informative Dirichlet process for **F** (a prior of the form Dirichlet(0, ..., 0) on the vector of all possible observations from **F**), then we can directly draw from  $P(\mathbf{F}|\mathbf{X})$  using a simple process called the Bayesian bootstrap. In the Bayesian bootstrap of a vector  $\mathbf{Z}$ , a weight vector  $\mathbf{W}$  of the same size is generated from a random variable with Dirichlet distribution with scale parameter of one ( $\mathbf{W} \sim Dirichlet(1, ..., 1)$ ), such that the *i*<sup>th</sup> element of **W** is the weight assigned to the *i*<sup>th</sup> element of **Z**. The probability distribution defined by putting point mass W<sub>i</sub> on Z<sub>i</sub> can be considered a random draw from the distribution of  $P(\mathbf{F}|\mathbf{Z})$ . It turns out that the conventional bootstrap can also be interpreted, with some approximation, in the same way; in this case the vector W is generated from a scaled multinomial distribution <sup>99(p125)</sup>. Although this sampling method does not correspond to a formal Bayesian inference, the similarity in the operation and results to the Bayesian bootstrap allows one to interpret the conventional bootstrap in the same way (see <sup>64</sup> for an extensive study of the numerical similarity of the two methods). Rubin called this the approximate

Bayesian bootstrap <sup>99(p124)</sup>. This method is used as an alternative to the Bayesian bootstrap in some circumstances like the imputation of missing data <sup>65,99</sup> and the weighted likelihood bootstrap <sup>100</sup>. As such, I adopt a general notation and use the term **Z**\* to indicate a bootstrap sample of a vector **Z** with the bootstrap weights generated using Bayesian or approximate Bayesian mechanisms. The *empirical distribution* of **Z**\* generated by putting probability mass  $W_i$  on  $Z_i$ , can be interpreted as a random sample from the posterior distribution of  $P(\mathbf{F}|\mathbf{Z})$ .

## 2.5 Cost-effectiveness analysis without the incorporation of external evidence

In a CEA in which we do not intend to incorporate the external evidence, we are interested in generating a random sample from the distribution of  $\varphi(\mathbf{F}|\mathbf{X})$ . This sample can then be used to estimate the expected value of cost and effectiveness outcomes across interventions as well as to characterize uncertainty in a variety of ways, including cost-effectiveness planes, acceptability curves, or intervals around the incremental cost-effectiveness ratio or incremental net benefit.

With bootstrapping acting as sampling from  $P(\mathbf{F}|\mathbf{X})$ , one can generate a random sample from the distribution of  $\varphi(\mathbf{F}|\mathbf{X})$  using a simple Monte Carlo approach:

- 1. For *q*=1, ..., *M*, where *M* is the number of simulations:
- Generate X\*, a (Bayesian) bootstrap sample of X with bootstrapping performed within each arm of the trial.

- 3. Calculate  $\varphi^* = \varphi(\mathbf{X}^*)$ , the cost and effectiveness outcomes from the bootstrap sample.
- 4. Store the value of  $\varphi^*$  and jump to 1.

This approach generates *M* random draws from the posterior distribution of the cost and effectiveness outcomes having observed the RCT data. If the conventional (approximate Bayesian) bootstrap is used, this algorithm becomes similar to the method commonly used to quantify uncertainty in RCT-based CEAs <sup>52,54</sup>.

#### 2.6 Incorporating external evidence

To incorporate external evidence, I specify the information on  $\theta$  by putting a further prior on **F** representing the prior knowledge on the value of  $\theta$  measured in **F** (note the requirement on the identifiability of  $\theta$  in **F** made earlier):

$$P(\mathbf{F}|\mathbf{X}, \pi_{\theta}) \propto \pi_{\theta}(\theta(\mathbf{F})) \cdot \pi_{\mathbf{F}}(\mathbf{F}) \cdot L(\mathbf{X}|\mathbf{F}) \propto \pi_{\theta}(\theta(\mathbf{F})) \cdot P(\mathbf{F}|\mathbf{X}),$$
(2)

Estimating  $\varphi(\mathbf{F})$  from a random sample from this distribution provides a random sample from  $P(\varphi|\mathbf{X}, \pi_{\theta})$ . What equation (2) tells us is that for any given  $\mathbf{F}$ , the term  $P(\mathbf{F}|\mathbf{X}, \pi_{\theta})$  is a weighted value of  $P(\mathbf{F}|\mathbf{X})$  with weights being  $\pi_{\theta}(\theta(\mathbf{F}))$ . With the empirical distribution of  $\mathbf{X}^*$ acting as a random sample from  $P(\mathbf{F}|\mathbf{X})$ , I only need to weight each  $\mathbf{X}^*$  by  $\pi_{\theta}(\theta(\mathbf{X}^*))$ , the probability of the value of  $\theta$  measured in the bootstrap sample given its distribution from the external evidence. Being able to sample from  $P(\mathbf{F}|\mathbf{X})$ , I propose two sampling schemes for generating samples from  $P(\mathbf{F}|\mathbf{X}, \pi_{\theta})$ : the vetted bootstrap, which is a form of rejection sampling <sup>101</sup>, and the weighted bootstrap, which is a form of importance sampling <sup>102</sup>. In both schemes, one desires to sample from a probability distribution with density function f(.), but is only able to generate random samples from the density function g(.), often called the instrumental or proposal distribution. Here,  $g(\mathbf{F}) = P(\mathbf{F}|\mathbf{X})$  and  $f(\mathbf{F}) \propto \pi_{\theta}(\theta(\mathbf{F})).g(\mathbf{F})$ . The idea is to 'weight' each sample from g(.) by a weight proportional to f(.)/g(.), which in this case is proportional to  $\pi_{\theta}(\theta(\mathbf{F}))$ .

#### 2.7 Algorithm: the vetted bootstrap

In this scheme, each  $\mathbf{X}^*$ , the entire bootstrap sample of RCT data, is accepted by a probability that is proportional to  $\pi_{\theta}(\theta(\mathbf{X}^*))$  (hence 'vetting' the bootstrap). To change weights to valid probabilities, I need only to multiply them by a constant to make sure that the weights will remain in the interval [0,1]. The optimal way to do so is to divide the weights over the maximum possible weight, i.e.,  $max_{\theta} \pi_{\theta}(\theta)$  (any constant larger than this value will be valid but larger values result in wasteful rejection of bootstrap samples). This results in the following algorithm:

1. Calculate  $\omega_{max} = max_{\theta} \pi_{\theta}(\theta)$  as the scaling factor for weights from the distribution of the external evidence.

2. For *q*=1, ..., *M*, where *M* is the number of simulations:

3. Generate  $X^*$ , a (Bayesian) bootstrap sample of X, with bootstrapping performed separately within each arm of the trial.

4. Calculate the statistics  $\theta^* = \theta(X^*)$  in this sample.

5. Calculate  $\omega = \pi_{\theta}(\theta^*)$ , the likelihood of the statistics according to external evidence.

6. Randomly draw *u* from a uniform distribution in the interval [0,1]. If *u* > *ω/ω<sub>max</sub>*, then ignore the bootstrap sample and jump to step 3.
7. Calculate φ<sup>\*</sup> = φ(X<sup>\*</sup>), the cost and effectiveness outcomes for each arm from the bootstrap sample.

8. Store the value of  $\varphi^*$  and jump to 2.

#### 2.8 Weighted bootstrap as an alternative to the vetted bootstrap

As an alternative to probabilistically accepting or rejecting bootstrap samples based on the weight  $\omega$ , one can assign the weights directly to each bootstrap sample and incorporate such weights in all downstream calculations. This importance sampling scheme will lead to the same results as the acceptance-rejection method employed in the vetted bootstrap <sup>98</sup>. This mechanism is especially helpful for the situations in which  $\omega_{max}$  cannot be determined. However, unlike the vetted bootstrap, which can generate a desired number of independent, identically distributed draws from the posterior distribution of cost-effectiveness outcomes, the unequal weights assigned to bootstrap samples in the weighted bootstrap scheme affect subsequent calculations and pose difficulties in presenting results graphically in the cost-effectiveness plane.

#### 2.9 Case studies

## 2.9.1 Analogy between the vetted bootstrap and parametric analysis of bivariate normal data

First, I use a highly stylized example that allows for an analytical solution and shows how the vetted bootstrap operates. Imagine we measure the paired values of  $\varphi$  and  $\theta$  in n=100 individuals and that we are interested in inference on  $E(\varphi)$  based on this sample and external information on  $\theta$ . We observe that  $\theta$  has a mean 1 and variance of 10 in the sample. Imagine we also observe the relation between  $\varphi$  and  $\theta$  having a linear form  $\varphi = 2\theta + 1 + e$ , where the error term e is assumed homoscedastic and normally distributed, with zero mean and variance 5. This yields the likelihood function for  $E(\varphi)$  as  $Normal \left(2 \times 1 + 1 = 3, \frac{4 \times 10 + 5}{n} = 0.45\right)$ , with the first and second parameters being mean and variance, respectively.

Now imagine there is external evidence on  $E(\theta)$  in the target population which is *Normal*(0.5, 0.1). Because  $\varphi$  is dependent on  $\theta$ , this external evidence will carry information about  $\varphi$ . To update the distribution of  $E(\varphi)$ , I first use parametric Bayesian inference on normal distribution (approximating the *t* distribution with normal, equal to assuming that the variance is known) to update the marginal distribution of  $E(\theta)$ <sup>103(p46)</sup>: the posterior distribution

of E(
$$\theta$$
) is  $Normal(\frac{1\times0.1+0.5\times\frac{10}{n}}{0.1+\frac{10}{n}} = 0.75, \frac{1}{\frac{1}{0.1}+\frac{10}{n}} = 0.05)$ . From this, the posterior distribution of E( $\varphi$ ) can be obtained as  $Normal(2\times0.75+1=2.5, 4\times0.05+5/n=0.25)$ . The prior distribution of E( $\theta$ ), the likelihood function of E( $\varphi$ ) based on the data, and the posterior distribution of E( $\varphi$ ) are shown in **Figure 2-1**.

Figure 2-1: An illustrative example of the parametric Bayesian inference vs. the vetted bootstrap approach.



Data are generated from 100 pairs of samples from  $\theta$  and  $\varphi$ . See text for details. (R code available in Appendix A.1)

To implement the vetted bootstrap to derive the distribution of  $E(\varphi)$ , I first generate bootstrap sets from the observed pairs of  $\theta$  and  $\varphi$  and calculate the mean of  $\theta$  and  $\varphi$  in each set (the conventional [approximate Bayesian] bootstrap is chosen for this example). The histogram of the mean of  $\varphi$  in the bootstrap sets (white bars in *Figure 2-1*), expectedly, matches the parametric likelihood function of  $E(\varphi)$ . Next, I calculate weights  $\omega$  for each bootstrap set given the prior normal distribution of  $\theta$  as  $\omega = \phi(\frac{\overline{\theta^*} - 0.5}{\sqrt{0.1}})$  with  $\phi$  being the normal probability density function and  $\overline{\theta^*}$  the mean of  $\theta$  in the bootstrap sample. Finally, I will accept each bootstrap set with a probability proportional to  $\omega/\phi(0)$ , with  $\phi(0) = 0.3989$  being the maximum possible value of weights. The histogram of the accepted bootstraps (gray bars, drawn under the axis for clarity in **Figure 2-1**) will resemble the posterior distribution of the expected value of *b* derived using parametric inference (the R code for this stylized example is available in **Appendix A.1**).

Note that the parametric inference in this example provides a valid answer only because the relation between  $\theta$  and  $\varphi$  was simple and data were generated to match the parametric assumptions. In reality, the nature of such relations can be very complex, especially when the external evidence and outcomes of interest are multidimensional. On the other hand, in drawing inference on  $\varphi$  using the vetted bootstrap, I did not rely on the assumed relation between  $\varphi$  and  $\theta$  and used the raw data to draw inference.

#### 2.9.2 A real-world RCT-based CEA

Here, I use data from the OPTIMAL trial to show the practical aspects of implementing the vetted bootstrap algorithm. I describe the original approach taken for the CEA of the RCT <sup>3</sup> and show that such steps can easily be modified to incorporate external information on treatment effect size. This case study is to demonstrate the operational aspects of implementing the algorithm and the exercise is undertaken only for pedagogical purposes.

Data on both resource use and quality of life were collected during the trial, which was used to carry out the CEA. The outcomes of the CEA were the incremental costs per exacerbation avoided and incremental costs per QALY gained. For the original analysis, we partitioned the time series data on resource use and exacerbations to 13 intervals and used a nested sequence

of bootstrapping, imputation of missing and censored values, and linear regression for adjusting QALYs for baseline utility. Since individual level resource use and effectiveness outcomes were available, the CEA was based on the direct inference on their distribution. No external information was incorporated into the analysis in the original CEA. The vector of data for an individual patient ( $X_{ij}$  in the notation developed earlier) used in the CEA consists of 13 cost values collected in each period, 13 values indicating the number of exacerbations in each period, 5 utility values measured at baseline and follow-up visits, and the baseline covariates used to adjust the QALY.

#### 2.9.2.1 Incorporating external evidence

To my knowledge, there is currently no other RCT published that provides evidence on the effects size of treatments used in the OPTIMAL study, but there are RCTs that have used drugs within the same classes. I used the results of a meta-analysis comparing exacerbation rates between COPD patients receiving tiotropium plus formoterol (the same class as salmeterol) versus tiotropium alone as the source of external evidence for the effect size between TS and TP <sup>104</sup>. For evidence on the effect size of TFS versus TP, I chose the results of a 12-week RCT on comparing budesonide (the same class as fluticasone) and formoterol added to tiotropium versus tiotropium alone in COPD patients <sup>105</sup>. The evidence was parameterized on the *log (OR)* scale. Because the external evidence was synthesized from the studies that used medications from a similar class yet different from those in the OPTIMAL study, I decided *a priori* that such historical evidence should be discounted by inflating the variance of *log (OR)* by 50% <sup>28(p151)</sup>. This reflects my desire to use historical data but to avoid the assumption that these data were

obtained from the same population that received the study drugs <sup>106</sup>. Because the distribution of external evidence was modeled as normal, this can also be seen as constructing a *power prior* (with  $\alpha$ =2/3) that is often used for discounting historical data <sup>107</sup>.

The meta-analysis pooled five studies and estimated an OR of 0.93 (95% CI 0.45 – 1.93). Because this meta-analysis was based on a random-effects model, the most relevant estimate for the effect size for the OPTIMAL trial is the predictive distribution of the log(OR) of treatment in a new RCT <sup>108</sup>. This quantity has an approximate normal distribution with mean equal to the pooled estimate of the effect size and a variance that is the sum of the estimated between-trial variance (0.34) and the variance of the pooled estimate <sup>28(p150)</sup>. This results in a normal prior for the *log(OR)* of TS vs. TP for experiencing at least one exacerbation (denoted by  $\theta_{TS/TP}$ ) with mean -0.073 and variance 1.03, equal to OR=0.93 (95% CI 0.13 – 6.84). The previous RCT reported a risk ratio of 0.38 (95% CI 0.25–0.57) for triple therapy versus monotherapy, which after discounting corresponds to a log(OR) of TFS vs. TP for experiencing at least one exacerbation (denoted by  $\theta_{TFS/TP}$ ) having a normal distribution with mean -0.97 and variance 0.09, equal to OR= 0.38 (95% CI 0.21 – 0.68). I note that the external evidence on TS vs. TP is relatively weak with a point estimate indicating near-equivalence accompanied by a large variance. The external evidence on TFS vs. TP, on the other hand, favors TFS more strongly than what the OPTIMAL results indicate.

Putting all these together, the likelihood function for the external evidence becomes

$$\pi_{\theta}(\theta_{TS/TP}, \theta_{TFS/TP}) \propto e^{-\frac{(\theta_{TS/TP}+0.073)^2}{2.06} - \frac{(\theta_{TFS/TP}+0.97)^2}{0.18}},$$

the product of two normal likelihoods representing our knowledge on the treatment effect. Note that  $\pi_{\theta}(.)$  is already scaled to have a maximum of 1, hence all weights generated from  $\pi_{\theta}(.)$  are valid probabilities without need for further manipulation.

The original algorithm for the CEA can now be updated to incorporate the external evidence as follows:

- 1. For q=1,2,...,10000.
- 2. Generate X\*, a (Bayesian) bootstrap sample within each of the three arms of the RCT.
- 3. Impute the missing values in costs, utilities, and exacerbations in X\*.
- 4. Calculate  $\theta_{TS/TP}^*$  and  $\theta_{TFS/TP}^*$  from the bootstrap sample, the *log(OR)* of experiencing at least one exacerbation during the follow-up period for TS vs. TP and TFS vs. TP, respectively.
- 5. Calculate  $\omega = \pi_{\theta} (\theta^*_{TS/TP}, \theta^*_{TFS/TP})$  using the distribution constructed for the external evidence.
- 6. Randomly draw u from a uniform distribution in the interval [0,1]. If  $u > \omega$ , then ignore the bootstrap sample and jump to step 2.
- 7. Calculate mean costs, number of exacerbations, and adjusted QALYs for each arm from  $\mathbf{X}^*$ .
- Store the average values for costs, number of exacerbations, and adjusted QALYs; then jump to 1.

The above algorithm was run separately using the Bayesian and approximate Bayesian bootstraps. In order to study the impact of applying the external information on TS/TP and TFS/TP separately, I repeated the analysis three times, once with evidence on TS/TP, once with evidence on TFS/TP, and once with full evidence (as described in the above algorithm). I also used the same data generated using this algorithm (including all the rejected and accepted bootstraps) for calculating the outcomes using the weighted bootstrap method, by also recording the value of the weight generated for each bootstrap.

#### 2.9.2.2 Results

*Table 2-1* presents the mean costs, mean number of exacerbations, and mean QALYs for each of the three arms of the trial. Each panel shows the results with and without incorporating the external evidence, with the former obtained through four different permutations of the bootstrap and weighting methods: the vetted Bayesian bootstrap (VB), the vetted approximate Bayesian bootstrap (VAB), the weighted Bayesian bootstrap (WB), and the weighted approximate Bayesian bootstrap (WAB). As this figure demonstrates, the four different methods for the most part generate very similar results. The incorporation of external evidence on TS/TP does not have a noticeable effect on the outcome, an expected finding given the weak prior on the effect size of TS vs. TP. The incorporation of evidence on TFS/TP, on the other hand, shifts the outcomes of the TFS arm in the favorable direction (lower costs, lower exacerbation rate, and higher QALYs), and shifts the outcomes of the TP arm in the opposite direction. This is an expected finding given the strong prior in favor of TFS for the effect size of TFS vs. TP.

	ТР				TS			TFS				
No external information												
Costs	2,660.2				2,821.0				4,071.3			
Exacerbations	1.578				1.696				1.345			
QALY	0.7071				0.7019				0.7129			
	VB	VAB	WB	WAB	VB	VAB	WB	WAB	VB	VAB	WB	WAB
External information on TS vs. TP												
Costs	2,662.3	2,654.3	2,663.8	2,654.3	2,818.2	2,822.1	2,818.1	2,820.5	4,072.7	4,083.0	4,071.3	4,082.2
Exacerbations	1.580	1.580	1.581	1.580	1.693	1.695	1.693	1.695	1.345	1.346	1.345	1.346
QALY	0.7070	0.7068	0.7070	0.7068	0.7020	0.7017	0.7020	0.7017	0.7128	0.7125	0.7129	0.7125
External information on TFS vs. TP												
Costs	2,753.6	2,722.7	2,737.8	2,736.8	2,820.9	2,817.8	2,824.2	2,827.3	3,960.2	4,012.4	3,970.7	3,996.8
Exacerbations	1.645	1.638	1.643	1.643	1.694	1.697	1.696	1.702	1.283	1.289	1.284	1.284
QALY	0.7058	0.7051	0.7056	0.7054	0.7026	0.7022	0.7022	0.7020	0.7151	0.7142	0.7148	0.7145
External information on TS vs. TP & TFS vs. TP												
Costs	2,759.4	2,726.1	2,738.8	2,738.1	2,820.7	2,816.3	2,823.2	2,826.2	3,961.4	4,011.5	3,971.5	3,997.4
Exacerbations	1.646	1.638	1.644	1.643	1.693	1.696	1.695	1.700	1.283	1.289	1.285	1.284
QALY	0.7058	0.7051	0.7056	0.7054	0.7027	0.7022	0.7022	0.7020	0.7151	0.7143	0.7148	0.7145

#### Table 2-1: Outcomes of the OPTIMAL CEA without and with the incorporation of external evidence

A total of 67,786 and 67,618 runs were required to obtain 10,000 accepted bootstraps using the Bayesian and approximate Bayesian techniques, respectively

*TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol. VBB: vetted Bayesian bootstrap, VAB: vetted approximate Bayesian bootstrap, WB: Weighted Bayesian, WAB: Weighted approximate Bayesian* 

Compared to the cost and effectiveness outcomes, the impact of incorporating external evidence on the ICER, presented in *Figure 2-2*, is more noticeable. This figure shows the ICERs for exacerbation avoided and for QALY gain for TS vs. TP and TFS vs. TP. Incorporating external evidence on TS/TP has virtually no impact on the ICERs. Likewise, incorporating external evidence on both TFS/TP and TS/TP gives similar results with incorporating external evidence on TFS/TP only, another demonstration of the little information provided by the external evidence on TS/TP. For TFS vs. TP comparison, both ICERs decrease by 40% after the incorporation of the external evidence. This is a change in the expected direction given that the external evidence strongly favours the TFS strategy in terms of the lower number of exacerbation, which intuitively translates into better quality of life outcome.





*Figure 2-3* presents the cost-effectiveness plane (CE-plane). I have provided the CE-plane for TFS vs. TP comparison (the TS vs. TP CE-plane is not noticeably affected by the incorporation of external evidence), and for two scenarios: one without incorporation of external evidence, and one with the incorporation of external evidence on both TS/TP and TFS/TP (the incorporation of

external evidence on TS/TP results in the CE-plane that is not distinguishable from the noninformative scenario). Although the overall shape of the CE-planes remain the same, the incorporation of external evidence shifts the cloud of data mainly to the right, corresponding mainly to an improvement in the effectiveness outcome.

### Figure 2-3: Cost-effectiveness plane (CE-plane) for TFS vs. TP without and with incorporation of external evidence



QALY: quality-adjusted life years

Finally, the cost effectiveness acceptability curve (CEAC) is provided in *Figure 2-4.* Again, I have focused on incorporating both effect sizes in the informative analysis, and have generated the curves from the output of the vetted Bayesian bootstrap method (the approximate Bayesian method gives virtually the same curve). The incorporation of the external evidence increased the probability of cost-effectiveness for TFS, especially with higher willingness-to-pay values. Without the incorporation of the external evidence, the probability of TFS being the cost-effective intervention surpassed that of TP at WTP values greater than \$200,000/QALY, while the incorporation of the external evidence moves this threshold to \$130,000/QALY. Changes in the CEAC were modest for the TS arm.





Square: TP, Diamond: TS, Triangle: TFS. Filled shapes: without external evidence. Empty shapes: with external evidence. QALY: quality-adjusted life year TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol.

A total of 67,786 and 67,618 runs were required to obtain 10,000 accepted bootstraps using the Bayesian and approximate Bayesian techniques, respectively. The incorporation of the external evidence on the TS/TP effect size resulted in a rejection of 5% of samples. Incorporating both sources of evidence, on the other hand, led to the rejection of 85% of the bootstraps. This relatively high rate of rejection can be interpreted as a manifestation of the dissimilarity of the prior (external information) and likelihood (RCT data), with regard to the effect size of TFS vs. TP.

#### 2.10 Discussion

In the health economics literature, when an economic evaluation is conducted alongside a single RCT, the practice of evidence synthesis is not currently an integral part of the analysis. This is partly because evidence synthesis can result in problem-specific and complex statistical models. There are simple methods for borrowing inference from external sources that result in straightforward calculations for the posterior distributions of costs and effectiveness <sup>38,49</sup>, but these methods are able to incorporate external evidence only with certain parametric assumptions and only if the external evidence is defined on costs and/or effectiveness outcomes. The vetted and weighted bootstrap methods, which provide the ability to incorporate evidence on any aspect of intervention within a semi-parametric framework, are therefore a practical, relevant extension of the current methods of evidence synthesis for RCT-based CEAs. Rejection sampling and importance sampling are popular methods in which sampling from a 'difficult' distribution is replaced by sampling from a proposal (or instrumental) distribution <sup>109</sup>. Here sampling from  $P(\mathbf{F}|\mathbf{X}, \pi_{\theta})$  is performed via the proposal

distribution  $P(\mathbf{F}|\mathbf{X})$ , and the latter can easily be sampled through (Bayesian) bootstrapping. This form of sampling has seldom been applied to a bootstrap. This uncommon mixture was employed here because of the need for evidence synthesis in CEAs and the popularity of bootstrap in RCT-based CEA.

In synthesizing evidence for RCT-based CEAs, a carefully crafted parametric model with comprehensive analysis of model convergence and sensitivity of results to parametric assumptions has some strengths over an acceptance/rejection approach, including the higher computational efficiency of MCMC or likelihood-based methods and the ability to synthesize and propagate all evidence in a single analytical framework <sup>26,110</sup>. Nevertheless, practical issues make the vetted bootstrap a competitive option. The vetted bootstrap is an intuitive and easy extension of the popular bootstrap method of RCT-based CEAs; it does not require specialist software and in-depth content expertise for its implementation. In addition to such practical advantages, this method connects the auxiliary parameters to the cost and effectiveness outcomes without an explicit model. Instead, it uses RCT data to connect related parameters (e.g., treatment effect size and quality of life), maintaining the correlation structure between cost and effectiveness outcomes and intermediary parameters.

A particularly useful application of the vetted bootstrap approach is to incorporate the use of so-called objective or structural priors in RCT-based CEAs <sup>40</sup>. These are priors expressing our confidence about the structure of the data coming from logical expectations whilst having relatively much weaker prior knowledge of the numerical values of the parameters <sup>40</sup>. For example, in a RCT-based CEA of chemotherapy versus combination of chemotherapy plus

radiotherapy for a malignancy, one can, based on basic laws of biology, expect that the combination therapy will not result in the higher recurrence rate of cancer, at least in the short term. Then, one can discard (or assign a very low weight to) the bootstrap sets in which the incidence of cancer in the combination therapy arm is higher.

This chapter provides a conceptual framework. Further research into theory, as well as practical issues in using this method in realistic situations, should follow. If the prior distribution and sample distribution of the external evidence substantially differ from each other (i.e., the prior and data are in conflict), then the weights assigned to the majority of bootstrap samples will be low, resulting in the rejection of many bootstraps. This was the case in our example study, as incorporating external evidence that assumed an effect size that conflict with the observed effect size resulted in the rejection of 85% of bootstraps. Similar situations may arise when there are too many parameters with external evidence, in which each bootstrap sample needs to be vetted by several distributions, resulting in a low yield for the whole process. Finally, how to perform the bootstrap and how to weigh the bootstrap sample against the external evidence might not be straightforward in some situations, such as cluster or cross-over RCTs.

The prior information for the vetted bootstrap was constructed by multiplying the Rubin's prior on the population distribution ( $\pi_{\mathbf{F}}$ ) with the informative prior on the auxiliary parameters ( $\pi_{\theta}$ ). An obvious consequence of this approach is that the marginal distribution of  $\theta$  is not exactly the same as the intended distribution,  $\pi_{\theta}$ , because  $\pi_{\mathbf{F}}$  carries information about  $\theta$ . Nevertheless, this should generate no misgivings. First,  $\pi_{\mathbf{F}}$  is a vague prior used merely to justify bootstrapping in a Bayesian context and does not carry strong information on the population

distribution (and hence on any parameter estimated from it), as evident by the proven performance and comparability of the bootstrap method for inference with the frequentist and non-informative Bayesian methods. Second, in a Bayesian CEA without the incorporation of external evidence one should ideally assign a flat (or otherwise a conventionally accepted noninformative) prior distribution to the cost and effectiveness outcomes, yet the use of  $\pi_{\rm F}$ translates into using an unknown (but not necessarily flat) prior on such outcomes. As such, using the bootstrap for inference implies accepting such a deviation from a strictly noninformative analysis (one can see this as the cost that comes with the convenience of using resampling methods for inference). From this perspective an informative prior of the form  $\pi_{\theta} \times \pi_{\rm F}$  is quite logical in that it incorporates the external evidence 'incrementally' on top of the non-informative scenario, therefore the difference in the results with and without the external evidence shows the true impact of our external knowledge. However, it might be possible to construct a prior on F that both enables bootstrapping and has a marginal distribution of  $\theta$  that matches the distribution from the external evidence. Such a prior will be more efficient as there will be no need for rejection sampling. However, specifying such information in the context of a Dirichlet process prior will lead to posterior distributions that put non-zero probability on unobserved values, therefore invalidating resampling methods for inference <sup>111</sup>.

Faced with the soaring costs of RCTs and the requirement by many decision-making bodies for formal economic evaluation of emerging health technologies, trialists and health economists are hard-pressed to generate as much relevant information for policy makers as possible. As such, and despite criticisms, it appears that RCT-based CEAs are here to stay. The incorporation

of external evidence should improve CEA validity and help optimize adoption decisions. The vetted bootstrap, aside from its theoretical contribution, provides the large camp of analysts using bootstrap for RCT-based CEAs with a statistically sound, easily implementable tool for such purpose.

# Chapter 3. Pulling EVI by its bootstrap: a framework for non-parametric expected value of information analysis of RCT-based CEAs

#### 3.1 Background

Methods have been developed for EVI analysis of model-based <sup>72,112–114</sup> and RCT-based <sup>70,79–81,84</sup> CEAs. For the latter group, methods are almost entirely based on the normal approximation of the distribution of the expected (incremental) net benefit. This turns the output of the trial into a multivariate normal distribution of net benefits for which the algorithms for EVI analysis for model-based CEAs can readily be applied (such a multivariate distribution can indeed be considered a simple decision-analytic model). For two-arm RCTs, equations based on such normal approximation can be expressed in closed form <sup>71</sup>. If the CEA and quantification of uncertainty is based on parametric inference on the net benefit scale <sup>55</sup>, this method of EVI analysis can be considered conjugate to parametric CEA <sup>61</sup>.

In this chapter I propose a method for the calculation of the EVPI and EVSI in the context of RCT-based CEAs based on the bootstrap. It has already been pointed out that the data generated by bootstrapping can be used to calculate the EVPI <sup>86,92</sup>. However, as mentioned by others, it is not clear if and how such methods for the CEA can be extended to calculate the EVSI <sup>86</sup>. In this chapter I propose a general, non-parametric method that can be considered the extension of the bootstrap method of the CEA. This method allows for calculating the maximum expected return of investment for a future RCT with similar design as the current RCT (expected value of perfect information, EVPI), or the expected return of investment for a future RCT with similar design.

the assumption of normality of net benefits, and has the capability of incorporating practical issues that arise in the analysis, such as missing values in the current and future RCTs and unbalanced distribution of covariates. Compared to the contemporary EVI methods for RCTbased CEAs, the incorporation of such practical aspects requires more elaborate statistical programming and is computationally more demanding. Nevertheless, as the case study shows, incorporating such aspects into the analysis may lead to results that are different from those obtained by simpler methods, with potentially significant impact on the design of the future RCTs.

The present chapter builds directly on the framework developed in the previous chapter, namely the Bayesian interpretation of the bootstrap. The remaining sections of this chapter are structured as follows: after describing the context, I show the similarity of the results between the two-level method and the normal approximation method using a stylized example. Next, I use the OPTIMAL trial data to perform EVI analysis under two scenarios: a simple scenario based on patient-level net benefits from the RCT, aimed at demonstrating the analogy between the parametric and non-parametric methods of EVI calculation, and a detailed scenario that tackles issues around EVI analysis, such as the appropriate handling of missing data and the need for covariate adjustment. I conclude by listing some other scenarios in which the two-level resampling methodology can remain valid, and elaborate on its strengths and limitations.

#### 3.2 Context

The context is similar to the general context outlined in the Introduction. In addition, only direct medical costs are considered in this analysis, and the decision maker has no prior
information about any aspects of the competing interventions. Having the raw data from a current RCT at hand, the decision maker is interested in estimating the value of another RCT with a similar or nested design; that is, with the same (or a subset of) interventions and also for the same follow-up period. As discussed in the original study, because the interventions considered in the OPTIMAL trial generate costs and benefits as long as medications are being used by patients (unlike, say, a surgical intervention that generates a one-time large cost and continuous benefits), there is no strong reason to believe a longer time horizon would change their cost-effectiveness; as such, I assume the future RCT and its CEA analysis will have one year follow-up. Given this, no discounting was applied to the outcomes of the analysis. All calculations were performed in the statistical programming environment R version 2.10.01<sup>115</sup>.

# 3.3 EVPI and EVSI definitions in a non-parametric framework

The equations for EVI measures (see the Introduction) are heavily based on 'parameters' that constitute the evidence. In the bootstrap method of RCT-based CEAs, however, there is no obvious parameter to begin with. Nevertheless, an analogy can be established by treating the distributions that have generated data for each arm of the RCT (*population distributions*) as a random quantity. Like the notations in the previous chapter, let **X** be the data of the RCT at hand. The individual-level data observed within each arm of a trial can be considered random draws from the *population distribution* of the respective arm. Let **F** be the *population distribution distribution* that generates data for a random individual from the population depending on which treatment that individual receives. One can think of such population distribution as the random quantity of interest for which one has only partial

information in the form of the data of the current RCT. I can write the formulas for EVPI and EVSI as

$$EVPI(\mathbf{X}) = E_{\mathbf{F}|\mathbf{X}}\{\max_{i}[NB_{i}(\mathbf{F})]\} - \max_{i}\{E_{\mathbf{F}|\mathbf{X}}[NB_{i}(\mathbf{F})]\}$$

$$EVSI(\mathbf{X}) = \mathbb{E}_{\mathbf{X}^{**}|\mathbf{X}}\left[\max_{i}\{\mathbb{E}_{\mathbf{F}|\{\mathbf{X},\mathbf{X}^{**}\}}[NB_{i}(\mathbf{F})]\}\right] - \max_{i}\{\mathbb{E}_{\mathbf{F}|\mathbf{X}}[NB_{i}(\mathbf{F})]\},$$

where  $NB_i(.)$  is the estimator of the population value of NB if the population receives the  $i^{th}$  treatment. The term  $\mathbf{X}^{**}$  is the data for the future RCT. These equations parallel the equations for EVI calculations for decision-analytic models, with the population distribution  $\mathbf{F}$  replacing model parameters <sup>67,68,81</sup>. If I can sample from  $P(\mathbf{F}|\mathbf{X})$ , then I can calculate  $\mathbf{E}_{\mathbf{F}|\mathbf{X}}$  and hence the EVPI using Monte Carlo techniques. Likewise, if I can generate samples from  $P(\mathbf{X}^{**}|\mathbf{X})$  and  $P(\mathbf{F}|\{\mathbf{X}^{**},\mathbf{X}\})$ , EVSI calculation can be performed through Monte Carlo simulation (estimating  $\mathbf{E}_{\mathbf{X}^{**}|\mathbf{X}}$  and  $\mathbf{E}_{\mathbf{F}|\{\mathbf{X}^{**}\}}$  by Monte Carlo). This is analogous to Monte Carlo approaches for EVI calculations in parametric models.

# 3.4 Bootstrap as a method of sampling from the "distribution of a distribution"

The critical question of how to generate samples from  $P(\mathbf{F}|\mathbf{X})$ ,  $P(\mathbf{X}^{**}|\mathbf{X})$ , and  $P(\mathbf{F}|\{\mathbf{X}^{**}, \mathbf{X}\})$  is answered by the Bayesian viewpoint to the bootstrap, described in the previous chapter. Again, I denote  $\mathbf{X}^*$  to be a (Bayesian or approximate Bayesian) bootstrap sample of  $\mathbf{X}$ . The empirical distribution associated with  $\mathbf{X}^*$  can be considered as a random draw from  $P(\mathbf{F}|\mathbf{X})$ . Having samples from the population distribution **F** at hand, a random draw from the future trial data, **X**<sup>\*\*</sup>, can be obtained by sampling from **F**, therefore  $P(\mathbf{X}^{**}|\mathbf{X})$  can be generated by twolevel resampling; the first level (which can be Bayesian or approximate Bayesian bootstrap) generates sample from  $P(\mathbf{F}|\mathbf{X})$ , and the second level (which is sampling with replacement with sample size equal to that of the future RCT) produces sample from  $P(\mathbf{X}^{**}|\mathbf{F})$ . Finally, generating samples from  $P(\mathbf{F}|\{\mathbf{X}^{**}, \mathbf{X}\})$  really depends on the analyst's plan for the CEA once the future RCT data are available. Given the similar design of the current and future RCTs, one option, which I will adopt here, is to simply merge the datasets of the current and future RCTs, in which case samples from  $P(\mathbf{F}|\{\mathbf{X}^{**}, \mathbf{X}\})$  can be obtained by the (Bayesian) bootstrap of the merged set. Other options include relying solely on the future RCT, or using a random-effects model for pooling data from current and future RCTs <sup>116</sup>.

# 3.5 Case studies

#### 3.5.1 A stylized example

The purpose of this stylized example is to show the analogy between the two-level resampling method and the normal approximation method for EVSI calculation. I use the summary data of the Early Cephalic Version trial used by Willan et al. <sup>71</sup>. In this RCT, 232 patients were equally randomized between the treatment and control groups. The outcome of interest (non-caesarean delivery) occurred in 41 patients in the treatment arm and 33 patients in the control arm. Assuming the decision maker wants to maximize the probability of successful outcome in the target population (regardless of any other aspect of treatment), and assuming a willingness-to-pay (WTP) value of \$1000 to achieve one favourable outcome, the current RCT provides an

estimate of the net benefit of  $(41/116 - 33/116) \times 1000 = 69.0$  with a variance of  $[41/116 \times (1 - 41/116)/116 + 33/116 \times (1 - 33/116)/116] \times 1000^2 = 3724.8$ . Now imagine we are interested in conducting another RCT with the similar design. The EVSI of the future RCT can be calculated in closed form as described by Willan et al. <sup>71</sup>. Note that the data has a very simple structure: each patient within each arm can be represented by a single binary variable indicating success or failure. A RCT therefore can be fully specified by the number of patients in each arm, and the proportion within each arm who had a successful outcome. The *population distribution* for each arm can be fully specified by a single value: the probability of successful outcome. This simple structure allows direct modeling of the outcomes in the bootstrap samples instead of actually performing the bootstraps. The sequence of sampling for EVSI calculations is provided in *Table 3-1*.

In order to calculate the EVSI non-parametrically using the conventional and Bayesian bootstraps, I should first obtain a bootstrap sample separately within each arm of the trial. Imagine we have *n* patients of which *m* had successful outcome. The number of successful outcomes in a conventional (approximate Bayesian) bootstrap sample obtained from this sample is a random variable with probability distribution binomial(n, m/n). The probability of success ( $p_s$ ) in the population is therefore a draw from this distribution, divided by *n*. Likewise, for the Bayesian bootstrap, the probability of success in the population if they receive the treatment is the sum of the weights assigned to *m* components of the *n* -variate Dirichlet variable with scale parameter of 1, which according to the relationship between the Dirichlet and Beta distributions <sup>117(pp189-90)</sup>, is a random variable with distribution *beta*(*m*, *n* - *m*). Once the population distribution is specified in this way, the number of patients in each arm of the

future RCT who experience the outcome (denoted by  $m_F$ ) can be modeled as

 $m_F \sim binomial(n_F, p_S)$ , where  $n_F$  is the sample size per arm of the future RCT. The updated estimate of the probability of success for treatment after observing the future trial is  $(m + m_F)/(n + n_F)$ . Once these calculations are performed for both arms, the decision can be revised. A single loop of non-parametric EVSI calculation can therefore be described as in **Table 3-1**.

Step	Bayesian	Approximate Bayesian <sup>*</sup>	Comments		
1	p <sub>st</sub> ∼beta(41,116 – 14)	p <sub>st</sub> ~binomial(116,41/116)/116	Random draw from the population distribution of treatment arm		
2	<i>p<sub>Sc</sub>∼beta</i> (33,116 – 33)	p <sub>sc</sub> ~ binomial(116,33/116)/116	Random draw from the population distribution of control arm		
3	$m_{Ft}{\sim}$ bino	Number of patients with successful outcome in the treatment arm of the future RCT			
4	$m_{Fc}{\sim}bino$	Number of patients with successful outcome in the control arm of the future RCT			
5	$INB = [(41 + m_{Ft})/(116 + n_F)]$	Expected incremental net benefit after merging the current and future RCTs			
6	EVSI = n	EVSI estimate for the current loop			
See Appendix A.2 for the K code that generates the results					

Table 3-1: Non-parametric EVSI calculation for the stylized example

<sup>&</sup>lt;sup>\*</sup> The abuse of notation in  $p_{St}$  and  $p_{Sc}$  is for brevity, and indicates randomly drawing from the binomial distribution and dividing the value by the denominator.

The EVPI using the parametric method, non-parametric method based on the Bayesian bootstrap, and non-parametric method based on the approximate Bayesian bootstrap are 3.945, 3.912, and 3.944, respectively. *Figure 3-1* presents the results of EVSI calculations for a range of sample sizes for the future RCT, averaged over 1,000,000 loops, as well as the results obtained using the normal approximation method using the equations provided by Willan et al. <sup>71</sup>(see *Appendix A.2* for the R code that generates the results). As can be seen in this figure, all three methods generate very similar EVSIs. While there is no doubt the closed-form equation provides a much faster method of EVSI calculation and does not suffer from Monte Carlo sampling error, the two-level resampling method has some potential advantages: it can readily be extended to RCTs with more than two interventions, and can be used flexibly in order to accommodate realistic aspects of RCTs, as will be described in the next section. Figure 3-1: Parametric vs. non-parametric EVSI calculations for the stylized example



Non-parametric calculations are average of 1,000,000 iterations. (R code available in **Appendix A.2**)

EVSI: expected value of sample information

#### 3.5.2 Example from the OPTIMAL trial

I use the individual-level data of the OPTIMAL trial in order to calculate the EVPI and EVSI for a future RCT with a similar or nested design (i.e., having a subset of intervention arms). I assume the future RCT will attempt to randomize patients with equal probability into each intervention. I perform the calculations using two methods: a simple scenario that is based on working with individual-level NBs and is a non-parametric analogous of the normal approximation methods, and a detailed scenario that demonstrates the flexibility of EVI analysis using patient-level data with regard to incorporating some realistic aspects of designing and conducting RCTs.

#### 3.5.2.1 Simple scenario: EVPI/EVSI calculation using individual net benefits

In the simple scenario, I work directly on individual-level NBs, assuming there is no uncertainty in their elicitation from each patient (that is, **X** is the set of observed NBs in the trial, and **F** is the probability distribution that generates NB for a random individual from the population depending on which treatment that individual receives). Because the simple scenario does not allow for the incorporation of uncertainties arising from missing value imputation, I replaced the missing cost and utility data with the sample average within the same time point and within the same arm of the trial. In this scenario, a sample from **F** can be obtained by bootstrapping NBs separately for each arm, and  $NB_i(.)$  simply calculates the sample mean of NBs for the *i*<sup>th</sup> arm.

Given that I am able to generate random draws from  $P(\mathbf{F}|\mathbf{X})$ ,  $P(\mathbf{X}^{**}|\mathbf{X})$ , and  $P(\mathbf{F}|\{\mathbf{X}^{**}, \mathbf{X}\})$ using the resampling techniques described earlier, I can repeat the same steps as the parametric EVPI and EVSI, working directly with the data instead of parameters. The algorithm can be followed as below:

- Generate X\*, a (Bayesian) bootstrap sample from X, separately for each trial arm. For the approximate Bayesian bootstrap, this is equal with sampling with replacement from the vector of NBs within each arm with the size equal to the sample size of the corresponding arm in the current trial.
- 2. Calculate the mean net benefit for each intervention using  $\mathbf{X}^*$  and pick the maximum mean NB.

- 3. Generate  $X^{**}$  through sampling with replacement from  $X^*$ , with the sample size within each arm equal to the sample size of the corresponding arm in the future RCT.
- 4. Merge the two datasets **X** and **X**<sup>\*\*</sup>.
- 5. Calculate the mean net benefit for each intervention from the merged data and pick the maximum mean NB.
- 6. Subtract from [2] the current maximum mean NB. This will be an estimate of EVPI. Subtract from [5] the current maximum mean NB; this will be an estimate of EVSI.

7. Repeat the cycle several times and average the results.

If the conventional (approximate Bayesian) bootstrap method is used for generating samples from **F**, the first two steps of this algorithm are similar to the bootstrap method of analysis of uncertainty in RCT-based CEAs  $^{52,118}$ .

# 3.5.2.2 Detailed scenario

The simple scenario is based on the assumption that conducting a clinical trial is akin to randomly drawing from the distribution of NBs from the population distribution within each arm of the trial. However, the reality is often more complicated, as individual NBs are estimated by following patients over time and collecting intermediate quantities such as health state utility weights and resource use records. Transforming such intermediate quantities to NBs often involves additional statistical inference. Such statistical inference is associated with uncertainty, and a comprehensive analysis has to take into account these aspects of the analysis. The steps that I explicitly consider for EVI analysis in the CEA of the OPTIMAL trial are

the imputation of missing values (costs and utilities) and the adjustment of calculated QALYs for the baseline estimate of health state utility values <sup>119</sup>. In addition, the future RCT will inevitably have missing data and some level of covariate imbalance, which will be taken into account in the detailed scenario. The analysis of the detailed scenario adopted here is based on the conventional bootstrap because first, it is consistent with the commonly used bootstrap method of RCT-CEA, and second, it can be performed by resampling the data (instead of assigning weights to data rows as in the Bayesian bootstrap), which facilitates subsequent analyses.

Unlike in the simple scenario, in the detailed scenario **X** is the set of all intermediate quantities that are used to calculate NBs. As such, this approach will generate an entire panel of data, instead of vectors of NBs, for the population distribution and for the future RCT. For the OPTIMAL study, this panel of data consists of baseline covariates (age, gender, and study site code), five estimates of health state utility values and 13 estimates of costs (accumulated in each 28-day period) per individual. In this context **F** is the multivariate distribution generating such a set of quantities for a random individual depending on the treatment the individual receives, and NB(.) represents the full process of cost-effectiveness analysis that calculates the expected value of NB from the intermediate quantities (including missing value imputation and covariate adjustment). A schematic flow-chart of the approach taken for the calculation of EVPI and EVSI for the future trial is provided in *Figure 3-2*.





*EVPI: expected value of perfect information, EVSI: expected value of sample information, RCT: randomized controlled trial, CEA: cost-effectiveness analysis* 

# 3.5.2.2.1 Missing values

In the OPTIMAL study, 10.0% of health resource data and 9.7% of health state utility values were missing in the original data. Calculating population values of NBs for each arm requires imputation of such missing values. In line with the non-parametric approach adopted in this analysis, I performed a non-parametric, two-level bootstrap imputation method <sup>65</sup>, stratified within the quintiles of propensity score for the presence of missing cost and QALYs. I used the same covariates in constructing the propensity score as in the original CEA, except the period-lagged estimates of costs and utilities which were excluded in order to reduce the computational demand.

#### 3.5.2.2.2 Covariate adjustment

In the CEA of OPTIMAL trial, as recommended elsewhere <sup>119</sup>, we adjusted the estimated QALY for each individual for the baseline utility, and used such adjusted difference in QALYs among arms for cost-effectiveness. I assume the future cost-effectiveness analysis of data will be based on the same approach.

The interpretation of the first level bootstrap as a random draw from the population distribution has an important bearing on the structure of the data. Because the interventions in the OPTIMAL trial compete with each other in the same population (e.g., patients with COPD in Canada), the population distribution is in fact the same across the three arms as long as factors not associated with treatment are considered; hence such factors should have the same distribution among the population distribution of each arm. While this argument is valid for all covariates not affected by treatment (e.g., gender and age), the focus here will be on the baseline utility as it is explicitly taken into account in the CEA and hence directly impacts the CEA results. When the bootstrap is used for calculating cost and effectiveness outcomes, or to calculate the EVPI, one common approach for adjusting for difference in baseline covariates is to use regression techniques. This approach is useful in such situations because the goal for the CEA and EVPI calculations is to calculate mean NB for each treatment arm. Here, however, the plan is to simulate the full panel of data for the future RCT in order to calculate the EVSI. Therefore adjusting mean NBs for baseline covariates using regression techniques is not enough.

Balance in the distribution of non-treatment-related variables in this context can conceivably be achieved in several ways. I propose two approaches, and perform simulations to compare their performance. The first approach is based on generating *k* sets of candidate bootstraps from **F** and picking the one in which baseline utilities across treatment arms are closest in distribution as measured by the Kolmogorov-Smirnov (K-S) test statistic <sup>120</sup>. The theoretical justification of this approach is based on the idea of the vetted bootstrap developed in the previous chapter. Here, that the population distribution of trial arms should have the same distribution of non-treatment related factors is considered external knowledge (we know this because we randomize patients from the same subject pool into each arm). This knowledge can be incorporated into the analysis by putting the following probability distribution on **F**:

$$\pi_{\mathbf{F}}(\mathbf{F}) \propto \begin{cases} 1 & F_1^z = F_2^z = \cdots = F_N^z \\ 0 & otherwise \end{cases},$$

with Z being the set of parameters that need to have equal distribution across treatment arms in **F**, and  $F_i^Z$  denoting the marginal distribution of such parameters in **F** if the patient receives the *i*<sup>th</sup> treatment. The vetted-bootstrap implementation of this idea means only accepting the bootstraps in which the population distributions of RCT arms are equal in the distribution of *Z*, and rejecting all other bootstrap sets. Of course, this process in its exact form is infinitely unwieldy. The proposed approach is an approximation that converges to the above scheme asymptotically with a RCT of infinite size and with *k* being infinite. The choice of the K-S statistics is somewhat arbitrary. Any measure of distance among two or more distributions that becomes smaller the more similar the distributions become could be a candidate. The second method follows a similar line of reasoning, but uses a more *ad hoc* approach. The use of linear regression for adjusting QALYs on baseline utility in current and future RCTs reflects our assumption that the expected value of QALY in each arm is a linear function of the expected value of baseline utility. Therefore, it is good enough to ensure random samples from the population distributions have the same mean baseline utility across treatment arms. To achieve this, I suggest generating *k* candidate bootstrap sets, and picking the one in which the F-test statistic of a one-way fixed-effects ANOVA for baseline utility across treatment arms has the smallest value (indicating smaller between-group difference in mean).

I compared the performance of the two approaches described above with three different values of *k* (10, 20, and 100). For each method, I calculated the reduction in the between-treatment sum-of-squares for baseline utility and the reduction in the K-S test statistic. Results were compared to a non-adjusted scenario that simply accepts the first bootstrap, regardless of the distribution of baseline utility. I also calculated EVPI based on bootstrapping within each arm, regressing NBs on baseline utility, and estimating the EVPI using the adjusted mean NB for each arm. This last approach provides a fully-adjusted EVPI, but because it does not generate panel data for the future RCT it cannot be used for EVSI calculations for the detailed scenario.

Results of the simulation analysis are provided in **Table 3-2**. As expected, when the selection of the best bootstrap set is based on the K-S test statistic, the reduction in the K-S statistic is higher than when the selection is based on ANOVA. Likewise, ANOVA results in better adjustment based on between-treatment sum of squares than the K-S statistic. In both methods, with k=100, the calculated EVPI becomes close to the fully adjusted EVPI.

EVPI	Sum of	K-S statistic
	squares	
349.1	0.017	0.212
211.0	0.007	0.175
189.5	<0.0001	0.112
175.9	0.0002	0.150
247.0	0.0006	0.130
207.1	0.003	0.147
176.8	0.002	0.125
167.2	NA	NA
	EVPI 349.1 211.0 189.5 175.9 247.0 207.1 176.8 167.2	EVPI         Sum of squares           349.1         0.017           211.0         0.007           189.5         <0.0001

### Table 3-2: The effect of adjusting for covariate imbalance on EVPI calculations

\* Full adjustment is based on regressing mean net benefit on baseline utility. There is no sampling from the population distribution in this case and measures of distribution similarity are irrelevant. EVPI: expected value of perfect information, K-S: Kolmogorov-Smirnov, ANOVA: analysis of variance

Based on these results, I picked the ANOVA method with k=20 for the EVSI calculations in the detailed scenario as I perceived that it provides a trade-off between computational efficiency and balancing the distribution of baseline utility.

# 3.5.2.2.3 Future RCT

A sampling with replacement stratified within each arm of the trial, with a sample size equal to the size of the corresponding arm in the future RCT was performed in order to generate the panel of the data for the future RCT. I also considered the fact that while the future RCT will be aimed at having equal number of patients within each arm, the blinded randomization results in each new patient being assigned with probability of 1/3 to each arm. Therefore, the final number of patients in each arm is itself a random variable with a multinomial distribution. In the next step, I artificially introduced missing values in the future RCT, assuming it would have a similar pattern of missing values as the current RCT. This was modeled by sampling from the current RCT data and applying the missing data pattern to each individual in the future RCT. Finally, the current and future RCT data were merged, and the resulting data was subject to a full CEA to calculate the expected net benefit of the combined results. The future CEA was again based on non-parametric imputation of the missing data and adjusting QALYs for baseline utility using a linear regression, as had been the case in the original analysis.

I used the above algorithm for calculation of the EVPI and EVSI for a range of sample sizes (50 to 1,500 patients per arm) for the future RCT. I compared the results obtained from the simple scenario using Bayesian bootstrap, conventional bootstrap, and the normal approximation of the sample mean of NB as described by Koerkamp et al. <sup>86</sup> with those estimated using the detailed scenario, based on 10,000 repetitions.

#### 3.5.2.2.4 Results

The calculated EVPIs are presented in *Table 3-3.* The EVPIs estimated using the three methods in the simple scenario were generally in the same range. The calculated EVPIs using the detailed scenario for the trial on all three treatments is \$191.3 per individual, which is 45% lower than the EVPI calculated using the conventional bootstrap method in the simple scenario. A substantially lower EVPI was observed for the detailed analysis compared with simple scenarios for the TP/TS comparison. On the other hand, the EVPI calculated based on the detailed analysis for the TS/TFS comparison was almost twice as its counterparts based on the simple scenario.

	Simple scenario			Detailed scenario		
Future RCT design	Normal	Bayesian bootstrap	Approximate Bayesian bootstrap			
TP/TS	331.9	329.8	331.7	164.5		
TP/TFS	70.3	62.8	64.0	58.7		
TS/TFS	62.7	51.2	56.5	114.3		
TP/TS/TFS	356.4	344.2	349.2	191.3		
* The standard errors of the estimates are around 5 after 10,000 simulations.						
All values are in net monetary benefit scale with 2008 \$CAD.						
TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol.						

#### Table 3-3: EVPI for various design and scenarios for the OPTIMAL trial

**Figure 3-3** presents the EVSIs for a range of sample sizes for the future RCT. Again, the estimated EVSIs using the three methods in the simple scenario (marked curves) are closely similar. However, except for trials on TP/TFS, the estimated EVSIs using the detailed scenario are substantially different than those estimated in the simple scenario (lower for the TP/TS/TFS and TP/TS comparisons, higher for the TS/TFS comparison). Such differences generally follow the pattern of the differences in the EVPI between simple and detailed scenarios.



Figure 3-3: EVSI per individual for a future study with similar design as the OPTIMAL study with a range of sample sizes

Each point in the graph is generated by averaging the results of 10,000 simulations. Continuous line without marks: detailed scenario. Dashed line: simple scenario (Bayesian bootstrap), dotted line: simple scenario (approximate Bayesian bootstrap), and continuous line with marks: simple scenario (normal).

TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol.

To further explore the reason for such a disparity between the results of the simple versus

detailed analyses, I calculated the EVPI using the detailed analysis, by turning off, one at a time,

one of the three features that distinguish the detailed analysis from the simple one (imputation of missing values, adjustment for baseline utility, and sample size per arm of the future RCT being a random variable). The result of such analysis is reported in *Table 3-4* and is contrasted against the EVPI estimated from the fully implemented detailed scenario and the simple scenario based on the approximate Bayesian bootstrap.

	Feature disabled*			Fully implemented	Simple analysis
	Missing value imputation	Adjusting QALYs for baseline utility	Random sample size per arm for future RCT	detailed analysis	based on approximate Bayesian bootstrap
тр/тс	168.2	33/1 8	171.6	164 5	331 7
TP/TFS	56.6	65.2	50.8	58.7	64.0
TS/TFS	93.7	56.9	110.0	114.3	56.5
TP/TS/TFS	190.9	353.9	193.3	191.3	349.2

#### Table 3-4: Impact of different aspects of the detailed analysis on the EVPI

\* Results are based on 1,000 simulations. All values are in net monetary benefit scale with 2008 \$CAD.

*TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol, EVPI: expected value of perfect information. QALY: quality-adjusted life year* 

Results of this analysis indicate that the feature of the detailed analysis that is the most responsible for the disparity in the results between the detailed and the simple scenarios is the adjustment of QALYs for baseline utility. When this feature is removed, the results of the detailed analysis became similar to those of the simple analysis.

# 3.6 Discussion

In this chapter I explained a generic, non-parametric framework for EVI analysis. The main

advantage of this method is its versatility which allows it to accommodate a spectrum of

intermediate steps. In its simplest form, two-level resampling for EVSI calculation can directly be performed on the vector of net benefits within each arm of the trial. This is an extension of the popular bootstrap method for the analysis of RCT-based CEAs and can be considered a nonparametric method analogous to methods based on normal approximation of mean net benefit <sup>71,86</sup>. Bootstrapping net benefits -albeit computationally more demanding than the normal approximation method- is still fast and operationally trivial. It is not based on parametric assumptions, and in terms of the underlying statistical assumptions, is conjugate with the popular bootstrap method of RCT-based CEA.

I described a more detailed scenario that illustrated one possible way for incorporating more realistic aspects of EVI analysis, and showed its impact on EVI calculations. The estimates of EVPI and EVSI using the detailed scenario showed significant differences from those in the simple scenario. While the estimates of EVPI using the three methods in the simple scenarios were generally close to each other, estimates of EVPI from the detailed scenario varied from being less than half (TP/TS) to more than twice (TS/TFS) as large as its counterparts in the simple analysis. EVSIs for any given sample size also followed the same pattern. These differences are large enough to have an impact on the sample size of the future RCT, if the sample size is informed by the EVSI, based on maximizing the difference between population EVSI and the budget of the trial <sup>88</sup>.

Because the estimated EVPIs and EVSIs using the simple two-level bootstrap are very close to their counterparts in the normal approximation method, it can be concluded that the different results estimated from the detailed scenario is not because of the non-parametric vs.

parametric nature of calculations, rather they are due to the incorporation of some realistic analytical aspects. Further analysis of the results after disabling features of the detailed scenario highlighted the role of adjusting the estimated QALYs for baseline utility. This should not come as a surprise given the substantial impact of such an adjustment on the estimated mean QALYs for each arm (refer to *Table 1-1*). The normal approximation method can conceivably be upgraded to incorporate adjustment for baseline covariates, for example, by estimating the mean and variance of NBs for each arm of the trial from a regression model with adjustment for baseline utility. But this is not as rigorous as the approach used in the detailed scenario as it fails to properly model the potential lack of balance in the distribution of baseline covariates in the future study (which is automatically modeled in the detailed scenario as part of the second-level sampling). More elaborate schemes such as joint modeling of baseline utility and NBs can enable the normal approximation method to more rigorously address this, but it will cost this method its main appeal: its simplicity. Meanwhile, there is little doubt that, in other situations, other aspects of the analysis such as handling of missing values, which can be realistically modeled in a two-level resampling scheme, will turn out to be important factors affecting the EVI results.

The questions that I answered in this work, estimating the value of conducting a future RCT with similar design to a current RCT, is not the only questions that can be explored using this approach. The two-level sampling paradigm is a framework to which more problem-specific details can be added. A relevant scenario is to model the future RCT to have a longer follow-up period. After drawing a sample from the population distribution **F**, one can use time series methods for extrapolating the data into future time points, considering the statistical

uncertainty in the predicted values, and perform other steps for EVI calculations accordingly. Other questions that can be explored using this framework include estimating the benefit of attempts to prevent attrition from the future RCT and the benefit of controlling for possible confounding effects at the design stage (e.g., by stratified randomization), to name a few. One can compare the costs and return on investment of RCTs with and without such design specifications, and the two-level method appears to be flexible enough to address these and similar scenarios. In a broader sense, this approach could have applications beyond economic evaluation of RCTs. It can be seen as a way to generate individual-level samples from a future clinical trial based on the data of a current RCT, which can be used to generate predictive probability distributions for the results of a planned analysis.

I acknowledge several limitations with this approach. The foremost drawback is the computational demand of a detailed analysis (it required 2,000 times more computational time than the simple scenario based on the normal method), and its requirement for context-specific statistical programming. Incorporating more advanced scenarios such as extrapolating behind the time horizon of the current RCT will only add to such computational burden. Secondly, merging the individual-level data of the current and future RCTs amounts to a fixed-effects analysis. This is not the only paradigm for evidence synthesis, and other approaches such as random-effects models should be evaluated in this context <sup>116</sup>.

A growing number of trials incorporate economic end-points at the design stage and there are established protocols and guidelines for conducting economic evaluation alongside a RCT<sup>1</sup>. With soaring costs, the design and conduct of a RCT is a formidable undertaking, and rigor and

objectivity in planning such studies in order to optimize the investment are worth taking. The stakes seems to be high enough to justify the time and resources required for conducting a realistic EVI analysis to fine-tune the design of a RCT. The present method adds to the available toolkit of EVI methods and sets the stage for future studies to compare and contrast different approaches from both theoretical and practical perspectives.

# Chapter 4. A heuristic algorithm for calculation of single-parameter expected value of partial perfect Information

# 4.1 Background

A robust, unbiased, and easy-to-implement method for calculating the expected value of the outcomes and quantifying uncertainty in CEAs is to perform a Monte Carlo simulation. In model-based CEAs, this is done by randomly drawing from the distribution of uncertain parameters and calculating cost and effectiveness outcomes. In RCT-based CEAs, this can be performed by bootstrapping within each arm of the trial, and calculating the mean cost and effectiveness outcomes for each arm from the bootstrap sample. Generally, this method in the health economics literature is referred to as probabilistic sensitivity analysis (PSA).

The EVPPI is the expected gain in benefit by completely resolving uncertainty around a subset of evidence <sup>73</sup>. The EVPPI can be used as a generic measure to compare the relative importance of uncertainty in parameters of a decision model. Population EVPPI sets an analytical upper limit on the budget of future research aimed at obtaining more information on those parameters.

Unfortunately, calculation of EVPPI is often computationally intensive as it generally requires a two-level nested Monte Carlo expectation <sup>73</sup>. For model-based CEAs, alternative methods for EVPPI calculation have been proposed; they are either based on parametric assumptions or work only in special cases (e.g., when the model is multi-linear on its parameters), while others have been proved to be incorrect (see <sup>112</sup> for a review). There are some meta-modeling

approaches in calculating the EVPPI but they too come with certain assumptions and require considerable expertise for implementation <sup>112,113</sup>. For RCT-based CEAs, the calculation of the EVPPI is much less discussed. Koerkamp et al. performed EVPPI calculations for a trial on endovascular revascularization or supervised exercise training for the treatment of intermittent claudication. This was performed by assuming joint normal distribution for the parameters of interest and the rest of parameters, which allowed them to implement the two-level Monte Carlo method <sup>86</sup>. But the parameters were all subsets if cost and effectiveness outcomes, and the method does not seem to be readily extendable to other parameters such as the effect size of treatment that might not be identified during the CEA.

The calculation of the two-level expectation using Monte Carlo requires that the inner expectation be calculated while  $\theta$  is fixed at a given value set at the outer expectation loop. For RCT-based CEAs, if the parameter of interest has an explicit probability distribution, such as a unit price, then the outer expectation can be performed by randomly drawing from the distribution of the parameter, while the inner expectation can be performed by bootstrapping and calculating mean NB for each arm with  $\theta$  fixed at the value set at the outer loop. When  $\theta$ itself is the function of the RCT data (such as the treatment effect size), then the random draw from  $\theta$  for the outer loop can be obtained by bootstrapping. But in this case, in the inner loop only the bootstrap sets in which the value of  $\theta$  is equal to the fixed value generated from the outer loop should be accepted. This is indeed an (infinitely) unwieldy process; therefore the bootstrap method cannot readily be used in two-level EVPPI calculation for RCT-based CEAs when the parameter of interest is a function of the RCT data.

This chapter presents a novel and simple method for calculating single-parameter EVPPI that is applicable to both model-based and RCT-based CEAs. The main advantages of the present method are its computational efficiency and that it only relies on the data generated through the PSA, which is a standard output of any stochastic CEA. During the final preparation of this thesis I encountered an unpublished report by Strong et al. presenting an approach for EVPPI calculation which is operationally somewhat similar to the present method <sup>121</sup>. Nevertheless, the present method is based on a different theoretical justification and seems to have important advantages to the method developed by Strong et al. which I will explain in the Discussion section.

I begin by defining the mathematical formulation of the EVPPI. Next, I outline the heuristic underlying the present method. Based on such heuristic, I propose an estimator for the EVPPI. The convergence in probability of the estimator to the true EVPPI under the necessary regularity conditions is formally proved. I explain a visual method for checking the performance of the algorithm. I then compare the numerical and computational performance of the new algorithm with those of the conventional two-level Monte Carlo simulation method using three exemplar decision-analytic models. The OPTIMAL clinical trial is used as a case study demonstrating the feasibility of such calculations for RCT-based CEAs.

#### 4.2 Methods

## 4.2.1 Context and notations

Let  $\Omega$  denote the set of stochastic quantities that inform the underlying decision task. In modelbased CEAs,  $\Omega$  is the set of all uncertain model parameters that are represented by probability

distributions. In RCT-based CEAs, it is the raw RCT data, plus any stochastic parameters that might be used in deriving cost-effectiveness outcomes. Examples of such stochastic parameters in the context of a RCT-based CEA include unit prices, if a probability distribution is assigned to their values, which will be multiplied by individual resource use data from the RCT, or if regression coefficients are used to convert a disease-specific measure of quality of life to health state utility values (if such coefficients are uncertain). I denote the single parameter whose EVPPI is of interest as  $\theta$ . Let  $NB_d(\Omega)$  represent the process that calculates the NB of the  $d^{th}$ strategy associated with a realized value of  $\Omega$ . One fundamental assumption in this work, as in any stochastic CEA, is that  $NB_d(\Omega)$  has an expected value and a finite variance, such that by many times sampling from  $\Omega$ , calculating  $NB_d(\Omega)$ , and averaging the results one obtains a value that converges to the expected net benefit of the  $d^{th}$  strategy. Let  $P_{\theta}(\theta)$  be the probability density function of  $\theta$ . The present method is applicable to the parameters with continuous probability distributions<sup>\*</sup>. In model-based CEAs,  $\theta$  is typically one of the input parameters. For RCT-based CEA, it can be one of the uncertain quantities (such as a unit price) that are used to carry out the CEA, or alternatively, it can be a summary statistic that is estimated from the RCT data, such as the effect size of treatment. Let  $\theta_L$  and  $\theta_U$  denote the lower and upper bounds (either or both can be infinite) of  $\theta$ .

The PSA is performed by randomly drawing from the distribution of  $\Omega$  and calculating the net benefit (NB) for all *D* strategies, repeating this process *n* times. The PSA data can be denoted by  $\{\widehat{\theta}, \widehat{NB}\}$  where  $\widehat{\theta} = (\widehat{\theta}_i, i = 1: n, \widehat{\theta}_1 \le \widehat{\theta}_2 \dots \le \widehat{\theta}_n)$  is the random draws from the distribution of

<sup>&</sup>lt;sup>\*</sup> When using conventional bootstrap for RCT-based CEAs, the distribution of any aggregate statistics based on the RCT data is inevitably discrete. The present algorithm therefore relies on asymptotics in this situation.

the parameter of interest (ordered ascendingly, without loss of generality), and  $\widehat{\mathbf{NB}} = (\widehat{\mathbf{NB}}_{i,d}, i = 1: n, d = 1: D)$  is the corresponding matrix of NBs (the draws from the other parameters are irrelevant and are omitted in this notation).

I define the function  $ENB_d(\theta)$  as the expected NB of the  $d^{th}$  strategy conditional on the parameter of interest being fixed at  $\theta$ :

$$ENB_{d}(\theta) \equiv E_{\mathbf{\Omega}|\theta}(NB_{d}(\mathbf{\Omega})).$$
<sup>(2)</sup>

In this case the EVPPI for the parameter  $\theta$  can be written as

$$EVPPI_{\theta} \equiv E_{\theta}\{\max_{d}[ENB_{d}(\theta)]\} - \max_{d}\{E_{\theta}[ENB_{d}(\theta)]\}.$$
(3)

The main difficulty in calculating (3) is the first term with two nested expectations separated by a maximization step. As these expectations are analytically intractable for all but the simplest situations, a two-level Monte Carlo simulation is often used for their calculation, such that the inner and outer levels perform, respectively, the inner and outer expectations in the left term of (3)<sup>73</sup>. There does not seem to be an overarching rule for the sample size of the simulations <sup>122</sup>, but even with a few thousand iterations at each level, the overall number of simulation runs required could easily become overwhelming.

#### 4.2.2 Concept

The concept underlying the present method can be described as 'data segmentation approach to EVPPI calculation'. I start by defining

$$\delta(\theta) \equiv \operatorname{argmax}_{d} ENB_{d}(\theta) , \qquad (4)$$

the function that returns the index of the strategy that has the highest ENB at a given value of  $\theta$ . The heuristic is that if a strategy has the maximum ENB at  $\theta$ , it probably has the maximum ENB at the vicinity of  $\theta$  as well. The  $\delta$  function is therefore, for the most realistic scenarios, piecewise constant with finitely many pieces (as shown in **Figure 4-1**). As such, it is good enough to restrict our attention to the set of such functions in calculating the EVPPI.



Figure 4-1: Schematic illustration of the segmentation approach to EVPPI calculation

The curved lines are the ENB functions for three hypothetical strategies at different values of the parameter of interest. In this example, the two points for which the strategy that has the maximum ENB changes divide the range of the parameter of interest into three segments. Within each segment, the best strategy remains the same.

The expectation that  $\delta$  is piecewise constant rests generally on logical relationship between the parameter of interest and the outcomes of decisions. The nature of such relationship can often be deduced even before looking at the data. For example, it is rationally expected that the higher the treatment effect size, the higher the incremental net benefit of treatment vs. no treatment; or the higher the prevalence of the disease, the higher the incremental net benefit of the benefit of a screening vs. no screening strategies. That is, the incremental net benefit function between

such pairs of decisions monotonically varies with the parameter of interest. Therefore the ENB of the treatment and no treatment (or screening and no screening) decisions are unlikely to cross too many times.

Restricting our attention to the set of piecewise constant functions proves advantageous: if  $d^*$  consists of M + 1 pieces created by M segmentation points ( $m_i$ ; i = 1: M;  $m_1 < m_2 < \cdots < m_M$ ), then the left term of the EVPPI can be rewritten as

$$\mathbb{E}_{\theta}\{\max_{d}[\mathbb{E}NB_{d}(\theta)]\} = \max_{m_{1},\dots,m_{M}}\left\{\sum_{i=0}^{M}\max_{d}\left[\int_{m_{i}}^{m_{i+1}}\mathbb{E}NB_{d}(x).p_{\theta}(x).dx\right]\right\},$$
(5)

with  $m_0 = \theta_L$  and  $m_{M+1} = \theta_U$ . The reason for using such a formula for the EVPPI is that the right side of equation (5) can be estimated from the PSA data. Define the vector **L** ( $l_i$ ;  $i = 1:L; l_1 < l_2 < \cdots < l_L$ ) as the  $L \times 1$  vector of candidate segmentation points. The elements of **L** correspond to the row indices of elements in  $\hat{\theta}$ . Define  $\psi(\mathbf{L})$  as

$$\psi(\mathbf{L}) = \frac{1}{n} \sum_{i=0}^{L} \max_{d} \left( \sum_{j=l_i+1}^{l_{i+1}} \widehat{NB}_{j,d} \right), \tag{6}$$

with  $l_0 = 0$  and  $l_{L+1} = n$ . Then define

$$\widehat{EVPPI} = \max_{l_1, l_2, \dots, l_L} \psi(\mathbf{L}) - \frac{1}{n} \cdot \max_d \sum_{i=1}^n \widehat{NB}_{i,d},$$
(7)

as an estimator of the EVPPI from the PSA data. The intuition in the above formula is that the quantity  $\max_{l_1, l_2, .., l_L} \psi(\mathbf{L})$  is analogous to the term in the right side of equation (5) but estimated from the PSA data.

Putting all these together, the heuristic algorithm for single-parameter EVPPI calculation suggests that one calculates (7) with a high enough L from the PSA data. The power of this method is its computational efficiency compared to the two-level Monte Carlo method as well as its ability to estimate EVPPI for all individual parameters from one set of PSA data.

# 4.2.2.1 Proof of the convergence of EVPPI to EVPPI

Here I prove that with a fixed size of L, and provided that L is equal or greater than M, the true number of segmentation points (discontinuity points on  $\delta$ ), the term  $\max_{l_1, l_2, ., l_L} \psi(\mathbf{L})$  converges in probability to the left term of EVPPI (i.e.,  $E_{\theta}\{\max_d[ENB_d(\theta)]\}$ ). This is proved in two stages.

**Lemma 4.1:** Let  $\mathbf{M}$   $(m_i; i = 1: M; m_1 < m_2 < \cdots < m_M)$  be the  $M \times 1$  vector of true segmentation points on  $\theta$  that maximizes equation (5). Each element of  $\mathbf{M}$  is a real value within the range of  $\theta$ . Let  $\mathbf{L}$   $(l_i; i = 1: L; l_1 \le l_2 \le \cdots \le l_L)$  be the  $L \times 1$  vector of candidate segmentation points. The elements in  $\mathbf{L}$  correspond to the row indices of elements in  $\widehat{\mathbf{\theta}}$ . If  $L \ge M$  then

$$\lim_{n \to \infty} \frac{1}{n} \max_{l_1, \dots, l_L} \left\{ \sum_{i=0}^{L} \max_d \left[ \sum_{j=l_i+1}^{l_{i+1}} ENB_d(\hat{\theta}_j) \right] \right\} = E_\theta \left\{ \max_d [ENB_d(\theta)] \right\},$$

with  $l_0 = 0$  and  $l_{L+1} = n_{.}$ 

**Proof:** Since by definition, within any segment created by **M** one decision has the maximum f at all values of  $\theta$ , we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{M}\max_{d}\left[\sum_{j=m_{i}+1}^{m_{i+1}}ENB_{d}(\hat{\theta}_{j})\right] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\max_{d}\left[ENB_{d}(\hat{\theta}_{i})\right] \xrightarrow{a.s.} E_{\theta}\left\{\max_{d}\left[ENB_{d}(\theta)\right]\right\},$$

with  $m_0 = \theta_L$  and  $m_{M+1} = \theta_U$ . The notation  $\xrightarrow{a.s.}$  implies almost sure convergence according to the strong law of large numbers <sup>123(p60)</sup>.

So by setting  $l_0 = 0$  and  $l_{L+1} = n$ , and all other  $l_i$  to the index of the largest element in  $\hat{\theta}$  that is smaller than  $m_i$  (and this is why we need  $L \ge M$ ; extra elements in **L** are all set to n) the right side term is achievable. Also, this is the maximum value a piecewise constant function can achieve as it picks the maximum NB at all points of  $\hat{\theta}$ . This guarantees the above equality.

**Lemma 4.2:** If  $L \ge M$  then

$$\max_{l_1,\dots,l_L} \psi(\mathbf{L}) \xrightarrow{p} E_{\theta} \left\{ \max_{d} [ENB_d(\theta)] \right\}$$

where  $\xrightarrow{p}$  indicates convergence in probability.

**Proof:** During the proof we use the following propositions (whose proofs are either very simple or well known) that are later referenced by their index.

For any real-valued matrix **X** of size  $n \times m$ :

1) 
$$\max_{j} (\sum_{i=1}^{n} X_{ij}) \le \sum_{i=1}^{n} \max_{j} (X_{ij})$$

2) 
$$\min_{j} (\sum_{i=1}^{n} X_{ij}) \ge \sum_{i=1}^{n} \min_{j} (X_{ij})$$

3) 
$$\min_{j} \left( \sum_{i=1}^{n} X_{ij} \right) = -\max_{j} \left( \sum_{i=1}^{n} - X_{ij} \right)$$

4) If **Y** is a vector of random variables of size *n* then  $P[\max_i(Y_i) > a] \leq \sum_{i=1}^n P(Y_i > a)$ .

5) Kolmogorov inequality <sup>124(sec22.4)</sup>: If **Y** is a vector of *n* independent random variables with finite (not necessarily equal) variance and zero expectation, for each  $\lambda > 0$ ,

$$P\left[\max_{1\leq m\leq n}\left(\left|\sum_{i=1}^{m}Y_{i}\right|\right)\geq\lambda\right]\leq\frac{1}{\lambda^{2}}\sum_{i=1}^{n}var(Y_{i}).$$

Now we proceed with the main proof:

$$\max_{l_1,\dots,l_L} \left[ \frac{1}{n} \cdot \sum_{i=0}^{L} \max_d \left( \sum_{j=l_l+1}^{l_{l+1}} \widehat{NB}_{j,d} \right) \right] = \frac{1}{n} \cdot \max_{l_1,\dots,l_L} \left( \sum_{i=0}^{L} \max_d \left\{ \sum_{j=l_l+1}^{l_{l+1}} \left[ ENB_d(\widehat{\theta}_j) + e_d(\widehat{\theta}_j) \right] \right\} \right)$$

(where  $e_d(\hat{\theta}_i)$  is an error term with zero expectation)

 $\stackrel{\#1 \text{ for inner maximization}}{\stackrel{=}{\longrightarrow}} \\ \leq \frac{1}{n} \cdot \max_{l_1, \dots, l_L} \left\{ \sum_{i=0}^{L} \left[ \max_{d} \sum_{j=l_i+1}^{l_{i+1}} ENB_d(\hat{\theta}_j) + \max_{d} \sum_{j=l_i+1}^{l_{i+1}} e_d(\hat{\theta}_j) \right] \right\} \\ \stackrel{\#1 \text{ for outer maximization}}{\stackrel{=}{\longrightarrow}} \\ \leq J + K,$ 

where

$$J = \frac{1}{n} \cdot \max_{l_1,..,l_L} \left\{ \sum_{i=0}^{L} \left[ \max_{d} \sum_{j=l_i+1}^{l_{i+1}} ENB(\hat{\theta}_j) \right] \right\}, \qquad K = \frac{1}{n} \cdot \max_{l_1,..,l_L} \left\{ \sum_{i=0}^{L} \left[ \max_{d} \sum_{j=l_i+1}^{l_{i+1}} e_d(\hat{\theta}_j) \right] \right\}.$$

*J* is already proved in lemma 4.1 to converge to  $E_{\theta}\{\max_{d}[ENB_{d}(\theta)]\}$  as *n* grows to infinity. So we proceed by proving that *K* converges in probability to zero. This term is the maximum sum of errors over *L* + 1 segments, and hence is not greater than *L* + 1 times the sum of errors of a segment with maximum sum of errors, and not less than *L* + 1 times the sum of errors of a segment with minimum sum of errors:

$$\frac{L+1}{n} \cdot \min_{d,1 \le a \le b \le n} \sum_{j=a}^{b} e_d(\hat{\theta}_j) \le K \le \frac{L+1}{n} \cdot \max_{d,1 \le a \le b \le n} \sum_{j=a}^{b} e_d(\hat{\theta}_j)$$

expanding the summands

$$\frac{L+1}{n} \cdot \min_{d,1 \le a \le b \le n} \left[ \sum_{j=1}^{b} e_d(\hat{\theta}_j) + \sum_{j=1}^{a} - e_d(\hat{\theta}_j) \right] \le K \le \frac{L+1}{n} \cdot \max_{d,1 \le a \le b \le n} \left[ \sum_{j=1}^{b} e_d(\hat{\theta}_j) + \sum_{j=1}^{a} - e_d(\hat{\theta}_j) \right]$$

Now we change the maximization and minimization condition  $1 \le a \le b \le n$  to  $\{1 \le a \le n, 1 \le b \le n\}$ . Since the new condition is less restrictive, the maximization cannot result in a smaller value, and the minimization not in a larger value, and we can continue

where  $Q \equiv \frac{1}{n} \max_{\substack{d, \\ 1 \le a \le n}} |\sum_{j=1}^{a} e_d(\hat{\theta}_j)|$  and  $R \equiv \frac{1}{n} \max_{\substack{d, \\ 1 \le a \le n}} |\sum_{j=1}^{a} - e_d(\hat{\theta}_j)|$ . We show that  $Q \xrightarrow{p} 0$  (the proof for

 $R \xrightarrow{p} 0$  follows by symmetry, and indeed  $Q \xrightarrow{p} 0$  and  $R \xrightarrow{p} 0$  implies  $Q + R \xrightarrow{p} 0$ ). With any positive value  $\lambda$  we can proceed:

$$P(Q > \lambda) = P\left\{\frac{1}{n} \max_{d} \left[\max_{1 \le a \le n} \left|\sum_{j=1}^{a} e_d(\hat{\theta}_j)\right|\right] > \lambda\right\}$$
$$\xrightarrow{\#4}$$
$$\leq \sum_{d=1}^{D} P\left[\max_{1 \le a \le n} \left|\sum_{j=1}^{a} \frac{1}{n} e_d(\hat{\theta}_j)\right| > \lambda\right].$$

Applying the Kolmogorov inequality (#5) with  $\lambda = n^{-1/4}$  for each term of the above summation, we can write

$$P\left(Q > n^{-\frac{1}{4}}\right) \le n^{\frac{1}{2}} \sum_{d=1}^{D} \sum_{j=1}^{n} \frac{1}{n^{2}} var[e_{d}(\hat{\theta}_{j})] = n^{-\frac{1}{2}} \sum_{d=1}^{D} \hat{X}_{d}$$

where

$$\hat{X}_d = \frac{1}{n} \sum_{j=1}^n var[e_d(\hat{\theta}_j)].$$

The proof is complete if we can show that  $\lim_{n\to\infty} (n^{-1/2} \sum_{d=1}^{D} \hat{X}_d) = 0$  as this will guarantee that for any positive value  $\lambda$ ,  $\lim_{n\to\infty} P(Q > \lambda) = 0$ , which implies convergence in probability. We note that  $\hat{X}_d$ is a Monte Carlo estimator of  $E_{\theta}\{var[e_d(\theta)]\}$ . Because  $var[e_d(\theta)] \equiv E_{\Omega^-|\theta}[NB_d(\Omega)^2] - f_d(\theta)^2$ , we have  $\hat{X}_d \xrightarrow{a.s.} E_{\theta}\{E_{\Omega^-|\theta}[NB_d(\Omega)^2] - f_d(\theta)^2\} \equiv E_{\Omega}[NB_d(\Omega)^2] - E_{\theta}[f_d(\theta)^2]$ . We note that  $E_{\Omega}[NB_d(\Omega)^2] \equiv$  $V_d + \mu_d^{-2}$ , where  $V_d$  and  $\mu_d$  are the variance and expected value, respectively, of the distribution of the NB for the  $d^{th}$  decision, and the remaining term  $E_{\theta}[f_d(\theta)^2]$  is non-negative. Therefore, and given that  $\hat{X}_d$  is non-negative,  $0 \leq \lim_{n\to\infty} \hat{X}_d \leq V_d + \mu_d^2$ . Because of the fundamental assumption of the finiteness of D,  $V_d$ , and  $\mu_d$  made in the introduction,  $\lim_{n\to\infty} (n^{-\frac{1}{2}} \sum_{d=1}^{D} \hat{X}_d) = 0$ .

# 4.3 Deciding on the number of segmentation points, and a visual tool for model checking

The convergence of EVPPI to the true EVPPI rests on the critical assumption that L, the number of fitted segmentation points, is at least as large as M, the number of true segmentation points. M is indeed unknown, but one can choose a very large L to ensure this condition is satisfied. However, I note that each additional segmentation point can cause overestimation of the EVPPI in finite PSA samples (because of the maximization step in (7)). Therefore a parsimonious choice for L is important for avoiding overtly overestimated EVPPIs.

Again, I refer to the logical expectation that the incremental net benefit as a function of  $\theta$  between any two strategies is, for the most likelihood, a monotonical function and as such the *ENBs* of any two decisions cross each other at most once. This line of reasoning suggests that an economical choice for *L* is *D*. (D - 1)/2, the total number of the pairs of decisions in a *D*-decision task.

Fortunately, there is a powerful visual tool for assessing such an assumption: for a pair of strategies a and b, define the quantity

$$\hat{S}(\theta) = \frac{1}{n} \sum_{i=1}^{n} I(\hat{\theta}_i < \theta) (\widehat{NB}_{i,a} - \widehat{NB}_{i,b})$$

where I(.) is the indicator function taking the value of one when the condition is satisfied, and zero otherwise.  $\hat{S}(\theta)$  is the running cumulative sum of the incremental net benefits from the PSA data (after creating an ordered version of the PSA by sorting the data on  $\hat{\theta}$ ), which can be
calculated very easily in a spreadsheet or using a simple function in any computer program. In addition,  $\hat{S}(\theta)$  is a Monte Carlo estimator of

$$S(\theta) = \int_{\theta_I}^{\theta} [ENB_{a}(x) - ENB_{b}(x)] \cdot P_{\theta}(x) \cdot dx.$$

Because any crossing point of  $ENB_a$  and  $ENB_b$  corresponds to an extremum on  $S(\theta)$ , one can expect to observe an extremum on its Monte Carlo estimator,  $\hat{S}(\theta)$ , around such values as well. This is clearly observable in *Figure 4-2*. These are the running cumulative sums of the incremental net benefit associated with the three parameters of a decision-analytic model (model 1 in subsequent sections). The left panel presents the scatter plot of the incremental net benefits against  $\hat{\theta}$  from the PSA data. The right panel is the  $\hat{S}(\theta)$  function for the same parameter. Clearly, an extremum is visible on  $\hat{S}(\theta)$  for the first and second parameters, indicating a positive EVPPI, while for the third parameter, the function does not have a clear extremum, suggesting that EVPPI=0.





Vertical lines in the second row are the global extrema of the function

The visual inspection of the performance of the algorithm therefore involves plotting the  $\hat{S}$  calculated from the PSA data for all pairs of decisions, and checking if an extremum is visible (or if there are more than one segmentation points). The software implementation of the algorithm provided in the **Appendix A.5**, under default settings, automatically generates such

pair-wise plots, and the user is able to force the algorithm to fit a given number of segmentation points between pairs of decisions after observing the plots.

#### 4.4 Simulation studies

### 4.4.1 Comparing the performance of the new algorithm and the conventional twolevel simulation method for model-based CEAs.

I compared the performance of this algorithm with that of the conventional two-level method using three decision-analytic models. Model 1(*Figure 4-3*, left panel) is a simple model for which the EVPPIs could be derived analytically (see *Appendix A.3.* for analytical derivation of EVPPIs for this model). Model 2 (*Figure 4-3*, right panel) is a relatively simple decision tree that, on top of model 1, will test the performance of the algorithm when more than two decisions are compared. Model 3 is a more complex, realistic decision tree and Markov model used by Brennan et al. for comparing different methods of EVPPI calculations <sup>73</sup>. All simulations were run on the default setting of fitting D. (D - 1)/2 segments for a *D*-decision task. Because the algorithm was automated to repeat the simulation 1,000 times, no visual evaluation of the segmentation was performed. All model outputs were converted to net benefits with a willingness-to-pay of 50,000 for unit of effectiveness for EVPPI calculations. I also compared the computational performance of the two methods. This was done on the R implementation of the algorithm on a personal computer (a typical setting for the application of this method) using model 1.

### Figure 4-3: Schematic illustration and parameter specification for model 1 (left) and model 2 (right)



Results are reported as the mean and standard deviation (SD) of the estimated EVPPI by each algorithm at various sizes of Monte Carlo simulations. For model 1 (for which EVPPIs could also be calculated analytically, see *Appendix A.3*), the root mean squared error (RMSE) is also reported. To calculate the mean, SD, and RMSE, I repeated all simulations 1,000 times.

Calculations were performed in MATLAB (version 7.6.0 for Linux, Matchwork Inc. Natick,

Massachusetts, USA) on a multi-processor computer running Red Hat Linux version 2.6.9 (Red

Hat, Inc. Raleigh, North Carolina, USA). The choice of the software and platform for this

simulation was for the sake of computational performance given the number of model runs required, especially for the two-level Monte Carlo method. Both methods can be implemented for use with personal computers using popular software (see section 4.6).

**Table 4-1** presents the mean and SD of the new and the two-level estimators. For the most part, the higher precision of the new estimator is obvious. An interesting comparison is between the new method with 1,000,000 and two-level method with 1,000×1,000 (number of outer simulation runs x number of inner simulation runs) iterations for model 1, as they involve equal number of model runs. The calculated RMSEs for the parameter *pSurvival\_Rx* and *pSurvival\_NoRx* were more than 20 times higher in the two-level method than the new method. Another interesting observation is that for the parameter *cRx*, which has a true EVPPI of zero, the reduction in its SD with successively increasing PSA sizes is greater than the inverse of the square-root of the PSA size (as one would expect for an estimator). This is likely due to the fact that the SD of the estimator is affected both by the chance of the algorithm in finding a segmentation point, as well as the sample standard error of the EVPPI given a fixed segmentation point, both of which decreases as the PSA size increases. For model 2, the ratio of the variance varies from more than 900 for *pSurvival* to 31.8 for *cRx1*. For model 3, the gain in performance is less spectacular, with the ratio of variance ranging from 1.6 for *P2* to 4.4 for *P3*.

On an average personal computer, it took 90 milliseconds (ms) to generate a PSA data with the size of 100,000 for model 1. The segmentation algorithm required a further 610 ms for each parameter. Therefore, around 1,920 ms was required to calculate the EVPPI for all three parameters. On the other hand, the two-level Monte Carlo with 10,000X1,000 samples (which

gives less precise results than the one-level method with PSA size of 10,000, as indicated by the results in *Table 4-1*) required 25,320 ms to estimate EVPPIs for all three parameters. This demonstrates the substantial computational advantage obtained by using the new algorithm.

### Table 4-1: Results of the simulation analysis comparing the performance of the novel and two-level Monte Carlo method for EVPPI calculation

			New method Monte Carlo sample size				Two-level method Monte Carlo sample size (Outer X Inner loops)			
		N								
		1,000	10,000	100,000	1,000,000	1,000X	10,000X	1,000X	10,000X	
						1,000	1,000	10,000	10,000	
Model 1 mean (SD) [RMSE]	Psurvival NoRx	3123.8	3130.4	3120.1	3121.6	3127.6	3125.5	3117.9	3123.5	
		(223.9)	(80.0)	(21.6)	(6.1)	(165.7)	(49.8)	(160.6)	(50.3)	
		[224]	[80]	[24]	[6.2]	[166]	[50]	[161]	[51]	
	PSurvival Rx	1655.1	1643.1	1622.0	1619.4	1613.7	1623.0	1612.5	1617.4	
		(232.2)	(68.6)	(20.1)	(6.5)	(161.4)	(52.3)	(170.5)	(53.6)	
		[235]	[70]	[29]	[7]	[162]	[53]	[171]	[54]	
	cRx	29.2	3.1	0.3	0.0	0.2	0.4	0.9	0.6	
	(EVPPI=0)	(32.0)	(3.0)	(0.3)	(0.03)	(13.7)	(8.3)	(12.1)	(7.8)	
		[43]	[26]	[2.9]	[0.3]	[14]	[8.3]	[12]	[7.8]	
Model 2 mean (SD)	pSurvival	1252.0	1281.5	1255.4	1277.2	1360.0	1362.4	1374.7	1365.8	
		(181.2)	(117.6)	(31.0)	(14.6)	(455.6)	(137.5)	(431.0)	(132.9)	
	OR1	15409.0	1419.1	1394.4	1397.1	1403.6	1369.0	1460.0	1358.0	
		(254.9)	(72.1)	(17.1)	(1.5)	(18.6)	(13.8)	(13.6)	(13.1)	
	OR2	3.4	2.1	1.1	0.8	-1.4	-2.1	-10.3	-7.9	
		(16.8)	(1.1)	(0.5)	(0.1)	(23.1)	(14.2)	(19.6)	(13.6)	
	cRx1	235.2	317.2	315.9	334.0	301.0	316.1	326.9	316.6	
		(121.3)	(56.5)	(9.2)	(3.3)	(18.0)	(13.6)	(13.6)	(13.1)	
	cRx2	16.2	9.0	4.3	0.9	-13.3	-9.3	-7.0	-8.3	
		(76.2)	(6.9)	(1.5)	(0.9)	(17.8)	(13.7)	(13.6)	(13.0)	
Model 3 mean (SD)	P1	48.4	217.3	327.0	317.3	303.8	334.1	290.6	305.9	
		(257.6)	(237.1)	(32.7)	(9.2)	(17.9)	(13.7)	(13.8)	(13.0)	
	P2	3249.0	3127.5	3113.1	3103.5	3115.9	3112.3	3188.1	3065.2	
		(487.9)	(142.9)	(46.2)	(14.5)	(18.6)	(13.7)	(13.7)	(13.1)	
	Р3	19.9	293.5	328.6	317.7	425.3	339.2	502.3	307.1	
		(154.0)	(71.1)	(27.9)	(8.5)	(17.8)	(13.6)	(13.6)	(13.1)	
	Ρ4	1368.7	1309.8	1216.9	1257.1	1116.6	1312.9	988.0	1314.8	
		(513.6)	(124.0)	(43.7)	(10.7)	(18.0)	(13.7)	(13.5)	(13.1)	
SD: standar	deviation DMSE: root maan cauged arror									

#### 4.5 Case study of EVPPI calculations for the OPTIMAL trial

The OPTIMAL trial's main outcomes were the effect size of TFS and TS versus TP for the prevention of respiratory exacerbations. Because the trial incorporated a prospective economic evaluation, it allowed for calculating cost-effectiveness outcomes. In the previous chapter, I calculated the overall EVPI and EVSI as the function of sample size for a future study with a similar (or nested) design to the OPTIMAL trial. However, conducting another study similar to the OPTIMAL trial is not the only way of obtaining evidence on the cost-effectiveness of combination therapies in COPD. For example, a very attractive alternative design is to use observational studies using electronic health databases to estimate the effectiveness of combination medications. The estimate of the effect size from such a study can be used to further inform cost-effectiveness of combination therapies. This can be done, for example, by using the vetted bootstrap approach discussed in Chapter 2 to update the measures of costeffectiveness from the OPTIMAL trial with the estimates of the effect size from the observational study as the external information. If such a design is in mind, the question would be how much the expected benefit of conducting a study that will estimate the effect size of TFS vs. TP, or TS vs. TP will be. This question can be answered by calculating the EVPPI of TFS vs. TP and TS vs. TP using the approach presented here.

To calculate the EVPPI using the algorithm described above, I generated PSA data through n=10,000 Bayesian bootstraps from the OPTIMAL study. Within the bootstrap loops, I followed the original CEA approach and adjusted the calculated net benefits for the baseline utility, and imputed the missing cost and effectiveness data using non-parametric missing value imputation

based on the propensity score. Single–parameter EVPPI was calculated for the OR of TFS vs. TP and TS vs. TP, with QALY as the effectiveness outcomes, for a range of willingness to-pay from 0 to 200,000. All the results were repeated using the conventional (approximate Bayesian) bootstrap as well.

For both the Bayesian and the approximate Bayesian bootstraps, and for the entire range of the WTP, the EVPPI for the effect size of TS vs. TP was close to zero, and no segmentation point was obvious in the visual inspection of the running cumulative sum. For the effect size of TFS vs. TP, the EVPPI using both methods started becoming positive at WTP of \$150,000 (*Figure 4-4*). The EVPPI was maximum at WTP=\$260,000 when the PSA data was based on the Bayesian bootstrap, and was maximum at WTP=\$250,000 when the PSA data was based on the approximate Bayesian bootstrap.

Figure 4-4: EVPPI as a function of willingness-to-pay for the effect size of TFS vs. TP, using Bayesian and approximate Bayesian bootstrap



EVPPI: expected value of partial perfect information. TP: Tiotropium + placebo, TS: tiotropium + salmeterol, TFS: tiotropium + fluticasone/salmeterol.

For illustration, the scatter plots of  $I\widehat{NB}$  vs.  $\hat{\theta}$  (left panel), and the  $\hat{S}$  (right panel) for WTP=\$50,000 and WTP=\$250,000 and for both the Bayesian and the approximate Bayesian methods are shown in *Figure 4-5*. Again, the  $\hat{S}$  plot provides a powerful visual tool for checking if the EVPPI is positive, with no extremum (indicating a zero EVPPI) on the top two panels and a clear extremum on the bottom two panels.





#### 4.6 Implementation

I have developed an add-in for Microsoft Excel (version 2007, Microsoft Corporation, Redmond, WA, USA) that performs EVPPI calculations from the PSA data. The program requires random samples from the parameter of interest as well as the corresponding vector of net benefits for each strategy (if one vector for net benefit is provided, the program assumes it is the vector of incremental net benefits for two strategies). The calculations are based on equation (7), finding up to one segmentation point on all possible pairs of strategies. I have also provided a set of R functions that carries out the analysis and provides graphical output including the plot of *S* for visual inspection. The Visual Basic program for the Excel add-in and the R function are provided in the *Appendix A.4* and *Appendix A.5*, respectively.

#### 4.7 Discussion

While EVPPI provides an easily interpretable measure of the decision uncertainty, the conventional two-step Monte Carlo simulation for EVPPI calculation is computationally intensive. In this chapter I explained a novel method for single-parameter EVPPI calculation, proved its consistency, and empirically demonstrated its performance. I consider this approach to be heuristic as it relies on some assumptions (namely, the  $\delta$  function being piecewise constant with finitely many pieces) on the shape of the net benefit function that, while plausible, are not guaranteed to hold for all instances. However, when these assumptions hold, this algorithm generally performs much faster than the two-level Monte Carlo method to achieve the same precision. An additional computational advantage is that once the suitable data are generated, it can be used for EVPPI calculations for all stochastic parameters of the

model. Such a computational advantage could be considerable especially for complex models with many parameters. Besides being more efficient, this approach is simpler, since it only requires data that are generated through PSA- data that are a standard output in any stochastic cost-effectiveness analysis. This is unlike the classic and almost all other alternative EVPPI calculation methods that need either to be built in the modeling software or be implemented by the analyst. For RCT-based CEAs, the present method provides an entirely new opportunity for calculating the EVPPI for aspects of the evidence, such as the treatment effect size, that are aggregate statistics of the RCT data. Overall, the proposed approach facilitates single-parameter EVPPI calculation as a by-product of the probabilistic sensitivity analysis with very little additional computational cost. I hope this will encourage researchers to report the results of EVPPI analysis which are the most analytically rigorous way of uncertainty analysis in decision making <sup>50</sup>.

The presented technique is also insensitive to first-level uncertainty, meaning that in modelbased CEAs it can equally be used in the individual-level (microsimulation) models. This is because the algorithm only needs an unbiased estimator of NB. In patient-level simulations, the NB generated for each simulated individual is an unbiased estimator of the population NB provided that the individual-level covariates are sampled from distributions that represent their variation among the target population.

The present method is not the first method for alternative computation of EVPPI. It can be contrasted with four different methods of EVPPI calculations reviewed by Coyle et al. <sup>112</sup>. Aside from the generic two-step Monte Carlo method, they also reviewed the Unit Normal Loss

Integral (UNLI) method, a single-step Monte Carlo method, and a quadrate method based on numerical integration of the outer expectation, which is especially wieldy for one-parameter EVPPI<sup>12</sup>. The UNLI and single-step simulation are only valid for special cases. The quadrature method seems to be especially comparable with our method. While the quadrature method requires less computation than the two-step approach, the outer integration part should yield high coverage of the probability distribution of the parameter of interest, and the inner Monte Carlo simulation sample should be large enough to minimize the bias caused by maximization. Meanwhile, unlike our method, the quadrature method cannot use the same set of data to calculate EVPPI for different parameters.

This approach could also be compared with the 'meta-modeling' techniques such as the Gaussian process modeling <sup>72,113,125</sup>. Implementation and interpretation of the results provided by such methods require considerable expertise. Meanwhile, for data of size n, the Gaussian process requires operation on matrices of size ( $n \times n$ ), making it practically impossible to work with large data (e.g. n=10,000) unless further approximations are made. However, the Gaussian process does have the advantage that once developed and tested, it can be used to calculate multi-parameter EVPPI.

Recently, Strong et al. developed a method for EVPPI calculation which is operationally very similar to the method presented here <sup>121</sup>. The idea in their approach is to partition the domain of the parameter of interest to an arbitrary number of intervals, and instead of performing a nested Monte Carlo simulation at a given value of  $\theta$ , simply average the associated NBs for all

<sup>&</sup>lt;sup>12</sup> the fifth, the difference method, as the authors explained, is not a valid method for EVPPI calculation

the  $\theta$  within the same segment (see equations (3) and (4) in their report). This obviates the need for the two-level Monte Carlo simulation and allows for the calculation of the EVPPI from the PSA data. They discuss that for the small (large) number of segments, their algorithm has a downward (upward) bias, but show that for a wide range of the number of segments, it generates estimated EVPPIs close to the real ones (see Figure 1 in their report). The method presented in this chapter can be seen as an improved version of this approach as instead of partitioning the domain of  $\theta$  into an arbitrary number of segments, it chooses the segmentation points only at relevant points (corresponding to discontinuity points in  $d^*$ ). I also believe the formal proof of the convergence in probability of the estimator presented in this chapter is more rigorous than the theoretical justification provided by Strong et al. based on the linear approximation of the net benefit function in small intervals (See Appendix B in their report) <sup>121</sup> which does not amount to a proof for the convergence of their proposed estimator to its true value.

The foremost shortcoming of this approach is that it can only be used for one parameter at a time. This is because with more than one parameter the net benefit is a 'surface' function of model parameters, not a one-dimensional curve, and crossing points of surfaces are not a finite set of points. As a note of caution, there is no additivity rule for multivariate EVPPI, meaning that one cannot simply sum (or use other arithmetic operations on) EVPPIs for individual parameters to calculate joint EVPPI for a set of parameters <sup>126</sup>. Convergence of the novel EVPPI estimator in probability to its true value means it is at least an asymptotically unbiased estimator. However, as the results of the simulations indicate, it has asymptotic upward bias, and such bias is larger in smaller samples. This bias is generated by the noise in the data around

the segmentation points, as the optimization step will inevitably capture some such noise towards overestimating EVPPI. Adjusting for such bias will remain the focus of further research. The empirical results indicate that with PSA sizes of 10,000 and above such bias has negligible effect.

There is broad consensus that EVPPI should be presented in the results of economic evaluations, but given complexities in calculation and computation, few studies have reported such results. I presented a practically simple method for calculating single-parameter EVPPI which we believe can eliminate these hurdles. Our method has been tested on only a small set of sample models, and so its efficiency in more complex models needs to be explored. The development of methods for bias adjustment and statistical selection of the number of segments could improve the results further. Extensions to this concept for calculation of similar metrics such as the expected value of sample information (EVSI) should also be added to the research agenda.

#### **Chapter 5. Integrated discussion and conclusions**

#### 5.1 Main contributions

The objective of this thesis has been to improve the methodology and accessibility of methods for the CEA and EVI analysis of health technologies. In this regard, several contributions were made, especially for the economic evaluations conducted alongside a single RCT. The vetted bootstrap, proposed in Chapter 2, allows for the incorporation of external evidence in a simpleto-implement scheme. The non-parametric EVI analysis method, explained in Chapter 3, provides a new method for EVSI calculation and a theoretical justification of using the bootstrap for EVPI calculation. The heuristic algorithm, explained in Chapter 4, is a very fast and easy-to-use method for single-parameter EVPPI calculation that is solely based on the PSA data. Each of the three studies in this thesis contributes a novel approach to the toolbox for the CEA and EVI analyses. Whenever the incorporation of external evidence in RCT-based CEAs has been sought, a parametric Bayesian framework has been chosen <sup>127</sup>, a departure from the popular bootstrap method in RCT-based CEAs. For the EVI analysis of RCT-based CEAs, previous authors have explicitly acknowledged the lack of non-parametric methods. Koerkamp et al. mention that "unfortunately, it is not obvious how bootstrapping should be implemented to estimate partial value of information and sample information"<sup>86</sup>. Meltzer et al. remark that "the absence of a decision model makes it impossible to calculate the expected value of perfect partial information on specific parameters that may be partial determinants of the outcomes of the interventions being considered"<sup>128</sup>. The methods presented in Chapters 3 and 4 are important steps towards addressing such problems. Meltzer et al. called minimal modeling approaches to

EVI analysis as "those that model [value of information] without constructing a decision model of the disease and treatment process to characterize the uncertainty in net benefit associated with an intervention" <sup>128</sup>. The content of Chapters 3 and 4 indeed fit well within this definition.

I believe the set of methods developed in this thesis have improved the accessibility of stochastic CEA and EVI analysis for the applied health economists. All three methods are based on the Monte Carlo simulation for estimating the cost and effectiveness outcomes using the bootstrap, an approach that is very common among practical economists to quantify uncertainty in RCT-based CEAs. Furthermore, the vetted and weighted bootstrap and the single-parameter EVPPI calculation methods can potentially be employed after the PSA data is generated, without any need for manipulating the process that has generated the data (the only extra step is that the relevant parameter should be recorded throughout the bootstrap loops). This allows for an 'offline' use of such methods without any tampering with the original RCT-based CEA, therefore making Bayesian evidence synthesis and EVI analysis available at low implementation or computation costs.

The fundamental building block of this thesis is the Bayesian interpretation of the bootstrap <sup>63,64</sup>. The Bayesian interpretation of the bootstrap has been the basis of several popular statistical procedures such as non-parametric missing-value imputation and the weighted likelihood approach <sup>99,100</sup>. In the CEA too, quantities such as the probability of cost-effectiveness in acceptability curves, or the probability of belonging to a particular quadrant in the cost-effectiveness plane, treat the unobserved population values of cost and effectiveness outcomes as random entities, and thus invite a Bayesian interpretation <sup>33,129</sup>. EVPI calculation using the

bootstrap strictly mandates a Bayesian interpretation <sup>86</sup>. Therefore, when bootstrapping is used in such contexts, it can be interpreted in its Bayesian paradigm. I explicitly quantified a population distribution, which allowed Bayesian evidence synthesis and justified subsequent resampling in two-level EVSI calculation.

This underlying common theme allows the vetted bootstrap approach to be readily incorporated in the EVPI and EVSI calculations (Chapter 3) and the single-parameter EVPPI calculation (Chapter 4). For example, the external information obtained from the literature in Chapter 2 regarding the effect size of the medications can be used in calculating the EVPI and EVSI, as outlined in Chapter 3. This can easily be performed by applying the acceptance/rejection algorithm in the first-level bootstrap, and performing the EVPI and EVSI calculations on the bootstrap sets that have been accepted in the first level. Similarly, one can generate PSA data consisting of all the bootstrap sets that have successfully passed through the acceptance/rejection step for EVPPI calculation without any modification to the EVPPI calculation method developed in Chapter 4.

#### 5.2 Limitations and future research

As with almost all other research, there are some limitations in the methods developed in this thesis. These methods are based on the non-parametric bootstrap. Nevertheless, the non-parametric nature of the bootstrap does not ensure robustness and freedom from assumptions. The apparent simplicity of the bootstrap has the potential to conceal the strong assumptions being made, especially with small datasets <sup>63,130</sup>. For one, both Bayesian and approximate Bayesian bootstrapping methods assume that the population distribution can only

generate the observed data and any other data has zero probability of occurrence <sup>63</sup>. There are modified versions of bootstrapping that can address this problem and might be considered in this context <sup>58,131</sup>. Generally, the bootstrap approach to statistical analysis is based on two asymptotics: the RCT data should be infinitely large, and the number of bootstrap loops should be infinite <sup>53</sup>. Departure from such asymptotic conditions in reality is inevitable and, as such, the results are always an approximate.

The vetted and weighted bootstrap methods can quickly become unwieldy if there are too many parameters for which external evidence is available, or when the distribution of the external evidence is in conflict with the evidence generated by the trial. It might be possible to develop more efficient bootstrap sampling methods. One possibility is to manipulate Rubin's prior distribution for the Bayesian bootstrap to carry the information on the parameters of interest, obviating the need for rejection or importance sampling used in Chapter 2. Unfortunately, this does not seem to be an easy extension, as the methods proposed so far result in posterior distributions that put non-zero probability on the unobserved values (therefore the bootstrap will no longer remain a valid sampling mechanism) or result in degenerate situations (e.g., posterior variance growing without bounds as the prior variance approaches infinity regardless of the amount of information in the data)<sup>111</sup>. Another interesting opportunity is to create auto-correlated Markov Chain bootstraps that tend to concentrate on the high probability areas of the posterior distribution, which result in fewer rejections. One interesting example is the work by Yu et al. in which a Markov chain bootstrap is constructed for efficient p-value evaluation in genetic association studies <sup>132</sup>.

An important area that needs further research is covariate adjustment of unbalanced covariates in the context of the bootstrap method of RCT-based CEA. The *ad hoc* process used for adjusting covariates in EVSI calculations in Chapter 3 is inefficient, only partially adjusts for the covariate imbalance, and can be justified only in the asymptotic case. Alternative candidates for covariate adjustment might be considered such as matching individuals between treatment arms based on important covariates, but such matching should be fully justifiable within the Bayesian framework for the EVI calculations. Yet another interesting alternative is to use ideas from the missing value imputation literature. One can assume for the *i*<sup>th</sup> patient in the *j*<sup>th</sup> arm, the data had the patient been assigned to the  $k^{th}$  ( $k \neq j$ ) arm is missing. Such missing data can be imputed from the sample in the  $k^{th}$  arm from the pool of patients that have similar values of the covariates.

For all three approaches presented in Chapters 2-4, there are alternative parametric methods (or one could conceive of such methods) to carry out the same objective. However, the proposed methods are novel enough that I felt the focus should be on developing the theory and detailing the context in which they are applicable, and simulation studies were mainly used to establish analogies with parametric methods and demonstrate the similarity in the results between the new and established techniques. The next logical step is to comprehensively examine the performance of the methods presented in this thesis versus the alternative approaches using simulation studies. The attractive feature of the simulation studies is that the analyst is in control of the data generation mechanism. The assumptions underlying the parametric model can be adjusted to match the mechanism that generates the data. Subsequently, the data generation mechanism can be tweaked to be intentionally in conflict

with the parametric model. This will allow comparison of the performance of the nonparametric versus both correctly and incorrectly specified parametric methods, and will help detect situations in which parametric or nonparametric methods are superior.

#### 5.3 Putting this research in context

For the EVPI and EVSI calculations, the method developed in Chapter 3 competes with normal approximation methods. For the specific case of two-arm trials, there are closed-form equations for EVPI and EVSI calculations that can be used for an efficient EVI analysis, and the results are not subject to Monte Carlo uncertainty <sup>71</sup>. This method, unfortunately, is not currently applicable to RCTs with more than two arms and the analyst should resort to Monte Carlo simulation. Still, Monte Carlo simulation based on sampling from a normal distribution will be faster than resampling-based methods that work on the raw RCT data. In addition, the normal approximation methods for EVI analysis are well integrated in realistic decision contexts and are extended to tackle issues such as imperfect implementation, transferability of results across jurisdictions, and so on <sup>92</sup>. There is no reason to believe non-parametric EVSI methods cannot be expanded likewise, but this remains to be explored in future studies.

Eckermann et al. consider the normal approximation approach to EVI calculation as the Occam's razor in EVI analysis <sup>92</sup>. This argument is made on the basis of the ease of use of this method and it being on par with or better than bootstrap methods in EVPI calculation, especially in small samples and because of *"their [more complex EVI methods such as bootstrapping] increased complexity and current limitations in informing decision making, with restriction to EVPI rather than EVSI and not allowing for important decision-making contexts"* <sup>92</sup>.

The method presented in Chapter 3 removes the limitation of non-parametric methods for EVSI calculation, and while I agree that the simplicity and ease of use of the normal methods are very attractive, their comparative accuracy for EVSI calculations in realistic scenarios remains to be studied. This task does not seem to be an easy undertaking given that there is no gold standard for EVSI calculations for any realistic data. All in all, as in many other areas in statistics, parametric and non-parametric methods can coexist and each can be used in certain situations. One example is the imputation of missing data in which methods based on normal approximation and non-parametric methods based on bootstrap are both popular. The non-parametric bootstrap method of missing data imputation <sup>65,99</sup> is specifically similar to the two-level approach for EVSI calculation outlined in Chapter 3. This is not a coincidence, as I believe EVI calculations are connected to the missing data problem; one can see the data of the future trial as missing data that can be imputed based on the current RCT.

The method presented in Chapter 4 can be compared to the method of EVPPI calculation based on partitioning the domain of the parameter of interest recently developed by Strong et al. <sup>121</sup>. The two methods are operationally similar. However, the method by Strong et al. is based on an arbitrary partitioning of the domain of the parameter of interest given that for a wide range of the number of partitions, the estimated EVPPI seems to remain close to the true EVPPI (see Figure 1 in their report). The present method can be seen as an improved version of the method by Strong et al. as partitioning is performed only at the 'necessary' points in which the net benefit functions of different decisions cross. This almost surely increases the precision of the estimates. However, finding such segmentation points requires a numerical maximization

algorithm which slightly increases its computational demand and the complexity of the software implementation compared to the method proposed by Strong et al.

#### 5.4 Knowledge transfer and exchange

Knowledge transfer and exchange (KTE) is an interactive interchange of knowledge between research users and researcher producers <sup>133</sup>. If the target audience does not become aware of the research results and put them into practice, then the time and resources spent on the research will be wasted. Given the importance of KTE, and to elaborate on the potential steps needed to be taken in order to ensure the uptake of the research developed in this thesis, I applied the framework suggested by Lavis et al. <sup>134</sup>. This framework provides a list of five elements to consider when organizing KTE: message, target audience, messenger, knowledge transfer process and support system, and evaluation strategy.

## What do research organizations transfer to their target audiences, and at what cost (the Message)?

The present thesis elaborates on the research that makes stochastic CEA and EVI analyses more accessible. The first core message to transfer is that the bootstrap method of RCT-based CEA has the capacity to be used in situations that so far have been considered to be exclusive to parametric modeling; and second, the set of specific methods that carry such additional information, as well as the computer programs developed for this purpose. The KTE literature emphasizes that the message should be "action-able" <sup>134</sup>. In this context, this can be interpreted as a message that results in the change of practice by the end user. This is most

relevant for the applied economist who can conceivably use the aforementioned methods instead of, or along with, alternatives.

# To whom do research organizations transfer research knowledge, and with what investments in targeting them (the Target)?

To improve the use of research, researchers must first decide who their audience is. Willson et al. suggest that each audience has different information needs and communication styles and therefore the information must be appropriately tailored <sup>135</sup>. In this regard, I recognize two distinct audience as the target of my research: the applied analysts (health economists and biostatisticians involved in RCT-based CEA research), and theoretical scientists (biostatisticians and statisticians developing methodology for RCT-based CEAs). The applied analyst uses the methods for the analysis of their data, and the theorist further develops the theory behind the proposed approaches to improve the methods and to create new ones.

## By whom is the research knowledge transferred, and with what investments in assisting them (the Messenger)?

The natural avenue for propagating this research to the target audience is the publication of the results in peer-reviewed journals and presentations at conferences. This is probably the most efficient method of dissemination of the knowledge generated in this work to the theoretical scientists. For the applied analysts, however, I will use alternative methods of communication and engagement. It is known that when researchers have the skills and experience to act as the principal messenger, their credibility will likely make them the ideal choice <sup>134</sup>. I can also seek the engagement of trusted intermediaries that can act as *knowledge brokers*. Among those candidates, I specially think professional bodies such as the *Canadian Agency for Drugs and Technologies in Health* (CADTH) and the *International Society For Pharmaco-economics and Outcomes Research* (ISPOR) are viable messengers for my research. These agencies periodically publish guidelines and recommendations and I hope the knowledge generated throughout my thesis research will be reflected in such reports in the future. It is already shown that the uptake of research is influenced by the reputation and credibility of the messenger such as the authoritative organizations representing professional groups <sup>136</sup>.

# How do research organizations engage target audiences in the research process (the Process and supporting systems)?

For the applied analysts, I will not restrict myself to the traditional academic language of peerreviewed publication. Non-traditional communication channels will be used. I have already presented the method in Chapters 2 and 4 in conference proceedings, which enables more face-to-face engagement of the applied analyst. I will avoid information overload, which is a known barrier for engaging the target audience; and will emphasize on the presentation of the summary results in simple language, and with clearly worded recommendations <sup>137</sup>. With the collaboration of some of my colleagues, I have already held a workshop about the method developed in Chapter 4 for the CADTH (<u>http://core.ubc.ca/software/voi/</u>), and will continue pursuing such opportunities to more directly engage with the relevant audience for this work.

In addition, I have developed a Microsoft Excel add-in and a set of R functions as a generic implementation of the algorithm that can work with PSA data of either RCT-based or model-

based CEAs with minimal requirement for the consideration of the specific settings of the original study. Excel is a very popular platform for decision analysis. R is a popular free and open source statistical computing program. The choice of these two software platforms is to ensure the maximum availability of the framework for the target audience.

# Do research organizations perform evaluative activities related to knowledge transfer (Evaluation strategy)?

This element of Lavis' framework for KTE mainly applies to the organizations that sponsor research at a larger scale <sup>134</sup>. Nevertheless, there are certain opportunities in getting feedback from the target audience in this research as well. Such activities can include face-to-face communication with the target users which makes the transfer of knowledge a two-way communication <sup>138</sup>. The target audience for my research is the relatively small community of theorists and applied health economists, and the usefulness of such approach in providing mutual feedback between the research users and research producers in such settings is already established <sup>138</sup>. In my future workshops and presentations, I will also try to implement a beforeafter survey to evaluate the uptake of the knowledge by the audience.

#### 5.5 Concluding remarks

In the world of frozen budgets, escalating costs of new health technologies, and soaring costs of RCTs, decision makers should increasingly be efficient in the decisions they make with regard to the allotment of resources. Making efficient adoption decisions is informed through CEA, while making efficient research prioritization is performed through the EVI analysis. Nevertheless,

stochastic CEA which allows the incorporation of evidence uncertainty in the analysis, as well the EVI analysis, are relatively young disciplines and there is much room for their improvement. This thesis has been an attempt in enriching the toolkit of stochastic CEA and EVI analysis, with an eye on the needs of practical economists that use resampling methods for RCT-based CEAs. Methods developed in this thesis seem to have applicability for the large camps of health economists who embark on the CEA and EVI analysis of RCTs. The research in this thesis was a bold step in territories unfamiliar to many health economists, and I hope the methods developed in this thesis, and the inevitable gaps remained, will stimulate further research in this area.

### Bibliography

1. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials; the ISPOR RCT-CEA Task Force report. *Value in health*. 2005;8(5):521-33.

2. Aaron S, Vandemheen K, Fergusson D, et al. Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: a randomized trial. *Ann. Intern. Med.* 2007;146(8):545-555.

3. Najafzadeh M, Marra C, Sadatsafavi M, et al. Cost effectiveness of therapy with combinations of long acting bronchodilators and inhaled steroids for treatment of COPD. *Thorax*. 2008;63(11):962-967.

4. Drummond M, O'Brien B, Stoddart G, Torrance G. *Methods for the Economic Evaluation of Health Care Programmes*. United Kingdom: Oxford University Press; 2005.

5. Palmer S, Byford S, Raftery J. Types of economic evaluation. BMJ. 1999;318(7194):1349.

6. Elliott R. *Essentials of economic evaluation in healthcare*. London: Pharmaceutical Press; 2005.

7. Neumann P, Goldie S, Weinstein M. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21:587-611.

8. Phelps C, Mushlin A. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *Int J Technol Assess Health Care*. 1991;7(1):12-21.

9. Gyrd-Hansen D. Willingness to pay for a QALY: theoretical and methodological issues. *Pharmacoeconomics*. 2005;23(5):423-432.

10. Mitra N, Indurkhya A. A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data. *Health Econ*. 2005;14(8):805-815.

11. O'Brien B. Economic evaluation of pharmaceuticals. Frankenstein's monster or vampire of trials? *Med Care*. 1996;34(12 Suppl):DS99-108.

12. Roland M, Torgerson DJ. Understanding controlled trials: What are pragmatic trials? *BMJ*. 1998;316(7127):285.

13. Glick H, Doshi J, Sonnad S, Polsky D. *Economic Evaluation in Clinical Trials*. New York: Oxford University Press; 2007.

14. Campbell D, Stanley J. *Experimental and Quasi-Experimental Designs for Research*. 1st ed. Chicago: Wadsworth Publishing; 1963.

15. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303):42-46.

16. Steckler A, McLeroy K. The importance of external validity. *Am J Public Health*. 2008;98(1):9-10.

17. Ades A, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Econ*. 2006;15(4):373-381.

18. Smith A, Ryan P, Evans J. The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Anal.* 1992;12(4):467-474.

19. Garber A, Phelps C. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*. 1997;16(1):1-31.

20. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. *BMJ*. 1993;306(6877):570-572.

21. Olsen J, Smith R. Theory versus practice: a review of "willingness-to-pay" in health and health care. *Health Economics*. 2001;10(1):39-52.

22. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a qualityadjusted life year: in search of a standard. *Med Decis Making*. 2000;20(3):332-42.

23. Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in costeffectiveness analysis. *Int J Technol Assess Health Care*. 1998;14(3):467-471.

24. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making*. 1998;18(2 Suppl):S68-80.

25. Sculpher M, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Econ.* 2006;15(7):677-687.

26. Ades A, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in costeffectiveness analysis. *Pharmacoeconomics*. 2006;24(1):1-19.

27. Spiegelhalter D, Freedman L, Parmar M. Bayesian Approaches to Randomized Trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 1994;157(3):357-416.

28. Spiegelhalter D, Abrams K, Myles J. *Bayesian approaches to clinical trials and health care evaluation*. Chichester: John Wiley & Sons; 2004.

29. Berry D. A case for Bayesianism in clinical trials. *Stat Med*. 1993;12(15-16):1377-1393; discussion 1395-1404.

30. Simon R. Bayesian design and analysis of active control clinical trials. *Biometrics*. 1999;55(2):484-487.

31. Brophy J, Joseph L. Placing trials in context using Bayesian analysis. GUSTO revisited by Reverend Bayes. *JAMA*. 1995;273(11):871-875.

32. O'Brien B, Drummond M, Labelle R, Willan A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Med Care*. 1994;32(2):150-163.

33. Briggs A. A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ*. 1999;8(3):257-261.

34. Heitjan D, Moskowitz A, Whang W. Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Econ*. 1999;8(3):191-201.

35. Willan A, Pinto E, O'Brien B, et al. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. *Health Econ.* 2005;14(4):327-338.

36. Manca A, Rice N, Sculpher M, Briggs A. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Econ.* 2005;14(5):471-485.

37. Al M, Van Hout B. A Bayesian approach to economic analyses of clinical trials: the case of stenting versus balloon angioplasty. *Health Econ*. 2000;9(7):599-609.

38. O'Hagan A, Stevens JW. A framework for cost-effectiveness analysis from clinical trial data. *Health Econ.* 2001;10(4):303-315.

39. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost-effectiveness analysis from clinical trial data. *Stat Med*. 2001;20(5):733-753.

40. O'Hagan A, Stevens JW. Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Stat Methods Med Res.* 2002;11(6):469-490.

41. Nixon R, Thompson S. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Econ*. 2005;14(12):1217-1229.

42. Heitjan D, Kim C, Li H. Bayesian estimation of cost-effectiveness from censored data. *Stat Med*. 2004;23(8):1297-1309.

43. Heitjan D, Li H. Bayesian estimation of cost-effectiveness: an importance-sampling approach. *Health Economics*. 2004;13(2):191-198.

44. Fenwick E, Wilson J, Sculpher M, Claxton K. Pre-operative optimisation employing dopexamine or adrenaline for patients undergoing major elective surgery: a cost-effectiveness analysis. *Intensive Care Med*. 2002;28(5):599-608.

45. UK BEAM Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: cost effectiveness of physical treatments for back pain in primary care. *BMJ*. 2004;329(7479):1381.

46. Shih Y, Mauskopf J, Borker R. A cost-effectiveness analysis of first-line controller therapies for persistent asthma. *Pharmacoeconomics*. 2007;25(7):577-90.

47. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000;10:325–337.

48. Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value Health*. 2009;12(4):409-418.

49. Briggs A. A Bayesian approach to stochastic cost-effectiveness analysis. An illustration and application to blood pressure control in type 2 diabetes. *Int J Technol Assess Health Care*. 2001;17(1):69-82.

50. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*. 1999;18(3):341-364.

51. Doubilet P, Begg CB, Weinstein MC, Braun P, McNeil BJ. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. *Med Decis Making*. 1985;5(2):157-177.

52. Briggs A, Wonderling D, Mooney C. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ*. 1997;6(4):327-340.

53. Willan A, Briggs A. Statistical analysis of cost-effectiveness data. John Wiley; 2006.

54. Chaudhary M, Stearns S. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Stat Med*. 1996;15(13):1447-1458.

55. Hoch J, Briggs A, Willan A. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ.* 2002;11(5):415-430.

56. Nixon R, Wonderling D, Grieve R. Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. *Health Econ*. 2010;19(3):316-333.

57. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ.* 1998;7(8):723-740.

58. Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. 1979;7(1):1-26.

59. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. 1st ed. New York: Chapman and Hall/CRC; 1994.

60. O'Brien B, Briggs A. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. *Stat Methods Med Res.* 2002;11(6):455-468.

61. Mihaylova B, Briggs A, O'Hagan A, Thompson S. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011;20(8):897-916.

62. Thompson S, Nixon R. How sensitive are cost-effectiveness analyses to choice of parametric distributions? *Med Decis Making*. 2005;25(4):416-423.

63. Rubin D. The Bayesian Bootstrap. Ann. Statist. 1981;9(1):130-134.

64. Lo A. A Large Sample Study of the Bayesian Bootstrap. Ann. Statist. 1987;15(1):360-375.

65. Schafer J. Multiple imputation: a primer. *Statistical Methods in Medical Research*. 1999;8(1):3-15.

66. Briggs A. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*. 2000;17(5):479-500.

67. Felli J, Hazen G. Sensitivity analysis and the expected value of perfect information. *Med Decis Making*. 1998;18(1):95-109.

68. Ades A, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making*. 2004;24(2):207-227.

69. Claxton K, Sculpher M. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics*. 2006;24(11):1055-1068.

70. Eckermann S, Willan A. Expected value of information and decision making in HTA. *Health Econ*. 2007;16(2):195-209.

71. Willan A, Pinto E. The value of information and optimal clinical trial design. *Stat Med*. 2005;24(12):1791-1806.

72. Tappenden P, Chilcott J, Eggington S, Oakley J, McCabe C. Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-beta and glatiramer acetate for multiple sclerosis. *Health Technol Assess*. 2004;8(27):iii, 1-78.

73. Brennan A, Kharroubi S, O'Hagan A, Chilcott J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med Decis Making*. 2007;27(4):448-70.

74. Philips Z, Claxton K, Palmer S. The half-life of truth: what are appropriate time horizons for research decisions? *Med Decis Making*. 2008;28(3):287-299.

75. Yokota F, Thompson K. Value of information literature analysis: a review of applications in health risk management. *Med Decis Making*. 2004;24(3):287-298.

76. Yokota F, Thompson K. Value of information analysis in environmental health risk management decisions: past, present, and future. *Risk Anal*. 2004;24(3):635-650.

77. Hammitt J, Shlyakhter A. The Expected Value of Information and the Probability of Surprise. *Risk Analysis*. 1999;19(1):135-152.

78. Raiffa H, Sohlaifer R. *Applied statistical decision theory*. Cambridge, MA: Harvard Business School; 1961.

79. Willan A, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ*. 2010;19(5):549-561.

80. Willan A, Kowgier M. Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clin Trials*. 2008;5(4):289-300.

81. Willan A. Clinical decision making and the expected value of information. *Clin Trials*. 2007;4(3):279-285.

82. Willan A. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clin Trials*. 2008;5(6):587-594.

83. Eckermann S, Willan A. The option value of delay in health technology assessment. *Med Decis Making*. 2008;28(3):300-5.

84. Eckermann S, Willan A. Globally optimal trial design for local decision making. *Health Econ*. 2009;18(2):203-216.

85. Koerkamp B, Hunink M, Stijnen T, Weinstein M. Identifying key parameters in costeffectiveness analysis using value of information: a comparison of methods. *Health Econ*. 2006;15(4):383-392.

86. Koerkamp B, Spronk S, Stijnen T, Hunink M. Value of Information Analyses of Economic Randomized Controlled Trials: The Treatment of Intermittent Claudication. *Value Health*. 2009.

87. Lachin J. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials*. 1981;2(2):93-113.

88. Pezeshk H. Bayesian techniques for sample size determination in clinical trials: a short review. *Stat Methods Med Res.* 2003;12(6):489-504.

89. Claxton K, Thompson K. A dynamic programming approach to the efficient design of clinical trials. *J Health Econ*. 2001;20(5):797-822.

90. Pindyck R. Irreversibility, Uncertainty, and Investment. *Journal of Economic Literature*. 1991;29(3):1110-1148.

91. Eckermann S, Willan A. Time and expected value of sample information wait for no patient. *Value Health*. 2008;11(3):522-526.

92. Eckermann S, Karnon J, Willan A. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics*. 2010;28(9):699-709.

93. Aaron S, Vandemheen K, Fergusson D, et al. The Canadian Optimal Therapy of COPD Trial: design, organization and patient recruitment. *Can. Respir. J.* 2004;11(8):581-585.

94. Meguro M, Jones P. Elicitation of Utility Weights for a COPD Specific Preference Measure. In: *ATS American Thoracic Society*. San Diego, CA; 2006.

95. O'Brien B. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. *Am J Manag Care*. 1997;3 Suppl:S33-39.

96. Eddy D. The Confidence Profile Method: A Bayesian Method for Assessing Health Technologies. *Operations Research*. 1989;37(2):210-228.

97. Obenchain R, Melfi C, Croghan T, Buesching D. Bootstrap analyses of cost effectiveness in antidepressant pharmacotherapy. *Pharmacoeconomics*. 1997;11(5):464-472.

98. Smith A, Gelfand A. Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *The American Statistician*. 1992;46(2):84-88.

99. Rubin D. Multiple imputation for nonresponse in surveys. New York: John Wiley; 1987.

100. Newton M, Raftery A. Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1994;56(1):3-48.

101. Von Neumann J. Various techniques used in connection with random digits. *Nat. Bureau Stand. Appl. Math. Ser.* 1951;12:36-38.

102. Marshall A. The use of multi-stage sampling schemes in Monte Carlo computations. In: *Meyer, M.A. (Ed.), Symposium on Monte Carlo Methods*. New York: Wiley; 1956:123-140.

103. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian Data Analysis*. London: Chapman & Hall; 1995.

104. Wang J, Jin D, Zuo P, et al. Comparison of tiotropium plus formoterol to tiotropium alone in stable chronic obstructive pulmonary disease: a meta-analysis. *Respirology*. 2011;16(2):350-358.

105. Welte T, Miravitlles M, Hernandez P, et al. Efficacy and tolerability of budesonide/formoterol added to tiotropium in patients with chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 2009;180(8):741-750.

106. Kass R, Greenhouse J. [Investigating Therapies of Potentially Great Benefit: ECMO]: Comment: A Bayesian Perspective. *Statist. Sci.* 1989;4(4):310-317.

107. Ibrahim J, Chen M. Power Prior Distributions for Regression Models. *Statistical Science*. 2000;15(1):46-60.

108. Ades A, Lu G, Higgins J. The Interpretation of Random-Effects Meta-Analysis in Decision Models. *Medical Decision Making*. 2005;25(6):646-654.

109. Robert C, Casella G. Monte Carlo statistical methods. Springer; 2004.

110. Cooper N, Sutton A, Abrams K, Turner D, Wailoo A. Comprehensive decision analytical modelling in economic evaluation: a Bayesian approach. *Health Econ*. 2004;13(3):203-226.

111. Walker S, Mallick B. A Note on the Scale Parameter of the Dirichlet Process. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*. 1997;25(4):473-479.

112. Coyle D, Oakley J. Estimating the expected value of partial perfect information: a review of methods. *Eur J Health Econ.* 2008;9(3):251-9.

113. Stevenson M, Oakley J, Chilcott J. Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. *Med Decis Making*. 2004;24(1):89-100.

114. Oakley J, Brennan A, Tappenden P, Chilcott J. *Sample Sizes for Monte Carlo Partial EVPI Calculations*. Department of Probability and Statistics, University of Sheffield, UK; 2007.

115. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011. Available at: http://www.r-project.org/. Accessed November 16, 2009.

116. Willan A, Eckermann S. Accounting for between-study variation in incremental net benefit in value of information methodology. *Health Economics*. 2011.

117. DasGupta A. *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*. New York: Springer; 2011.
118. Briggs A, Mooney C, Wonderling D. Constructing confidence intervals for costeffectiveness ratios: an evaluation of parametric and non-parametric techniques using Monte Carlo simulation. *Stat Med*. 1999;18(23):3245-3262.

119. Manca A, Hawkins N, Sculpher M. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487-496.

120. Durbin J. *Distribution theory for tests based on the sample distribution function*. No 9 in: SIAM regional conference series in applied mathematics (Arrowsmith, Bristol); 1973.

121. Strong M, Oakley J. An efficient method for computing partial expected value of perfect information for correlated inputs. United Kingdom: University of Sheffield; 2011. Available at: http://sheffield.ac.uk/polopoly\_fs/1.95249!/file/StrongOakleyEVPIWebVersion.pdf. Accessed October 2, 2011.

122. Oakley J, Brennan A, Tappenden P, Chilcott J. Simulation sample sizes for Monte Carlo partial EVPI calculations. *J Health Econ*. 2010;29(3):468-477.

123. Rosenthal J. A first look at rigorous probability theory. New York: World Scientific; 2006.

124. Billingsley P. In: Probability and Measure. 3rd ed. New York: John Wiley & Sons; 1995.

125. Rojnik K, Naversnik K. Gaussian process metamodeling in Bayesian value of information analysis: a case of the complex health economic model for breast cancer screening. *Value Health*. 2008;11(2):240-50.

126. Samson D, Wirth A, Rickard J. The value of information from multiple sources of uncertainty in decision analysis. *European Journal of Operational Research*. 1989;39(3):254-260.

127. McCarron E, Pullenayegum E, Marshall D, Goeree R, Tarride J. Handling Uncertainty in Economic Evaluations of Patient Level Data: A Review of the Use of Bayesian Methods to Inform Health Technology Assessments. *International Journal of Technology Assessment in Health Care*. 2009;25(04):546-554.

128. Meltzer D, Hoomans T, Chung J, Basu A. Minimal Modeling Approaches to Value of Information Analysis for Health Research. *Med Decis Making*. 2011.

129. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*. 2001;10(8):779-787.

130. Beran R. The Impact of the Bootstrap on Statistical Algorithms and Theory. *Statistical Science*. 2003;18(2):175-184.

131. Silverman B, Young G. The bootstrap: To smooth or not to smooth? *Biometrika*. 1987;74(3):469 -479.

132. Yu K, Liang F, Ciampa J, Chatterjee N. Efficient p-value evaluation for resampling-based tests. *Biostatistics*. 2011;12(3):582-593.

133. Kiefer L, Frank J, Di Ruggiero E, et al. Fostering evidence-based decision-making in Canada: examining the need for a Canadian population and public health evidence centre and research network. *Can J Public Health*. 2005;96(3):I1-40.

134. Lavis J, Robertson D, Woodside J, McLeod C, Abelson J. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Q*. 2003;81(2):221-248, 171-172.

135. Willison D, MacLeod S. The role of research evidence in pharmaceutical policy making: evidence when necessary but not necessarily evidence. *J Eval Clin Pract*. 1999;5(2):243-249.

136. Shonkoff J. Science, policy, and practice: three cultures in search of a shared mission. *Child Dev*. 2000;71(1):181-187.

137. Reimer B, Sawka E, James D. Improving research transfer in the addictions field: a perspective from Canada. *Subst Use Misuse*. 2005;40(11):1707-1720.

138. Mitton C, Adair CE, McKenzie E, Patten SB, Waye Perry B. Knowledge transfer and exchange: review and synthesis of the literature. *Milbank Q*. 2007;85(4):729-768.

# Appendices

## Appendix A.1: R code for stylized example of chapter 2.

#### Copying this code in the R environment will automatically generate the graph.

n<-100; #Number of data points</pre> mulTheta<-1; #Mean of theta in the observed sample
vLTheta<-10/n; #Variance of E(theta) in the observed sample</pre> slope<-2; intercept<-1; vE<-5/n; #Variance of the error around E(b); muLB<-intercept+slope\*muLTheta;</pre> vLB<-vLTheta\*slope^2+vE; #In order to generate data that exactly matches the properties described in the text, I found it easier to get samples from the joint distribution of (theta,b) which is bivariate normal rather than theta as normal and b as linear on theta. Consult any stat text book for who to derive parameters of the bivariate normal (theta,b) from a linear regression model; rho<-slope/sqrt(slope^2+vE/vLTheta); #correlation coefficient between (theta,b) muOTheta<-0.5; #Mean of prior distribution on theta vOTheta <- 0.1; #Variance of the prior distribution on theta mulTheta <- (muOTheta \*vLTheta +muLTheta \*vOTheta) / (vLTheta +vOTheta); #Fixed-effect normal method for the mean of posterior distribution of E(theta) vlTheta<-1/(1/v0Theta+1/vLTheta); #Fixed-effect normal method for the variance of posterior distribution of E(theta) mulB<-intercept+slope\*mulTheta; #Mean of the posterior distribution for E(b) with updated information on theta vlB<-(1-rho\*rho)\*vLB+vLB/vLTheta\*rho\*rho\*vlTheta; #Variance of the posterior distribution for E(b) with updated information on theta ##############generating sample of theta and b with desired properties <math display="inline">##################nSim<-10000; #Number of requested vetted bootstraps #Here I generate random samples of (theta,b) with sample mean and variance as described in the text. Note that the linear model of b on theta is transformed to the joint bivariate normal distribution of (theta,b) so that direct manipulation of the covariance matrix gives the desired properties for the sample; theta<-rnorm(n,0,1);</pre> theta<-theta-mean(theta); #Set mean to zero m<-solve(cov(x));</pre> m.eig <- eigen(m) m.sqrt <- m.eig\$vectors %\*% diag(sqrt(m.eig\$values)) %\*% solve(m.eig\$vectors)</pre> x<-x\*\*%m.sqrt; #This transformation generates a bivariate sample with zero mean and covariaqnee [1 0; 0 1]. type cov(x) to check; m<-c(vLTheta\*n, rho\*sqrt(vLB\*n\*vLTheta\*n), rho\*sqrt(vLB\*n\*vLTheta\*n), vLB\*n);</pre>  $\dim(m) < -c(2,2);$ m.eig <- eigen(m) m.sgrt <- m.eigSvectors %\*% diag(sgrt(m.eig\$values)) %\*% solve(m.eig\$vectors)</pre> x<-x%\*%(m.sqrt); x[,1]<-x[,1]+muLTheta; x[,2]<-x[,2]+muLB;</pre> f = 1 have finally generated bivariate sample x for (theta,b) with the exact distributional assumptions outlines in the text. Type r < -lm(x[,2] ~ x[,1]); then type coefficients(r); then type var(residuals(r)); to check #Creating the vectors for data W<-rep(0,nSim); B<-rep(0,2\*nSim);</pre> #B will hold the mean of bootstrap samples  $\dim(B) < -c(nSim, 2)$ ; VB<-B; #VB will hold the mean of vetted bootstrap samples; #Regular bootstrap; count < -0: while (count < nSim) xBS<-x[sample(1:n,n,replace=TRUE),]; #Bootstrap sample of x</pre> count<-count+1; B[count,]<-colMeans(xBS); #record the mean of each bootstrap sample</pre> #Vetted bootstrap; wMax<-dnorm(mu0Theta,mean=mu0Theta,sd=sqrt(v0Theta)); #Maximum weight (wMax in the text) count<-0; while(count<nSim) xBS<-x[sample(1:n,n,replace=TRUE),];</pre> #Bootstrap sample of x w<-dnorm(mean(xBS[,1]),mean=mu0Theta,sd=sqrt(v0Theta));#w
if(runif(1)<w/wMax) #This is the vetting part</pre>

```
{
       count<-count+1;
       VB[count,]<-colMeans(xBS);</pre>
   }
}
#Graph space. (note that if you hange values earlier in the code then xlim and ylim should change to accomodate the graphs) xlim<-c(-1,6);
ylim<-c(-1,1.5);
curve(dnorm(x,mean=muLB,sd=sqrt(vLB)), from=muLB-3*sqrt(vLB),to=muLB+3*sqrt(vLB), xlim=xlim, ylim=ylim, lty=2, xlab="", main="");
#Likelihood function of the mean of b
lines(xlim,c(0,0));
par(new=T);
#Prior info dist
par(new=T);
curve(-dnorm(x,mean=mulB,sd=sqrt(v1B)), from=mulB-3*sqrt(v1B),to=mulB+3*sqrt(v1B), xlim=xlim, ylim=ylim, xlab="", ylab="",
main="");
par(new=T);
hist(B[,2],20, freq=FALSE, xlim=xlim, ylim=ylim, xlab="", ylab="", main="");
z<-hist(VB[,2],20, plot=FALSE);
#Reverse the histogram to show it below the axis;
for(i in 1:(length(z$breaks)-1))
{
   z$density[i]<--z$density[i];
z$intensities[i]<--z$intensities[i];
z$counts[i]<--z$counts[i];</pre>
par(new=T);
plot(z, col=rgb(0.8,0.8,0.8), freq=FALSE, xlim=xlim, ylim=ylim, xlab="", ylab="", main="");
```

## Appendix A.2: R code for stylized example of chapter 3.

Copying this code in the R environment will automatically generate the graph.

```
BS_TYPE_MULTINOMIAL<-1;
BS_TYPE_DIRICHLET<-2;
#Willan-Pinto method for EVSI calculation;
evsiWP<-function(b0,v0,sigma,n)
     v1=1/(1/v0+n/2/sigma^2)
    #b1<-v1*(b0/v0+n*b_hat/2/sigma^2);
sigma_b_hat<-sqrt(v0+2*sigma^2/n);</pre>
    A<-v1*n^2/4/sigma^4+1/sigma_b_hat^2;
    B<-(v1*n/v0/sigma^2-2/sigma_b_hat^2)*b0;
C<-(v1/v0^2+1/sigma_b_hat^2)*b0^2;</pre>
    D<-1/2/pi*sqrt(v1/sigma_b_hat^2);
    I1<-D*sqrt(2*pi/A)*exp(B^2/8/A-C/2);
    a<-n*sqrt(v1*sigma_b_hat^2)/2/sigma^2;
b<--n*sqrt(v1)*sigma_b_hat^2*b0/2/sigma^2/v0;</pre>
    I2<-sqrt(v1)*(b*pnorm(b/sqrt(a^2+1))+a^2/sqrt(2*pi*(a^2+1))*exp(-b^2/2/(a^2+1)));
    I3<-b0*pnorm(-b0*sqrt(sigma_b_hat^2)/v0)-v1*n*sqrt(sigma_b_hat^2)/(2*sigma^2*sqrt(2*pi))*exp(-b0^2*sigma_b_hat^2/2/v0^2);
    EVSI<-I1+I2+I3;
    return(Di(b0,v0)-EVSI);
}
#Willan-Pinto method for EVPI calculation;
Di<-function(b0,v0)
{
     return(sqrt(v0/2/pi)*exp(-b0^2/2/v0)-b0*(pnorm(-b0/sqrt(v0))-(b0<=0)*1));
b0<-68.97:
v0<-3724.78;
sigma<-sqrt(217227);
n<-100;
wtp<-1000;
#Non-parametric EVPI and EVSI calculations using the two-level resampling;
eviNP<-function(n,nSim,bsType=BS_TYPE_MULTINOMIAL)
    if(bsType==BS_TYPE_MULTINOMIAL)
    {
         ptDataBS<-rbinom(nSim,116,41/116)/116;
         pcDataBS<-rbinom(nSim,116,33/116)/116;
     else
         ptDataBS<-rbeta(nSim, 41, 116-41);</pre>
         pcDataBS<-rbeta(nSim, 33, 116-33);
     }
    evpi=(sum((ptDataBS-pcDataBS)[ptDataBS>pcDataBS])-sum(ptDataBS-pcDataBS))/nSim*wtp;
    ptDataBSBS<-rbinom(nSim,n,ptDataBS);</pre>
    pcDataBSBS<-rbinom(nSim,n,pcDataBS);
b<-wtp*((41+ptDataBSBS)/(116+n)-(33+pcDataBSBS)/(116+n));</pre>
    res<-sum(b[b>0])/nSim;
evsi=res-wtp*(41/116-33/116);
    return(list(evpi=evpi,evsi=evsi));
}
#Note that we use the calls to the eviNP function to calculate both EVPI and EVSI. Because evpi is the same regardless of the
sample size of the future RCT, we average the EVPIs for
different sample sizes to gain precision:
main<-function()
     sampleSize<<-(0:10)*100;
    resWP<<-rep(0,length(sampleSize));
resNPMulti<<-resWP;</pre>
    resNPDiri<<-resWP;
    evpiWP<<-Di(b0,v0);
    evpiMulti<<-0;
```

evpiDiri<<-0;

```
for(i in 1:length(sampleSize))
{
    resWP[i]<<-evsiWP(b0,v0,sigma,sampleSize[i]);
    temp<-eviNP(sampleSize[i],1000000,bsType=BS_TYPE_MULTINOMIAL);
    evpiMulti<-evpiMulti+tempSevpi;
    resNPMulti[i]<<-temp$evsi;
    temp<-eviNP(sampleSize[i],1000000,bsType=BS_TYPE_DIRICHLET);
    evpiDiri<<-evpiDiri+tempSevpi;
    resNPDiri[i]<<-temp$evsi;
}
evpiMulti<-evpiMulti/length(sampleSize);
evpiDiri<<-evpiDiri/complexize);
resWP[1]<<-0;
yLim<-c(0,Di(b0,v0));
g_range <- range(yLim, resWP, resNPMulti, resNPDiri);
plot(sampleSize,resWP,type='b',ylim=yLim,pch=21,ann=FALSE);
lines(sampleSize,resNPDiri,type='o',ylim=yLim,pch=23,lty=3);
</pre>
```

legend(1, g\_range[2], c("Parametric","Bayesian","Approximate Bayesian"), cex=0.8, pch=21:23, lty=1:3); title(ylab="EVSI",xlab="Sample size");

main();

#END;

۱

## Appendix A.3: Exact calculation of the Expected Value of Partial Perfect

### **Information for model 1**

Let  $P_0$ ,  $P_1$  and C be the probability of survival without treatment, probability of survival with treatment, and cost of treatment, respectively. Also let  $\lambda$  be the willingness-to-pay. The model inputs are represented by the following distributions:

$$P_0 \sim Beta(2,2)$$
  
 $P_1 \sim Beta(6,4)$  ,  
 $C \sim Unif(1000,2000)$ 

with  $\lambda = 50000$ . The incremental net benefit between the treatment and no-treatment decisions can be written as

$$INB = E(\lambda(P_1 - P_0) - C)$$

and its expected value as

$$E_{INB} = \lambda (\overline{P_1} - \overline{P_0}) - \overline{C}$$
  

$$\overline{P_1} = 0.5$$
  

$$\overline{P_0} = 0.6$$
  

$$\overline{C} = 1500$$
  

$$\Rightarrow E_{INB} = 3,500$$

**EVPPI for**  $P_0(pSurvival_NoRx)$ :  $E_{\Omega^-|P_0}(INB(P_0)) = \lambda . \overline{P_1} - \lambda . P_0 - \overline{C}$  is a linearly descending

function of  $P_0$  and hence it has one root, denote by  $R_{P_0}$ , on  $P_0$  which is:

$$\lambda(\overline{P_1} - P_0) - \overline{C} = 0 \Longrightarrow R_{P_0} = \overline{P_1} - \frac{\overline{C}}{\lambda} = 0.57$$

140

The *EVPPI* is the negative of the area under the segment of the curve of  $E_{\Omega^-|P_0}$  that is negative. Therefore,

$$EVPPI_{P_0} = -\int_{0.57}^{1} [50000.(0.6 - P_0) - 1500].d_{G(p_0)}$$
  
= -28500.[I(1,2,2) - I(0.57,2,2)] + 50000  $\int_{0.57}^{1} x^2.(1-x)/B(2,2).dx$ 

Where  $B(\alpha, \beta)$  and  $I(x, \alpha, \beta)$  indicate the beta function and the cumulative distribution function of the beta distribution, respectively, with parameters  $\alpha$  and  $\beta$ . Noting that I(1,2,2)=1, I(0.57,2,2)= 0.604314, B(2,2)=1/6, and  $\int x^2 (1-x) = \frac{(4-3x)x^3}{12}$ , we have

$$EVPPI_{P_0} = 3120.650$$

Note that the above calculations can easily be verified numerically in R, by calculating the EVPPI through numerical integration, and finding the area under the negative part of  $E_{\Omega^-|P_0}(INB(P_0))$  by subtracting the integral of the function from the integral of its absolute value (paste the below code in R command prompt):

```
f<-function(x) {return((50000*(0.6-x)-1500)*dbeta(x,2,2));}
af<-function(x) { return(abs(f(x)));}
(integrate(af,0,1)$value-integrate(f,0,1)$value)/2</pre>
```

**EVPPI for**  $P_1$  (*pSurvival\_Rx*):  $E_{\Omega^-|P_1}(INB(P_1)) = \lambda \cdot P_1 - \lambda \cdot \overline{P_0} - \overline{C}$  is a linearly ascending function of  $P_1$  and hence it has one root on  $P_1$ , denoted by  $R_P$ :

$$\lambda(P_1 - \overline{P_0}) - \overline{C} = 0 \Longrightarrow R_{P_1} = \frac{\overline{C}}{\lambda} + \overline{P_0} = 0.53$$

$$EVPPI_{P_1} = -\int_{0}^{0.53} [50000.(P_1 - 0.5) - 1500].d_{G(p_1)}$$
  
= 26500.[I(0.53,6,4) - I(0,6,4)] - 50000  $\int_{0}^{0.53} x^6.(1-x)^3 / B(6,4).dx$ 

With I(0,6,4)=0, I(0.53,6,4)= 0.3163517, B(6,4)=0.001984127, and  $\int x^6 (1-x)^3 = \frac{x^7}{7} - \frac{3x^8}{8} + \frac{x^7}{7} - \frac{x^7}{8} + \frac{x^7}{7} - \frac{x^8}{8} + \frac{x^8}{7} - \frac{x^8}{7} - \frac{x^8}{8} + \frac{x^8}{7} - \frac{x^8}{8} + \frac{x^8}{7} - \frac{x^8}{8} + \frac{x^8}{7} - \frac{x^8}{7} - \frac{x^8}{7} - \frac{x^8}{8} + \frac{x^8}{7} - \frac{x^8}{7$ 

 $\frac{x^9}{3} - \frac{x^{10}}{10}$ , we have

 $EVPPI_{P_1} = 1618.275$ 

Again, the above calculations can easily be verified numerically in R (paste the below code in R

command prompt):

f<-function(x) {return((50000\*(x-0.5)-1500)\*dbeta(x,6,4));}
af<-function(x) { return(abs(f(x)));}
(integrate(af,0,1)\$value-integrate(f,0,1)\$value)/2</pre>

EVPPI for C (cRx):

The root is

$$R_{c} = \lambda (\overline{P_{1}} - \overline{P_{0}}) = 5000$$

which is outside the range of the distribution of this parameter, therefore

 $EVPPI_{C} = 0$ 

The numerical verification in R (paste the below code in R command prompt):

```
f<-function(x) {return((50000*(0.6-0.5)-x)*dunif (x,1000,2000));}
af<-function(x) { return(abs(f(x)));}
(integrate(af,1000,2000)$value-integrate(f,1000,2000)$value)/2</pre>
```

## Appendix A.4: Excel add-in for single parameter EVPPI calculation

This is the Visual Basic (VB) code that can be copied into the VB editor in an Excel file. A fully functional add-in, with sample data and instructions to use is available from <a href="http://core.ubc.ca/software/voi">http://core.ubc.ca/software/voi</a>

# Note that the code also generates a p-value for segmentation. This is an experimental feature and is not discussed in this thesis.

```
' VOI Excel add-in by Mohsen Sadatsafavi
' Last update 04/04/2011
' Please cite 'Sadatsafavi M, Najafzadeh M, Bansback M, Sizto M, Sun H, Lynd LD, Marra C. A NOVEL METHOD FOR THE CALCULATION OF
THE EXPECTED VALUE OF PARTIAL PERFECT INFORMATION. 30th Annual Meeting of the Society for Medical Decision Making Abstracts.
(October 19-22, 2008). Philadelphia, USA. Available online at: http://smdm.confex.com/smdm/2008pa/webprogram/Paper4378.html*
 This program is free for use and distribution. In case you found a bug or other problems with the code please email msafavi at
interchange dot ubc dot ca
Option Explicit
Public Function EVPI(NBs As Range)
    Dim nbsLength, nNBs, y, temp, i, j, k, runSum, eMaxNb
     If IsNumeric(Right(NBs.Address, 1)) Then nbsLength = NBs.Rows.Count Else nbsLength = NBs.Cells(NBs.Rows.Count,
1).End(xlUp).Row
    nNBs = NBs.Columns.Count
    If Not IsNumeric(NBs(1, 1)) Then
  y = NBs.Range("A2:" & Chr(Asc("A") + nNBs - 1) & nbsLength)
Else
    y = NBs.Range("A1:" & Chr(Asc("A") + nNBs - 1) & nbsLength)
End If
    nbsLength = UBound(y, 1)
    If nbsLength < 100 Then
         EVPI = "Warning: too few samples"
         Exit Function
     End If
     If UBound(y, 2) = 1 Then
         temp = y
         ReDim v(nbsLength, 2)
         For i = 1 To nbsLength
           y(i, 1) = temp(i, 1)
             y(i, 2) = 0
         Nexti
         nNBs = nNBs + 1
    End If
    runSum = 0
    For i = 1 To nbsLength
        For k = 1 To nNBs
             If k = 1 Then temp = y(i, k) Else temp = WorksheetFunction.Max(temp, y(i, k))
         Next
         runSum = runSum + temp
    Next
    For i = 1 To nNBs
        t = 0
temp = 0
For k = 1 To nbsLength
             temp = temp + y(k, i)
         Next
         If i = 1 Then eMaxNb = temp Else If temp > eMaxNb Then eMaxNb = temp
    Next
    EVPI = (runSum - eMaxNb) / nbsLength
    Exit Function
kooni:
    MsgBox Err.Description, vbCritical
End Function
Public Function EVPPI(param As Range, NBs As Range, Optional sigLevel = 0.1)
    Dim minP, i, j, k, temp, v, minX, maxX, minWhere, maxWhere, x0, x1, zetaMin, zetaMax, zeta, p, runSum, eMaxNb, localMax
Dim nbsLength
    Dim nNBs
    Dim parmsLength
    Dim roots(100)
    Dim rootsN
     Dim rootsLeft(100)
    Dim rootsRight(100)
    rootsN = 1
```

```
minP = 1
```

```
If IsNumeric(Right(param.Address, 1)) Then parmsLength = param.Rows.Count Else parmsLength = param.Cells(param.Rows.Count,
1).End(xlUp).Row
    If ISNumeric(Right(NBS.Address, 1)) Then nbsLength = NBs.Rows.Count Else nbsLength = NBs.Cells(NBs.Rows.Count,
1).End(xlUp).Row
    If Not parmsLength = nbsLength Then
         EVPPI = "Error: input parameters not the same siZe"
         Exit Function
    End If
    nNBs = NBs.Columns.Count
    Dim x, y
    If Not IsNumeric(param(1, 1)) Then
    x = param.Range("A2:A" & parmsLength)
    y = NBs.Range("A2:" & Chr(Asc("A") + nNBs - 1) & parmsLength)
     Else
        x = param.Range("A1:A" & parmsLength)
y = NBs.Range("A1:" & Chr(Asc("A") + nNBs - 1) & parmsLength)
    End If
    parmsLength = UBound(x, 1) 'update the estimat eo flength given the above code
    If parmsLength < 100 Then
EVPPI = "Warning: too few samples"
         Exit Function
     End If
     If UBound(y, 2) = 1 Then
         temp = v
         ReDim y(parmsLength, 2)
For i = 1 To parmsLength
    y(i, 1) = temp(i, 1)
    v(i, 2)
         y(i, 2) = 0
Next i
         nNBs = nNBs + 1
     End If
     'ReDim x(parmsLength, 1)
     'ReDim y(parmslengh, nNBs)
    QuickSort x, y
    ReDim nbDiffs(parmsLength, nNBs * (nNBs - 1) / 2), partialSums(parmsLength, nNBs * (nNBs - 1) / 2)
    Dim counter
    counter = 1
     'MsgBox param(1, 1)
     On Error GoTo kooni
    For i = 1 To nNBs
For j = i + 1 To nNBs
v = 0
              For k = 1 To parmsLength
                  nbDiffs(k, counter) = y(k, i) - y(k, j)
If k = 1 Then
                       partialSums(k, counter) = nbDiffs(k, counter)
                        minX = partialSums(k, counter)
                        minWhere = 1
                        maxX = partialSums(k, counter)
                        maxWhere = 1
                   Else
                        partialSums(k, counter) = partialSums(k - 1, counter) + nbDiffs(k, counter)
If partialSums(k, counter) < minX Then</pre>
                            minX = partialSums(k, counter)
                            minWhere = k
                        End If
                       If partialSums(k, counter) > maxX Then
    maxX = partialSums(k, counter)
                             maxWhere = k
                       End If
                   End If
                  v = v + nbDiffs(k, counter) ^ 2 / parmsLength
              Next
              x0 = partialSums(1, counter)
              x1 = partialSums(parmsLength, counter)
              zetaMin = Application.WorksheetFunction.Min(x1 - minX, x0 - minX)
              zetaMax = Application.WorksheetFunction.Min(maxX - x0, maxX - x1)
              zeta = Application.WorksheetFunction.Max(zetaMin, zetaMax)
              p = 2 * Application.WorksheetFunction.NormDist(-zeta / Math.Sqr(v * parmsLength), 0, 1, True)
              If p < minP Then minP = p
If p < sigLevel Then
                   rootsN = rootsN + 1
If zetaMin > zetaMax Then
roots(rootsN) = minWhere
                   Else
                        roots(rootsN) = maxWhere
                   End If
              End If
              counter = counter + 1
```

```
Next
     Next
     roots(1) = 0
     rootsN = rootsN + 1
     roots(rootsN) = parmsLength
     runSum = 0
For i = 1 To rootsN - 1
For j = 1 To nNBs
              temp = 0
For k = roots(i) + 1 To roots(i + 1)
    temp = temp + y(k, j)
               Next
               If j = 1 Then localMax = temp Else If temp > localMax Then localMax = temp
          Next
          runSum = runSum + localMax
     Next
     For i = 1 To nNBs
          temp = 0
For k = 1 To parmsLength
              temp = temp + y(k, i)
          Next
          If i = 1 Then eMaxNb = temp Else If temp > eMaxNb Then eMaxNb = temp
     Next
     EVPPI = runSum - eMaxNb
     ReDim results(3, 3)
     results(0, 0) = EVPPI / parmsLength
results(1, 0) = minP
     results(0, 1) = results(0, 1)
results(2, 0) = zeta / Math.Sqr(v * parmsLength)
results(0, 2) = results(2, 0)
     EVPPI = results
     Exit Function
kooni:
     MsgBox Err.Description, vbCritical
End Function
Private Sub QuickSort (ByRef Values As Variant, Optional ByRef Attached As Variant, Optional ByVal Left As Long, Optional ByVal
Right As Long)
  Dim i As Long
  Dim j As Long
Dim k As Long
  Dim Item1 As Variant
  Dim Item2 As Variant
  Dim II
  On Error GoTo Catch
  If IsMissing(Left) Or Left = 0 Then Left = LBound(Values)
  If IsMissing(Right) Or Right = 0 Then Right = UBound(Values)
  i = Left
j = Right
  Item1 = Values((Left + Right) \setminus 2, 1)
  Do While i < j
    Do While Values(i, 1) < Item1 And i < Right
i = i + 1
     Loop
Do While Values(j, 1) > Item1 And j > Left
     j = j - 1
Loop
If i < j Then
      Call Swap(Values, i, j)
Call Swap(Attached, i, j)
     End If
If i <= j Then
     i = i + 1
j = j - 1
End If
  Loop
If j > Left Then Call QuickSort(Values, Attached, Left, j)
If i < Right Then Call QuickSort(Values, Attached, i, Right)
Catch:
  MsgBox Err.Description, vbCritical
End Sub
Private Sub Swap(ByRef Values As Variant, ByVal i As Long, ByVal j As Long)
  Dim Temp1 As Double
  Dim d, C
  d = UBound(Values, 2)
  d = UBound(Values, 2)
For C = 1 To d
Temp1 = Values(i, C)
Values(i, C) = Values(j, C)
Values(j, C) = Temp1
  Next C
End Sub
```

```
Sub ParseRange(Optional RefAddress As String = vbNullString, _
      Optional LeftColumn As String = vbNullString, _
Optional LeftRow As Long = 0, _
Optional RightColumn As String = vbNullString, _
      Optional RightRow As Long = VDWullstring, _
Optional RightRow As Long = 0)
Dim Aryl As Variant, Ary2 As Variant, N As Integer, Msg As String
Const Title As String = "Procedure 'ParseRange'"
On Error GoTo ErrMsg
If RefAddress = vDNullString _
Then RefAddress = ActiveWindow.RangeSelection.Address(, False)
       'Convert the address to column-absolute, row-absolute format
RefAddress = Application.ConvertFormula(Formula:=RefAddress, _
       FromReferenceStyle:=xlA1, _
       ToReferenceStyle:=xlA1, _
       ToAbsolute:=xlAbsolute)
      Ary1 = Split(RefAddress, "$")
Ary2 = Split(Ary1(2), ":")
LeftColumn = Ary1(1)
       LeftRow = Ary2(0)
       On Error Resume Next
       RightColumn = Ary1(3)
       RightRow = Ary1(4)
      GoTo Finish
ErrMsg:
      Select Case Err.Number
Case 438, 1004, 9: Msg = "A range is not currently selected or specified." & vbCr
Case Else: Msg = "An unexpected error occurred in macro 'ParseRange'." & vbCr
       End Select
       Msg = Msg & "Error number: " & Err.Number & vbCr & _
"Descrip: " & Err.Description
Resume Contin
Contin:
      MsgBox Msg, vbCritical, Title
       LeftColumn = vbNullString
      LeftRow = -1
RightColumn = vbNullString
      RightRow = -1
Finish:
      On Error Resume Next
Erase Ary1 'Release memory
On Error Resume Next
      Erase Ary2
End Sub
```

## Appendix A.5: R code for the efficient algorithm for EVPPI calculation

This is the function implemented in R for the method developed in Chapter 4. An exemplary analysis is embedded in the code and the results will be displayed after copying the code in the R environment.

#### Another copy of the code is available from http://www.core.ubc.ca/~msafavi/thesis

```
# R code for EVPPI calculation by Mohsen Sadatsafavi;
# Version 1.00
# Last update: 18/12/2011
# Usage: evppi(param,nbs,...);
# param: the nXl vector of the stochastic parameter (_theta_hat in the manuscript).
# nbs: the nXm matrix of corresponding net benefits for m decisions (NB_hat in the manuscript). If m=1, then the it is assumed
nbs are incremental net benefit between two decisions.
# other parameters: use to specify the number of segmentation points between pairs of decision. This information should be
# other parameters: use to specify the number of segmentation points between pairs of decision. This information should be submitted in the form of c(a,b,k) where a and b specify the pairs of strategies, and k (which should be 0, 1, or 2), is the number of segmentation points. For example, for a three-decision model, using the function evppi(x1,nbs, c(1,2,0), c(2,3,2)) forces the algorithm not to fit any segmentation points between decisions 1 and 2, and 2 segmentations between decisions 2 and 3, for parameter x1. The default is to fit 1 segmentation point for each pairs of decisions.
# code for model 1, and an example of evppi calculation using the evppi function are appended to the end;
evppi<-function(param, nbs, ...)
      n<-length(param);</pre>
     o<-order(param);
     param<-param[o];
     d<-dim(nbs)[2];</pre>
      if(is.vector(nbs))
           nbs<-cbind(nbs[0],0);</pre>
           d<-2;
     }
     if(d==1)
      {
           nbs<-cbind(nbs[0,],0);</pre>
           d<-2;
      else
           nbs<-nbs[o,];</pre>
      nSeqs<-matrix(1,d,d);</pre>
      exArgs<-list(...);
      for(obj in exArgs)
           if(is.null(names(obj)))
                if (length(obj)=
                      nSegs[1,2]<-obj
                 else
                      nSegs[obj[1],obj[2]]<-obj[3];</pre>
      }
     seqPoints<-c();
      for(i in 1:(d-1))
            for(j in (i+1):d)
                 message(paste('Fitting ',nSegs[i,j],' segmentation points for decisions ',i,j));
                 cm<-cumsum(nbs[,i]-nbs[,j])/n;</pre>
                 if(nSegs[i,j]==1)
                      l<-which.min(cm);</pre>
                      u<-which.max(cm);
                      if(cm[u]-max(cm[1],cm[n])>min(cm[1],cm[n])-cm[1])
                                        segPoint<-u
                               else
                                        segPoint<-l;
                      if (segPoint>1 && segPoint<n)
                            segPoints<-c(segPoints, segPoint);
                 }
                 if(nSegs[i,j]==2)
                      minL<-0;
                            maxL<-0;
```

```
minR<-0:
              maxR<-0;
                   aMinL<-array(0,n);</pre>
                   aMaxL<-array(0,n);
aMinR<-array(0,n);</pre>
                   aMaxR<-array(0,n);</pre>
                   for(k in 1:n)
                       minL<-min(minL,cm[k]);</pre>
                     maxL<-max(maxL, cm[k]);
minR<-min(minR, -cm[n]+cm[n-k+1]);</pre>
                       maxR<-max(maxR, -cm[n]+cm[n-k+1]);</pre>
                     aMinL[k]<-minL;
aMaxL[k]<-maxL;
aMinR[n-k+1]<-minR;
aMaxR[n-k+1]<-maxR;</pre>
              }
              sP<-aMaxL-aMinR;
                   sN<-aMaxR-aMinL:
              if(max(sP)>max(sN))
                   br<-which.max(sP);</pre>
                   segPoint<-c(which.max(cm[1:br]),br+which.min(cm[br:n]));</pre>
              else
              {
                   br<-which.max(sN);</pre>
                   segPoint<-c(which.min(cm[1:br]),br+which.max(cm[br:n]));</pre>
              }
              if (segPoint[1]>1 && segPoint[1]<n)
                   segPoints<-c(segPoints, segPoint[1]);</pre>
              if (segPoint[2]>1 && segPoint[2]<n)
    segPoints<-c(segPoints, segPoint[2]);</pre>
          }
         x11();
plot(param,cm);
          for(k in 1:length(segPoint))
              title(paste("Decision",i,"vs.",j));
     }
if(length(segPoints)>0)
     segPoints2<-c(0, segPoints[order(segPoints)], n);</pre>
     evppi<-0;
       for(j in 1:(length(segPoints2)-1))
    evppi<=evppi=max(colSums(nbs[(1+segPoints2[j]):segPoints2[j+1],]))/n;
evppi<=evppi=max(colMeans(nbs));</pre>
else
     evppi<-0;
  return(list(evppi=evppi,segPoints=segPoints));
```

#This is the model currently in the manuscript for which EVPPIs can be #calcuated analytically.

model1<-function(n,parmsIn=c(NA, NA, NA),exParms)</pre>

}

```
if(is.vector(parmsIn))
    parmsIn=matrix(parmsIn,1,3);
if(!exists('evppi.wtp'))
    evppi.wtp<<-50000;
if(length(parmsIn)>3 && length(parmsIn)!=n*3)
{
    cat('error: when input parameters are explicit they should agree with n','\n');
    return(NA);
}
if(length(parmsIn)==3)
```

```
{
    temp<-parmsIn;
    if(is.na(temp[,1]))
        p0<-repta(n,2,2)
    else
        p0<-rep(temp[,1],n);
    if(is.na(temp[,2]))
        p1<-rbeta(n,6,4)
    else
        p1<-rep(temp[,2],n);
    if(is.na(temp[,3]))
        cRx<-1000+runif(n)*1000
    else
        cRx<-rep(temp[,3],n);
    }
    else
    {
        p0<-parmsIn[,1];
        p1<-parmsIn[,2];
        cRx<-parmsIn[,2];
        cRx<-p0;
    uNoRx<-0;
    uNoRx<-0;
    uNoRx<-p0;
    uRx<-p1;
    parmsOut<-cbind(p0, p1, cRx);
    nmbs<-cbind(uNoRx*evpi.wtp-cNoRx,uRx*evpi.wtp-cRx);
    return(list(parmsOut,nmbs));
}
temp<-model1(10000);
params<_temp[[1]];
    nbs<-temp[[1];
    evpi(params[,1],nbs);
</pre>
```

```
#END;
```